



Centrum voor Wiskunde en Informatica

REPORTRAPPORT

MAS

Modelling, Analysis and Simulation



Modelling, Analysis and Simulation

Positivity for explicit two-step methods in linear multistep
and one-leg form

N.N. Pham Thi, W.H. Hundsdorfer, B.P. Sommeijer

REPORT MAS-E0522 OCTOBER 2005

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).

CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

Copyright © 2005, Stichting Centrum voor Wiskunde en Informatica

P.O. Box 94079, 1090 GB Amsterdam (NL)

Kruislaan 413, 1098 SJ Amsterdam (NL)

Telephone +31 20 592 9333

Telefax +31 20 592 4199

ISSN 1386-3703

Positivity for explicit two-step methods in linear multistep and one-leg form

ABSTRACT

Positivity results are derived for explicit two-step methods formulated in linear multistep form and in one-leg form. It turns out that the latter formulation allows a slightly larger step size with respect to positivity.

2000 Mathematics Subject Classification: 65L06

1998 ACM Computing Classification System: G.1.7

Keywords and Phrases: Positivity, Multistep Methods, One-Leg Form.

Note: The work of N.N.P.T and B.P.S was carried out under subtheme MAS1.1 - Applications from the Life Sciences. The work of W.H was carried out under theme MAS3 - Nonlinear Dynamics and Complex Systems.

POSITIVITY FOR EXPLICIT TWO-STEP METHODS IN LINEAR MULTISTEP AND ONE-LEG FORM

N. N. Pham Thi, W. Hundsdorfer, B. P. Sommeijer
N.N.Pham.Thi@cwi.nl, Willem.Hundsdorfer@cwi.nl, B.P.Sommeijer@cwi.nl

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Abstract

Positivity results are derived for explicit two-step methods formulated in linear multistep form and in one-leg form. It turns out that the latter formulation allows a slightly larger step size with respect to positivity.

2000 Mathematics Subject Classification: 65L06.

1998 ACM Computing Classification System: G.1.7.

Keywords and Phrases: Positivity, Multistep Methods, One-Leg Form.

Note: The work of N.N.P.T and B.P.S was carried out under subtheme MAS1.1 - Applications from the Life Sciences. The work of W.H was carried out under theme MAS3 - Nonlinear Dynamics and Complex Systems.

1 Introduction

We consider the initial value problem for a positive system of ordinary differential equations (ODEs) in \mathbb{R}^m

$$\begin{aligned}\mathbf{w}'(t) &= \mathbf{F}(t, \mathbf{w}(t)), \\ \mathbf{w}(0) &= \mathbf{w}_0 \geq 0.\end{aligned}$$

With positivity (actually, non-negativity) we mean that the solution vector $\mathbf{w}(t) \geq 0$, $\forall t > 0$ if $\mathbf{w}_0 \geq 0$. Here, and in the sequel, such inequalities are to be understood componentwise. For such systems of ODEs we will study whether we can obtain a similar property for the numerical solutions $\mathbf{W}_n \approx \mathbf{w}(t_n)$, $t_n = n\Delta t$, Δt being the time step. In [4], the related concept of monotonicity with semi-norms for linear multistep methods has been studied. Here we focus on positivity and adapt the results obtained in [4]. In Section 2 we will present an extension in the case of explicit two-step methods with forward Euler start-up (to compute \mathbf{W}_1), and we will point out the best method with respect to positivity, i.e. $\mathbf{W}_n \geq 0$ for $n \geq 1$, whenever $\mathbf{W}_0 \geq 0$. In Section 3 we consider the corresponding one-leg formulation and show that this allows a slightly larger step size.

2 Positivity for linear two-step methods

Consider the following explicit linear two-step scheme

$$\mathbf{W}_{n+2} = \sum_{j=0}^1 \left[-\alpha_j \mathbf{W}_{n+j} + \beta_j \Delta t \mathbf{F}(t_{n+j}, \mathbf{W}_{n+j}) \right]. \quad (1a)$$

Observe that the freedom in scaling the coefficients has been used to set the coefficient in front of \mathbf{W}_{n+2} equal to 1. In the one-leg formulation we will use a different scaling.

The scheme (1a) is of second-order accuracy if

$$\alpha_0 = 1 - \xi, \quad \alpha_1 = \xi - 2, \quad \beta_0 = \frac{\xi}{2} - 1, \quad \beta_1 = \frac{\xi}{2} + 1, \quad (1b)$$

where ξ is a free parameter. We note that the scheme is zero-stable (stable for the trivial equation $\mathbf{w}'(t) = \mathbf{0}$, see [5]) if the condition $-1 \leq \alpha_0 < 1$ is satisfied, i.e. if $0 < \xi \leq 2$. In the remainder of this paper we shall always deal with methods that are second-order accurate and zero-stable. In [4], both implicit and explicit methods have been analyzed. In this section we will extend the results obtained in that paper for the explicit methods. For monotonicity results with higher-order methods, we refer to [2, 3].

Following Shu [7], the step in (1a) is written as a linear combination of scaled forward Euler steps yielding

$$\mathbf{W}_{n+2} = - \sum_{j=0}^1 \alpha_j \left[\mathbf{W}_{n+j} + c_j \Delta t \mathbf{F}(t_{n+j}, \mathbf{W}_{n+j}) \right], \quad c_j = -\frac{\beta_j}{\alpha_j}. \quad (2)$$

We define Δt_{FE} to be the largest time step for which the forward Euler method, starting from a positive value, yields a positive result, i.e.

$$\mathbf{v} + \Delta t \mathbf{F}(t, \mathbf{v}) \geq 0 \quad \text{for all } \mathbf{v} \geq 0, \quad t \geq 0, \quad 0 \leq \Delta t \leq \Delta t_{FE}. \quad (3)$$

Then, if

$$\beta_j \geq 0 \quad \text{and} \quad \alpha_j \leq 0, \quad \text{i.e. } c_j \geq 0, \quad \text{for } j = 0, 1, \quad (4)$$

the terms within the square brackets in (2) are non-negative under the step size restriction $0 \leq c_j \Delta t \leq \Delta t_{FE}$, $j = 0, 1$. Therefore, $\mathbf{W}_{n+2} \geq 0$ for all $\Delta t \leq \min(\frac{1}{c_0}, \frac{1}{c_1}) \Delta t_{FE}$, for arbitrary values of $\mathbf{W}_0, \mathbf{W}_1, \dots, \mathbf{W}_{n+1} \geq 0$.

However, for the class of explicit second-order two-step methods, condition (4) for β_0 leads to $\xi \geq 2$. Combining this with the zero-stability requirement $0 < \xi \leq 2$ gives $\xi = 2$ as the only possible value. This, however, results in $c_1 = \infty$ and hence $\Delta t \leq 0$. Indeed, for $\xi = 2$ we obtain

$$\mathbf{W}_{n+2} = \left[\mathbf{W}_n - \mathbf{W}_{n+1} \right] + \left[\mathbf{W}_{n+1} + 2\Delta t \mathbf{F}(t_{n+1}, \mathbf{W}_{n+1}) \right].$$

Although the second term gives a positive contribution for $\Delta t \leq \frac{1}{2} \Delta t_{FE}$, the first term can be negative for arbitrary positive \mathbf{W}_n and \mathbf{W}_{n+1} which may result in $\mathbf{W}_{n+2} < 0$.

Fortunately, if we consider appropriate starting conditions, a positive result can be obtained [4, 3]. If \mathbf{W}_1 is obtained by the forward Euler method, i.e.

$$\mathbf{W}_1 = \mathbf{W}_0 + \Delta t \mathbf{F}(t_0, \mathbf{W}_0), \quad (5)$$

we have $\mathbf{W}_1 \geq 0$ for all $\Delta t \leq \Delta t_{FE}$ (see (3)). By introducing a non-negative parameter θ , which is specified later, and subsequently subtracting and adding $\theta^j \mathbf{W}_{n+2-j}$, $j = 1, 2, \dots, n+1$, in (1a), in which the added terms with $j = 1, 2, \dots, n$ are again written in the form of (1a), we arrive at

$$\begin{aligned} \mathbf{W}_{n+2} &= (-\alpha_1 - \theta) \mathbf{W}_{n+1} + \beta_1 \Delta t \mathbf{F}_{n+1} \\ &\quad + \sum_{j=0}^{n-1} \theta^j \left[(-\alpha_0 - \theta \alpha_1 - \theta^2) \mathbf{W}_{n-j} + (\beta_0 + \theta \beta_1) \Delta t \mathbf{F}_{n-j} \right] \\ &\quad + \theta^{n-1} \left[\theta^2 \mathbf{W}_1 - \theta \alpha_0 \mathbf{W}_0 + \theta \beta_0 \Delta t \mathbf{F}_0 \right], \quad n \geq 0, \end{aligned} \quad (6)$$

where \mathbf{F}_j denotes $\mathbf{F}(t_j, \mathbf{W}_j)$. Since \mathbf{W}_1 was calculated by the forward Euler method and $\alpha_1 = -1 - \alpha_0$ (see (1b)), this relation can be written as

$$\begin{aligned} \mathbf{W}_{n+2} &= (-\alpha_1 - \theta) \mathbf{W}_{n+1} + \beta_1 \Delta t \mathbf{F}_{n+1} \\ &\quad + \sum_{j=0}^{n-1} \theta^j \left[(1 - \theta)(\theta - \alpha_0) \mathbf{W}_{n-j} + (\beta_0 + \theta \beta_1) \Delta t \mathbf{F}_{n-j} \right] \\ &\quad + \theta^n \left[(\theta - \alpha_0) \mathbf{W}_0 + (\theta + \beta_0) \Delta t \mathbf{F}_0 \right], \quad n \geq 0. \end{aligned}$$

Considering this step as a linear combination of scaled forward Euler steps, we see that $\mathbf{W}_{n+2} \geq 0$ if all coefficients are non-negative, i.e.

$$-\alpha_1 - \theta \geq 0, \quad \beta_1 \geq 0, \quad (1 - \theta)(\theta - \alpha_0) \geq 0, \quad \beta_0 + \theta \beta_1 \geq 0, \quad \theta - \alpha_0 \geq 0, \quad \theta + \beta_0 \geq 0. \quad (7)$$

These conditions imply the step size restriction $\Delta t \leq \gamma(\theta) \Delta t_{FE}$, where

$$\gamma(\theta) = \min \left(\frac{-\alpha_1 - \theta}{\beta_1}, \frac{(1 - \theta)(\theta - \alpha_0)}{\beta_0 + \theta \beta_1}, \frac{\theta - \alpha_0}{\theta + \beta_0} \right) =: \min(A(\theta), B(\theta), C(\theta)). \quad (8)$$

Obviously, the larger $\gamma(\theta)$, the better are the positivity properties of the scheme.

The conditions (7) define an eligible θ -interval, viz. $\theta \in [\theta_{min}, \theta_{max}]$, where

$$\begin{aligned} \theta_{min} &= \max(\alpha_0, -\frac{\beta_0}{\beta_1}, -\beta_0) = -\beta_0, \\ \theta_{max} &= \min(-\alpha_1, 1). \end{aligned}$$

Observe that $A(\theta)$, $B(\theta)$ and $C(\theta)$ are monotonic decreasing functions of θ (recall the condition $0 < \xi \leq 2$). Therefore, we obtain the maximal $\gamma(\theta)$ -value

$$\gamma_{max} = \min(A(\theta_{min}), B(\theta_{min}), C(\theta_{min})) = \begin{cases} B(\theta_{min}) = \frac{\xi}{2-\xi} & \text{if } 0 < \xi \leq \frac{2}{3}, \\ A(\theta_{min}) = \frac{2-\xi}{2+\xi} & \text{if } \frac{2}{3} \leq \xi \leq 2. \end{cases} \quad (9)$$

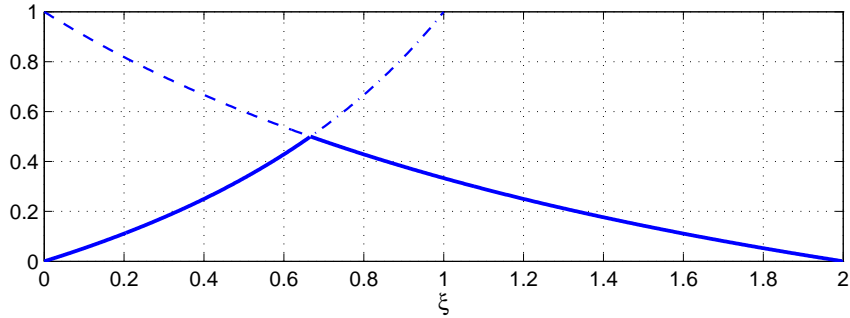


Figure 1: γ_{max} (solid), $A(\theta_{min})$ (dashed), and $B(\theta_{min})$ (dash-dotted) as functions of ξ .

The result is plotted in Figure 1. The ascending part of the γ_{max} -curve (i.e. for $0 < \xi < \frac{2}{3}$) is an extension to the work in [4]. We note that in that paper only the minimum of $A(\theta)$ and $B(\theta)$ was considered in (8), leading to a different value of θ_{min} . The forward Euler starting procedure (5) was introduced afterwards, but this does not lead to a positivity result for $0 < \xi < \frac{2}{3}$.

From Figure 1 we see that, within the class of explicit second-order two-step method, the optimal method with respect to positivity is the $\xi = \frac{2}{3}$ method (known as the extrapolated BDF2 method [5]). The resulting value for γ_{max} is $\frac{1}{2}$.

Remark. In (6), the sequence of subtracting and adding $\theta^j \mathbf{W}_{n+2-j}$ was performed until $j = n + 1$. In [4] these terms were subtracted and added up to $j = n$. It has been proved [6] that the latter choice has no advantages compared with the choice made in (6), i.e., does not lead to a more relaxed condition on Δt . The proof is rather lengthy and technical and therefore is not included in this paper.

3 Positivity for one-leg methods

One-leg schemes were introduced by Dahlquist [1] to facilitate the analysis of linear multistep methods. Therefore, it is of interest to study the positivity properties of methods when formulated in the one-leg form. Similar to the preceding section, we will consider explicit methods. We will see that the results are slightly better than those derived for the linear multistep formulation.

A natural scaling for one-leg methods is to require $\beta_0 + \beta_1 = 1$. Starting from the linear multistep formulation (1) we multiply the coefficients by a factor $\frac{1}{\xi}$ to obtain

$$\alpha_2 \mathbf{W}_{n+2} = \sum_{j=0}^1 \left[-\alpha_j \mathbf{W}_{n+j} + \beta_j \Delta t \mathbf{F}(t_{n+j}, \mathbf{W}_{n+j}) \right], \quad (10a)$$

where

$$\alpha_0 = \frac{1}{\xi} - 1, \quad \alpha_1 = 1 - \frac{2}{\xi}, \quad \alpha_2 = \frac{1}{\xi}, \quad \beta_0 = \frac{1}{2} - \frac{1}{\xi}, \quad \beta_1 = \frac{1}{2} + \frac{1}{\xi}. \quad (10b)$$

Since $\xi > 0$ we have

$$0 < \alpha_2 = -(\alpha_1 + \alpha_0). \quad (11)$$

The one-leg form of (10a) reads

$$\begin{aligned} \alpha_2 \mathbf{W}_{n+2} &= -\alpha_1 \mathbf{W}_{n+1} - \alpha_0 \mathbf{W}_n + \Delta t \mathbf{F}(\bar{t}, \overline{\mathbf{W}}_{n+2}), \\ \overline{\mathbf{W}}_{n+2} &= \beta_1 \mathbf{W}_{n+1} + \beta_0 \mathbf{W}_n, \end{aligned} \quad (12)$$

where $\bar{t} = \beta_1 t_{n+1} + \beta_0 t_n = t_n + \beta_1 \Delta t$. This one-leg formulation is second-order accurate if the coefficients satisfy (10b).

Let us define

$$\mathbf{V}_n = \mathbf{W}_n - \theta \mathbf{W}_{n-1}, \quad \theta \in [0, 1), \quad n \geq 1. \quad (13)$$

Furthermore, we introduce the coefficients

$$\begin{aligned} \alpha_1^* &= -\alpha_1 - \alpha_2 \theta, & \alpha_2^* &= -\alpha_0 - \alpha_1 \theta - \alpha_2 \theta^2 = (1 - \theta)(\alpha_2 \theta - \alpha_0), \\ \beta_1^* &= \beta_1, & \beta_2^* &= \beta_0 + \beta_1 \theta. \end{aligned} \quad (14)$$

The parameter θ in (13) and (14) will be chosen such that the coefficients in (14) satisfy

$$\alpha_j^* \geq 0, \quad \beta_j^* \geq 0, \quad j = 1, 2. \quad (15)$$

Assuming positive starting values

$$\mathbf{V}_1 \geq 0 \text{ and } \mathbf{W}_1 \geq 0, \quad (16)$$

we have the following theorem.

Theorem 1. *Suppose that $\Delta t \leq \mathcal{C} \Delta t_{FE}$, with $\mathcal{C} = \min\left(\frac{\alpha_1^*}{\beta_1^*}, \frac{\alpha_2^*}{\beta_2^*}\right)$, and θ is such that the conditions (15) and (16) are satisfied. Then $\mathbf{V}_n \geq 0$ and $\mathbf{W}_n \geq 0$ for all $n \geq 1$.*

Proof. The formulae (12)–(13) give

$$\alpha_2 \mathbf{V}_{n+2} = \alpha_1^* \mathbf{V}_{n+1} + \alpha_2^* \mathbf{W}_n + \Delta t \mathbf{F}(\bar{t}, \overline{\mathbf{W}}_{n+2}), \quad (17)$$

$$\overline{\mathbf{W}}_{n+2} = \beta_1^* \mathbf{V}_{n+1} + \beta_2^* \mathbf{W}_n. \quad (18)$$

Adding $\mathcal{C} \overline{\mathbf{W}}_{n+2}$ to both sides in equation (17) we obtain

$$\alpha_2 \mathbf{V}_{n+2} = (\alpha_1^* - \mathcal{C} \beta_1^*) \mathbf{V}_{n+1} + (\alpha_2^* - \mathcal{C} \beta_2^*) \mathbf{W}_n + \mathcal{C} \overline{\mathbf{W}}_{n+2} + \Delta t \mathbf{F}(\bar{t}, \overline{\mathbf{W}}_{n+2}).$$

The coefficients in this relation are non-negative, due to the definition of \mathcal{C} and (11). Therefore, $\mathbf{V}_{n+2} \geq 0$ if

$$\mathbf{V}_{n+1} \geq 0, \quad \mathbf{W}_n \geq 0, \quad \mathcal{C} \overline{\mathbf{W}}_{n+2} + \Delta t \mathbf{F}(\bar{t}, \overline{\mathbf{W}}_{n+2}) \geq 0. \quad (19)$$

The term $\mathcal{C} \overline{\mathbf{W}}_{n+2} + \Delta t \mathbf{F}(\bar{t}, \overline{\mathbf{W}}_{n+2})$ can be seen as a scaled forward Euler step. Thus, it is non-negative if $\overline{\mathbf{W}}_{n+2} \geq 0$ and $\Delta t \leq \mathcal{C} \Delta t_{FE}$. From (18) and (15) we see that $\overline{\mathbf{W}}_{n+2} \geq 0$ if

$$\mathbf{V}_{n+1} \geq 0 \quad \text{and} \quad \mathbf{W}_n \geq 0. \quad (20)$$

Combining (19) and (20) we have

$$\mathbf{V}_{n+2} \geq 0 \text{ if } \mathbf{V}_{n+1} \geq 0 \text{ and } \mathbf{W}_n \geq 0. \quad (21)$$

By assumption, we know that $\mathbf{V}_1 \geq 0$, $\mathbf{W}_1 \geq 0$ (see (16)) and $\mathbf{W}_0 \geq 0$. Thus, (21) yields $\mathbf{V}_2 \geq 0$. As a result, relation (13) gives $\mathbf{W}_2 = \mathbf{V}_2 + \theta\mathbf{W}_1 \geq 0$. Having $\mathbf{V}_2 \geq 0$ and $\mathbf{W}_1 \geq 0$, we obtain $\mathbf{V}_3 \geq 0$ (again by (21)) which results in $\mathbf{W}_3 = \mathbf{V}_3 + \theta\mathbf{W}_2 \geq 0$, etc. for all $n \geq 4$. \square

Let us now return to assumption (16) on the starting values. If \mathbf{W}_1 is calculated by the forward Euler method then we have $\mathbf{W}_1 \geq 0$ for all $\Delta t \leq \Delta t_{FE}$. Moreover, $\mathbf{V}_1 = \mathbf{W}_1 - \theta\mathbf{W}_0 = (1 - \theta)\mathbf{W}_0 + \Delta t\mathbf{F}_0 \geq 0$ under the additional step size restriction $\Delta t \leq (1 - \theta)\Delta t_{FE}$.

Using the above considerations we can formulate the following theorem on the positivity condition for the one-leg method.

Theorem 2. *If \mathbf{W}_1 is obtained by the forward Euler method (5) and θ is such that condition (15) is satisfied, then the one-leg method (12) is positive under the step size restriction $\Delta t \leq \gamma^{OL}(\theta)\Delta t_{FE}$ where*

$$\gamma^{OL}(\theta) = \min(\mathcal{C}, 1 - \theta) = \min\left(\frac{-\alpha_1 - \alpha_2\theta}{\beta_1}, \frac{(1 - \theta)(\alpha_2\theta - \alpha_0)}{\beta_0 + \beta_1\theta}, 1 - \theta\right). \quad (22)$$

It is illustrative to compare this $\gamma^{OL}(\theta)$ with the $\gamma(\theta)$ derived in (8): Condition (15) gives $\theta \in [\theta_{min}, \theta_{max}]$, where

$$\begin{aligned} \theta_{min} &= \max\left(\frac{\alpha_0}{\alpha_2}, -\frac{\beta_0}{\beta_1}\right) = -\frac{\beta_0}{\beta_1}, \\ \theta_{max} &= \min\left(-\frac{\alpha_1}{\alpha_2}, 1\right). \end{aligned}$$

Observe that the terms in the minimum function in (22) are monotonic decreasing functions of θ . Therefore, the optimal $\gamma^{OL}(\theta)$ -value is obtained at $\theta = \theta_{min} = \frac{2-\xi}{2+\xi}$ and is given by

$$\gamma_{max}^{OL} = \min\left(\frac{2(1 + \xi)(2 - \xi)}{(2 + \xi)^2}, \frac{2\xi}{2 + \xi}\right). \quad (23)$$

The result is plotted in Figure 2. From this figure we see that the best method with respect to positivity is no longer the method with $\xi = \frac{2}{3}$. The optimal method with respect to positivity is now the method with $\xi = \frac{1}{4}(\sqrt{17} - 1) \approx 0.78$. The corresponding γ_{max}^{OL} is then $\frac{1}{2}(\sqrt{17} - 3) \approx 0.56$. Comparing (9) and (23) we see that the one-leg method allows a slightly larger time step than the linear two-step method.

Acknowledgement. The investigations of N.N.P.T. were supported by the Computational Science program, which is subsidized by the Netherlands Organization for Scientific Research (NWO). W.H. and B.P.S. acknowledge support from the Dutch BSIK/BRICKS project.

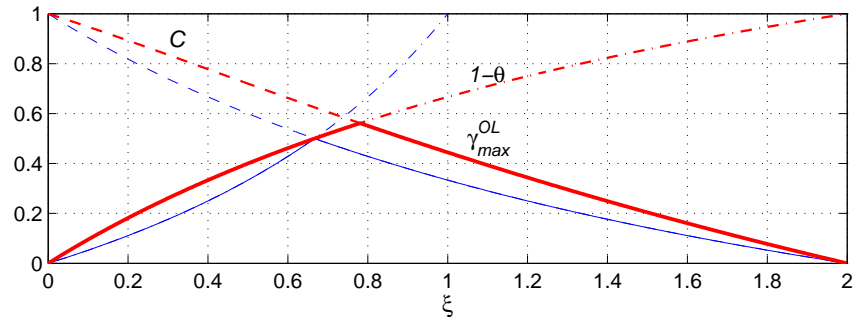


Figure 2: Step size restriction for positivity of the one-leg methods (thick lines) and of the linear two-step methods (thin lines, obtained from Figure 1).

References

- [1] G. Dahlquist, *Error analysis for a class of methods for stiff non-linear initial value problems*, Procs. Dundee Conf. 1975, Lecture Notes in Math. 506, G.A. Watson (ed.), Springer, Berlin (1976) 60-74.
- [2] S. Gottlieb, C.-W. Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Review 43 (2001) 89-112.
- [3] W. Hundsdorfer, S. J. Ruuth, *On monotonicity and boundedness properties of linear multistep methods*, Report MAS-E0404 (2004), CWI, Amsterdam, to appear in Math. Comp.
- [4] W. Hundsdorfer, S. J. Ruuth, and R. J. Spiteri, *Monotonicity-preserving linear multistep methods*, SIAM J. Numer. Anal. 41 (2003) 605-623.
- [5] W. Hundsdorfer, J. G. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, Springer Series in Computational Mathematics 33, Springer-Verlag (2003).
- [6] N. N. Pham Thi, *On the optimal way to rewrite the recursion of a linear multistep method to include the starting values*, private communication.
- [7] C.-W. Shu, *Total-variation-diminishing time discretizations*, SIAM J. Sci. Stat. Comput. 9 (1988) 1073-1084.