Extending nonstandard finite difference scheme rules to systems of nonlinear ODEs with constant coefficients

Marc E. Songolo[a*] and Brigitte Bidégaray-Fesquet[b]

[a]Department of Mathematics and Computer Science,
University of Lubumbashi, 1825 Lubumbashi, Democratic Republic of the Congo;
[b]Univ. Grenoble Alpes, Grenoble INP[†], LJK, 38000 Grenoble, France

Abstract. In this paper, we present a reformulation of Mickens' rules for nonstandard finite difference (NSFD) schemes to adapt them to systems of ODEs. This leads to exact schemes in the linear case, and also improves the accuracy in the nonlinear case. In the Hamiltonian nonlinear case, it consists in adding correction terms to schemes derived by Mickens.

Keywords. Ordinary Differential Equations, Matrix exponential, Exact finite difference schemes, Nonstandard finite difference schemes.

1 Introduction

The nonstandard finite difference (NSFD) method was created to overcome the numerical instabilities that exist in the construction of finite difference schemes for ordinary differential equations (ODE). The numerical instabilities indicate that the discrete equations are not able to model correctly the qualitative properties of the solutions to the differential equations [Mic94]. The rules for constructing NSFD models were derived from the construction of exact schemes for certain equations. These rules were then applied to many types of ODEs and to partial derivatives in order to obtain stable schemes which preserve the qualitative properties of the equations. The reference [Pat16] reviews many recent developments and applications of NSFD schemes.

In this paper, we propose to revisit some of Mickens' rules in the light of recent works [Cie13, CR10, CR11, QT18, SBF18, SBF21] to be able to treat properly the case of systems of ODEs. Especially the second rule, that relates to the renormalization of the discretization step-size, does not *a priori* take into account coupling factors between the equations in a differential system. Indeed the renormalization of the time-step derived in [Mic94], and with

[*] Corresponding author. Email: marc.songolo@gmail.com

[†] Institute of Engineering Univ. Grenoble Alpes

which we will compare the method we derive, is a scalar one, that can take carefully into account the time evolution of at most one of the equations in a system of ODEs.

The paper is organized as follows: we present Mickens' basic rules in Section 2. Section 3 displays the construction and the analysis of NSFD schemes for systems of ODEs, from a matrix formulation to scalar forms. This leads to define two correctors, the effect of which we explore on two examples in Section 4. The discussion then sets out the possible difficulties in developing this strategy and the way to overcome them.

2 The Nonstandard Finite Difference Context

2.1 Nonstandard Finite Difference Rules

In this section, we first give the rules for the construction of NSFD schemes as proposed by Mickens [Mic94].

Rule 1. *The order of the discrete derivatives must be exactly equal to the order of the corresponding derivatives of the differential equations.*

Rule 2. *Denominator functions for the discrete derivatives must, in general, be expressed in terms of more complicated functions of the step-sizes than those conventionally used.*

Rule 3. *Nonlinear terms must, in general, be modeled non-locally on the computational grid or lattice.*

Rule 4. *Special solutions of the differential equations should also be special (discrete) solutions of the finite difference models.*

Rule 5. *The finite difference equations should not have solutions that do not correspond exactly to solutions of the differential equations.*

These rules initially apply to single differential equations. Already in [Mic94] the case of Hamiltonian equations treated as systems of two first order equations makes use of a slightly modified version of Rule 2. Indeed, the derivatives are approximated by

$$\frac{\mathrm{d}x}{\mathrm{d}t} \simeq \frac{x_{k+1} - \psi(\Delta t)x_k}{\phi(\Delta t)}, \tag{1}$$

where ψ and ϕ are the functions of the step-size Δt and the parameters of the equations. As suggested by Rule 2, the denominator $\phi(\Delta t)$ plays the role of the step size and is such that

$$\phi(\Delta t) = \Delta t + \mathcal{O}(\Delta t^2) \text{ as } \Delta t \to 0.$$

There is an additional function ψ, which is not mentioned in Rule 2 and is close to identity, namely

$$\psi(\Delta t) = 1 + \mathcal{O}(\Delta t^2) \text{ as } \Delta t \to 0.$$

2.2 Nonstandard, Exact, and Best Finite Difference Schemes

Definition 1 (Nonstandard finite difference scheme [Mic00])**.** *A nonstandard finite difference scheme is any discrete representation of a system of differential equations that is constructed based on the above rules.*

This was the definition of a *best finite difference scheme* in [Mic94]. Originally, a best scheme referred to an exact scheme in [Pot82a]. An exact scheme has the same general solutions as the associated differential equations [Pot82a, Mic94]. The Mickens' rules have been defined to have *exact finite difference schemes* thus avoiding the usual questions about consistency, stability and convergence. This more or less involves that we know exact solutions of the equations (see Rules 4 and 5), which is of course not the case in general. It is however expected that schemes constructed with these rules would lead to the "best" schemes.

As in [AL01], we will study the stability of nonstandard finite difference schemes in the light of elementary stability. This concept combines classical stability with the fact that the fixed point solutions of the continuous and the discretized equations should coincide.

3 Nonstandard Finite Difference Schemes for systems

We address systems of ordinary differential equations which we split in a linear and a nonlinear part:

$$X'(t) = AX(t) + B(X(t)), \tag{2}$$

where $t \in [0,T]$, $X(t) \in \mathbb{R}^n$, $A \in \mathcal{M}_{n\times n}(\mathbb{R})$, and $B \in \mathcal{C}^0(\mathbb{R}^n, \mathbb{R}^n)$.

For the linear system $X'(t) = AX(t)$, it is known that exact finite difference methods may exist depending on the chosen step-size [Pot82b, RM13]. In what follows, we will suppose that A is invertible (see however Section 5.3).

We describe the method in the autonomous case, but this can be extended to non autonomous systems as the example in Section 4.4. Of course to ensure the well-posedness of the system, we suppose that B is Lipschitz, at least locally.

The analytical solution to system (2) can be expressed in integral form by

$$X(t+\Delta t) = e^{\Delta t A} X(t) + \int_t^{t+\Delta t} e^{(t+\Delta t - s)A} B(X(s)) \mathrm{d}s. \tag{3}$$

To go further in the explicit computations, we approximate $B(X(s))$ on the time interval $[t, t+\Delta t]$ by a function of $X(t)$ and $X(t+\Delta t)$:

$$B(X(s)) \simeq \mathcal{B}(X(t), X(t+\Delta t)).$$

Inserting this in (3)

$$\begin{aligned} X(t+\Delta t) &\simeq e^{\Delta t A} X(t) + \int_t^{t+\Delta t} e^{(t+\Delta t - s)A} ds \ \mathcal{B}(X(t), X(t+\Delta t)) \\ &= e^{\Delta t A} X(t) + (e^{\Delta t A} - I) A^{-1} \mathcal{B}(X(t), X(t+\Delta t)), \end{aligned}$$

where I is the identity matrix in $\mathcal{M}_{n\times n}(\mathbb{R})$. The simplest numerical method obtained by this formula is the *exponential Euler approximation* for which $\mathcal{B}(X(t), X(t+\Delta t)) = B(X(t))$:

$$X_{k+1} = e^{\Delta t A} X_k + (e^{\Delta t A} - I) A^{-1} B(X_k),$$

where X_k is approximating $X(t_k)$, $t_k = k\Delta t$ for $k \in \mathbb{N}$. This method makes use of a matrix exponential and is hence called *exponential integrator* [HO10]. Such an approximation for the nonlinear part does however not fulfill Rule 3 which advocates for a nonlocal discretization of the nonlinear part. We will therefore prefer the more general form

$$X_{k+1} = e^{\Delta t A} X_k + (e^{\Delta t A} - I) A^{-1} \mathcal{B}(X_k, X_{k+1}). \tag{4}$$

The Exponential integrator method has many advantages. Indeed, it is explicit, accurate and possesses good stability properties, even for large time steps. The basic idea of the exponential integrator method is to integrate exactly the linear part of the problem (which can be a stiff term), and then to use an appropriate approximation of the nonlinear part [HO10]. Thus, we can highlight a similarity between the NSFD method and the Exponential Integrator method, in the sense that the form of the denominator function for the discrete derivatives, given in the second Mickens rule, aims at obtaining an exact scheme for the linear term in the equation, while the non-local discretization of the nonlinear term, given in the third Mickens rule, is a way to obtain an adequate discretization for the nonlinear term. Besides, the Exponential Integrator method involves the computation of a matrix exponential or an exponential function of a matrix. So, the difficulties for the computation of the matrix exponential are largely responsible for the lack of interest in this method, especially for large systems.

3.1 Matrix formulation

In view of (4), we define the renormalisation matrix $\Phi(\Delta t) = (e^{\Delta t A} - I)A^{-1}$, and we can replace the exponential $e^{\Delta t A}$ by $I + \Phi(\Delta t)A$ in (4) to obtain

$$X_{k+1} = (I + \Phi(\Delta t)A)\, X_k + \Phi(\Delta t)\mathcal{B}(X_k, X_{k+1}),$$

or equivalently (if $\Phi(\Delta t)$ is invertible, which could be false for some exceptional values of Δt even if A is invertible, see the example of Section 4.1.1)

$$\Phi^{-1}(\Delta t)\,(X_{k+1} - X_k) = AX_k + \mathcal{B}(X_k, X_{k+1}), \tag{5}$$

where the renormalization matrix verifies the property

$$\Phi(\Delta t) = \Delta t I + \mathcal{O}(\Delta t^2) \text{ as } \Delta t \to 0.$$

Let us generalize (1) and hence Rule 2 to systems of ordinary differential equations. The scalar functions ϕ and ψ are then replaced by matrix-valued functions Φ and Ψ.

Rule 2'. *The first order derivatives in a nonstandard scheme for a system of ordinary differential equations should be approximated as*

$$\frac{\mathrm{d}X}{\mathrm{d}t} \simeq \Phi(\Delta t)^{-1}(X_{k+1} - \Psi(\Delta t)X_k),$$

where

$$\Phi(\Delta t) = \Delta t I + \mathcal{O}(\Delta t^2) \text{ as } \Delta t \to 0,$$

and

$$\Psi(\Delta t) = I + \mathcal{O}(\Delta t^2) \text{ as } \Delta t \to 0.$$

Scheme (5) is nonstandard. In particular, the discretization of the first order derivative corresponds to the above generalized rule, with $\Psi \equiv I$. The fact that we are able to write an exact or only a best scheme depends on the nonlinearity.

The major drawback of such a scheme is that we have to evaluate the exponential of matrix ΔtA. This can prove to be an expensive computation [MV03]. To overcome this difficulty we propose in the next section to reformulate scheme (5) in a scalar way.

3.2 Scalar formulation

3.2.1 Construction

To reformulate scheme (5), we consider the Cayley–Hamilton theorem, which implies that the exponential matrix can be rewritten as a finite expansion in powers of A:

$$e^{\Delta tA} = \alpha_0(\Delta t)I + \alpha_1(\Delta t)A + \alpha_2(\Delta t)A^2 + \cdots + \alpha_{n-1}(\Delta t)A^{n-1}, \tag{6}$$

where $\alpha_0(\Delta t), \alpha_1(\Delta t), \ldots, \alpha_{n-1}(\Delta t) \in \mathbb{R}$. The construction of these coefficients in the general case can be found in [MV03].

Introducing expansion (6) in the exponential integration scheme (4) yields

$$\begin{aligned} X_{k+1} = {} & \alpha_0(\Delta t)X_k + \alpha_1(\Delta t)[AX_k + \mathcal{B}(X_k, X_{k+1})] \\ & + \sum_{j=2}^{n-1} \alpha_j(\Delta t)A^{j-1}[AX_k + \mathcal{B}(X_k, X_{k+1})] \\ & + (\alpha_0(\Delta t) - 1)A^{-1}\mathcal{B}(X_k, X_{k+1}), \end{aligned}$$

which also reads

$$\begin{aligned} \frac{X_{k+1} - \alpha_0(\Delta t)X_k}{\alpha_1(\Delta t)} = {} & [I + R_1(\Delta t, A)][AX_k + \mathcal{B}(X_k, X_{k+1})] \\ & + R_0(\Delta t, A)\mathcal{B}(X_k, X_{k+1}), \end{aligned}$$

where we define the two correction factors

$$R_0(\Delta t, A) = \frac{\alpha_0(\Delta t) - 1}{\alpha_1(\Delta t)} A^{-1}, \quad R_1(\Delta t, A) = \sum_{j=2}^{n-1} \frac{\alpha_j(\Delta t)}{\alpha_1(\Delta t)} A^{j-1}. \tag{7}$$

In addition, we also introduce the notion of correction vectors,

$$\begin{aligned} T_0(\Delta t, A, X_k, X_{k+1}) &= R_0(\Delta t, A)\mathcal{B}(X_k, X_{k+1}), \\ T_1(\Delta t, A, X_k, X_{k+1}) &= R_1(\Delta t, A) \left[AX_k + \mathcal{B}(X_k, X_{k+1}) \right], \end{aligned}$$

to write the NSFD scheme as

$$\begin{aligned} \frac{X_{k+1} - \alpha_0(\Delta t)X_k}{\alpha_1(\Delta t)} &= AX_k + \mathcal{B}(X_k, X_{k+1}) \\ &\quad + T_0(\Delta t, A, X_k, X_{k+1}) + T_1(\Delta t, A, X_k, X_{k+1}). \end{aligned} \tag{8}$$

With regard to Mickens' second rule, we identify $\psi(\Delta t) = \alpha_0(\Delta t)$ and $\phi(\Delta t) = \alpha_1(\Delta t)$.

Remark 1. *If the system dimension is $n = 2$, $R_1(\Delta t, A) = 0$. In the case of a single equation ($n = 1$), the above formulation is not valid since $\alpha_1 \equiv 0$. For linear systems, the correction $T_0(\Delta t, A, X_k, X_{k+1})$ vanishes.*

3.2.2 Order estimate

Proposition 1. *The coefficients $\alpha_j(\Delta t)$ occurring in (6) verify*

$$\alpha_j(\Delta t) = \frac{\Delta t^j}{j!} + \mathcal{O}(\Delta t^n).$$

Proof. Let $S_{n-1}(\Delta t A)$ be the truncated expansion of $\exp(\Delta t A)$ in terms of powers of $\Delta t A$:

$$S_{n-1}(\Delta t A) = \sum_{j=0}^{n-1} \frac{\Delta t^j}{j!} A^j.$$

Then

$$\exp(\Delta t A) - S_{n-1}(\Delta t A) = \sum_{k=n}^{+\infty} \frac{\Delta t^k}{k!} A^k = \Delta t^n A^n \sum_{k=0}^{+\infty} \frac{\Delta t^k}{(n+k)!} A^k,$$

$$\| \exp(\Delta t A) - S_{n-1}(\Delta t A) \| \leq \Delta t^n \|A\|^n \sum_{k=0}^{+\infty} \frac{\Delta t^k}{k!} \|A\|^k = \Delta t^n \|A\|^n \exp(\Delta t \|A\|).$$

For $\Delta t \in]0, \Delta t_0]$, setting $C = \|A\|^n \exp(\Delta t_0 \|A\|)$,

$$\| \exp(\Delta t A) - S_{n-1}(\Delta t A) \| \leq C \Delta t^n.$$

The construction in [MV03] is based on the Cayley–Hamilton theorem. Matrix A^n can be written as a finite expansion in lower powers of A, defining coefficients c_j, $j = 0, \ldots, n-1$:

$$A^n = \sum_{j=0}^{n-1} c_j A^j.$$

This allows to define coefficients β_{kj}, $k \geq 0$, $j = 0, \dots, n-1$, such that

$$A^k = \sum_{j=0}^{n-1} \beta_{kj} A^j,$$

and the β_{kj} can be computed iteratively

$$\beta_{kj} = \begin{cases} \delta_{kj} & k < n, \\ c_j & k = n, \\ c_0 \beta_{k-1,n-1} & k > n, j = 0, \\ c_j \beta_{k-1,n-1} + \beta_{k-1,j-1} & k > n, j > 0. \end{cases}$$

Plugging this in the expansion of $\exp(\Delta t A)$ in terms of powers of $\Delta t A$, we obtain coefficients α_j:

$$\alpha_j(\Delta t) = \sum_{k=0}^{n-1} \frac{\Delta t^k}{k!} \beta_{kj} + \frac{\Delta t^n}{n!} \beta_{nj} + \sum_{k=n+1}^{\infty} \frac{\Delta t^k}{k!} \beta_{kj} = \frac{\Delta t^j}{j!} + \frac{\Delta t^n}{n!} c_j + \sum_{k=n+1}^{\infty} \frac{\Delta t^k}{k!} \beta_{kj}.$$

This implies that the expansion (6) is exactly $S_{n-1}(\Delta t A)$ at the precision $\mathcal{O}(\Delta t^n)$. □

Proposition 2. *The correction factors R_0 and R_1 have the following series expansion*

$$\begin{aligned} R_0(\Delta t, A) &= \frac{\Delta t^{n-1}}{n!} (-1)^{n-1} \det(A) A^{-1} + \mathcal{O}(\Delta t^n), \\ R_1(\Delta t, A) &= \sum_{j=2}^{n-1} \frac{\Delta t^{j-1}}{j!} A^{j-1} + \mathcal{O}(\Delta t^{n-1}). \end{aligned}$$

Proof. We compute

$$\frac{\alpha_0(\Delta t) - 1}{\alpha_1(\Delta t)} = \frac{\frac{\Delta t^n}{n!} c_0 + \mathcal{O}(\Delta t^{n+1})}{\Delta t + \mathcal{O}(\Delta t^n)} = \frac{\Delta t^{n-1}}{n!} c_0 + \mathcal{O}(\Delta t^n).$$

Besides $c_0 = (-1)^{n-1} \det(A)$. For $j \geq 2$,

$$\frac{\alpha_j(\Delta t)}{\alpha_1(\Delta t)} = \frac{\frac{\Delta t^j}{j!} + \mathcal{O}(\Delta t^{n+1})}{\Delta t + \mathcal{O}(\Delta t^n)} = \frac{\Delta t^{j-1}}{j!} + \mathcal{O}(\Delta t^{n-1}).$$

□

Remark 2. *In the same way that nothing ensures a priori that $\Phi(\Delta t)$ is invertible for all times, $\alpha_1(\Delta t)$ can vanish for some values of Δt. However the above expansions, and in particular the fact that $\alpha_1(\Delta t) \simeq \Delta t$, ensure that this is not the case for at least small enough values of Δt.*

3.2.3 Correction of the right-hand side

In Equation (8), Rule 2 in its generalized scalar form (1) is untouched, but now the discretization of the right-hand side is modified.

Rule 3'. *In an NSFD scheme for the nonlinear system* (2) *satisfying the classical Rule 2* (1)*, the usual nonlocal discretization of the right-hand side* $AX_k + \mathcal{B}(X_k, X_{k+1})$ *should be supplemented with correction terms:*

$$AX_k+\mathcal{B}(X_k, X_{k+1})+R_1(\Delta t, A)\left[AX_k + \mathcal{B}(X_k, X_{k+1})\right]+R_0(\Delta t, A)\mathcal{B}(X_k, X_{k+1}),$$

where R_0 *and* R_1 *are given by* (7)*.*

3.3 Stability issues

We follow here the same sketch of proof as in [AL01] adapting it to systems. They consider nonstandard schemes that derive from standard linear multistep methods. On the one hand we consider here a simpler case since we only consider one-step methods, but on the other hand we extend the time difference operator to functions $\Phi(\Delta t)$ which are more general than $\varphi(\Delta t)I$.

The condition for a classical linear one-step method to be consistent with equation $Y'(t) = F(Y)$ and stable is that it is a θ-scheme, i.e. there exists θ such that

$$Y_{k+1} - Y_k = \Delta t\left(\theta F(Y_{k+1}) + (1-\theta)F(Y_k)\right) \equiv \Delta t F_{\Delta t}(Y_k). \qquad (9)$$

The corresponding nonstandard scheme is

$$X_{k+1} - X_k = \Phi(\Delta t)\tilde{F}_{\Delta t}(X_k), \qquad (10)$$

where $\tilde{F}_{\Delta t}(X_k)$ can be nonlocal but is also an approximation of $F(X(t_k))$.

3.3.1 Consistency and zero-stability

The first issue is that both schemes have similar properties as $\Delta t \to 0$. More precisely we have the following theorem.

Theorem 1. *Equation* (10) *is consistent with equation* $Y'(t) = F(Y)$*. It is moreover stable if* $\tilde{F}_{\Delta t} = F_{\Delta t}$ *and is Lipschitz independently of* Δt *for bounded sequences of* Y_k*: for any* M *there exists* L *such that for all* Y_k^1*,* Y_k^2 *such that* $\|Y_k^1\| \leq M$ *and* $\|Y_k^2\| \leq M$

$$\sup_k \|F_{\Delta t}(Y_k^1) - F_{\Delta t}(Y_k^2)\| \leq L \sup_k \|Y_k^1 - Y_k^2\|.$$

Proof. Consistency. Let $t^* > 0$ a fixed time such that $t_k = t^*$ (which means that $k \to \infty$ as $\Delta t \to 0$). We compute

$$\begin{aligned}
&\Phi(\Delta t)^{-1}\Big(Y(t^*+\Delta t) - Y(t^*)\Big) - \tilde{F}_{\Delta t}(Y(t^*)) \\
&\qquad = \Phi(\Delta t)^{-1}\Delta t\left(\frac{Y(t^*+\Delta t) - Y(t^*)}{\Delta t} - F_{\Delta t}(Y(t^*))\right) \\
&\qquad\quad + \Phi(\Delta t)^{-1}\Delta t F_{\Delta t}(Y(t^*)) - \tilde{F}_{\Delta t}(Y(t^*)).
\end{aligned}$$

As $\Delta t \to 0$, $\Phi(\Delta t)^{-1}\Delta t \to I$. Hence the first term in the right-hand side goes to zero due to the fact that the standard scheme is consistent, and $\Phi(\Delta t)^{-1}\Delta t F_{\Delta t}(Y(t^*))$ and $\tilde{F}_{\Delta t}(Y(t^*))$ both tend to $F(Y(t^*))$, which proves the consistency of the nonstandard scheme.

Stability. Let δ_k and $\tilde{\delta}_k$ be two perturbations of the nonstandard scheme:

$$\begin{aligned} X_{k+1} - X_k &= \Phi(\Delta t)\big(F_{\Delta t}(X_k) + \delta_k\big), \\ \tilde{X}_{k+1} - \tilde{X}_k &= \Phi(\Delta t)\big(F_{\Delta t}(\tilde{X}_k) + \tilde{\delta}_k\big). \end{aligned}$$

We can write

$$\begin{aligned} X_{k+1} - X_k &= \Delta t F_{\Delta t}(X_k) + \left(\frac{\Phi(\Delta t)}{\Delta t} - I\right) F_{\Delta t}(X_k) + \frac{\Phi(\Delta t)}{\Delta t}\delta_k \\ &\equiv \Delta t F_{\Delta t}(X_k) + \gamma_k. \end{aligned}$$

In the same way, we can cast the equation for $\tilde{X}_k$ as

$$\tilde{X}_{k+1} - \tilde{X}_k = \Delta t F_{\Delta t}(\tilde{X}_k) + \tilde{\gamma}_k.$$

Since the standard scheme is stable, for Δt sufficiently small, there exists K such that for γ_k and $\tilde{\gamma}_k$ such that $\|\gamma_k - \tilde{\gamma}_k\| \le \varepsilon$, then $\|X_{k+1} - X_k\| \le K\varepsilon$. There remains to estimate $\|\gamma_k - \tilde{\gamma}_k\|$:

$$\gamma_k - \tilde{\gamma}_k = \left(\frac{\Phi(\Delta t)}{\Delta t} - I\right)\left(F_{\Delta t}(X_k) - F_{\Delta t}(\tilde{X}_k)\right) + \frac{\Phi(\Delta t)}{\Delta t}(\delta_k - \tilde{\delta}_k).$$

To estimate the first term in the right-hand side, we use the Lipschitz property (where the supremum is useful because $F_{\Delta t}$ can be nonlocal) and the fact that $\Phi(\Delta t) = \Delta t I + \mathcal{O}(\Delta t^2)$, i.e. that there exists C such that for all $X \in \mathbb{R}^n$, $\|\Phi(\Delta t)X - \Delta t X\| \le C\Delta t^2\|X\|$. Hence

$$\begin{aligned} \left\|\left(\frac{\Phi(\Delta t)}{\Delta t} - I\right)\left(F_{\Delta t}(X_k) - F_{\Delta t}(\tilde{X}_k)\right)\right\| &\le C\Delta t\|F_{\Delta t}(X_k) - F_{\Delta t}(\tilde{X}_k)\| \\ &\le CL\Delta t \sup_j \|X_j - \tilde{X}_j\|. \end{aligned}$$

For Δt sufficiently small, we can also ensure that the second term in the right-hand side is bounded:

$$\left\|\frac{\Phi(\Delta t)}{\Delta t}(\delta_k - \tilde{\delta}_k)\right\| \le 2\varepsilon.$$

We now take Δt sufficiently small to have $CL\Delta t \le 1/2K$ where K is anew the constant for the stability of the standard scheme. Eventually we have

$$\sup_k \|X_k - \tilde{X}_k\| \le K\left(\frac{1}{2K}\sup_j \|X_j - \tilde{X}_j\| + 2\varepsilon\right),$$

and hence

$$\sup_k \|X_k - \tilde{X}_k\| \le 4K\varepsilon,$$

which mean that $4K$ is a stability constant for the nonstandard scheme. □

3.3.2 Elementary stability

Although the standard and nonstandard schemes have the same properties as $\Delta t \to 0$, we moreover expect the nonstandard scheme to behave better for larger values of Δt, which is described by elementary stability.

Definition 2. *A scheme is elementary stable, if the fixed points of the scheme are those of the continuous equation and their linear stability is the same.*

We hence denote by $\tilde{Y} \in \mathbb{R}^n$ a fixed point of the continuous equation, i.e. $F(\tilde{Y}) = 0$. Fixed points $\tilde{X}$ of the non-standard scheme (10) should be solution to $\Phi(\Delta t)\tilde{F}_{\Delta t}(\tilde{X}) = 0$. Provided $\Phi(\Delta t)$ is invertible (which is true at least for small values of Δt) and the zeros of $\tilde{F}_{\Delta t}$ are those of $F(\Delta t)$, the continuous equation and the nonstandard scheme have the same fixed points.

4 Numerical tests

We will now study the impact of the corrections in various situations. According to Remark 1 we can find contexts where one or the other correction vanishes.

4.1 Impact of R_0

4.1.1 A quadratic nonlinear oscillator

We first study the impact of R_0. According to Remark 1, we therefore consider a system of two differential equations ($n = 2$), so that $R_1 \equiv 0$. In [Hu06] the quadratic nonlinear differential equation is presented as a good benchmark for numerical schemes

$$\begin{aligned} &x'' + x + x^2 = 0, \\ &x(0) = x_0 > 0, \\ &x'(0) = 0. \end{aligned} \tag{11}$$

This equation occurs for example in human eardrum oscillation modeling. It has the significant advantage to have a known exact solution for $x_0 < 1/2$, namely

$$x(t) = x_0 + a\,\mathrm{sn}^2(\omega t, m),$$

where

$$\begin{aligned} a &= \frac{-12x_0(1+x_0)}{\sqrt{3(1-2x_0)(3+2x_0)} + 3(1+2x_0)}, \\ \omega &= \frac{1}{2}\sqrt{\frac{1}{2} + x_0 + \frac{1}{6}\sqrt{3(1-2x_0)(3+2x_0)}}, \\ m &= \frac{1}{2} + \frac{3(2x_0^2 + 2x_0 - 1)}{3 + (1+2x_0)\sqrt{3(1-2x_0)(3+2x_0)}}, \end{aligned}$$

and sn is the Jacobi sine fonction. We write the second order equation (11) in the Hamiltonian form

$$\begin{cases} x' = y, \\ y' = -x - x^2. \end{cases} \tag{12}$$

We thus obtain a system of the form (2), with matrices

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \; B(X) = \begin{pmatrix} 0 \\ b(x) \end{pmatrix},$$

where $b(x) = -x^2$ and with initial data $x(0) = x_0$ and $y(0) = 0$.

Besides the comparison with an exact solution, two properties can be used to evaluate the quality of numerical methods. First the exact solutions to (11) are periodic with period

$$P = 4 \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - m^2 \sin^2 \theta}}.$$

Second, the differential equation (11) satisfies a conservation law:

$$E(t) \equiv \frac{1}{2}(x'(t))^2 + \frac{1}{2}x(t)^2 + \frac{1}{3}x(t)^3 = \frac{1}{2}x_0^2 + \frac{1}{3}x_0^3. \tag{13}$$

Several methods exist to obtain efficient schemes (called *best schemes* by Mickens [Mic94]) for a harmonic oscillator. We can cite the Gautschi type method [HL99], the exponential integration method [HO10], the gradient method [Cie11, Cie13, CR10, CR11], the NSFD method [Mic94, MOR05, MR94, Roe08]. Here, we are interested in the NSFD method, and the Potts method [Pot82b]. In [SBF18] the computation of α_0 and α_1 is made explicit for two-dimensional matrices in terms of the eigenvalues, namely $\pm i$ for the matrix involved in (12), leading to $\alpha_0(\Delta t) = \cos(\Delta t)$ and $\alpha_1(\Delta t) = \sin(\Delta t)$. This is coherent with the predictions of Proposition 1. Indeed, we have here $c_0 = -1$ and $c_1 = 0$, and this yields $\alpha_0(\Delta t) = 1 - \Delta t^2/2 + \mathcal{O}(\Delta t^3)$ and $\alpha_1(\Delta t) = \Delta t + \mathcal{O}(\Delta t^3)$. We are in a typical situation where $\alpha_1(\Delta t)$ can vanish for some values of Δt.

4.1.2 Mickens' scheme for Hamiltonian systems

In [Mic94], Mickens discretized Equation (12) as

$$\begin{cases} \dfrac{x_{k+1} - \cos(\Delta t)x_k}{\sin(\Delta t)} = y_k, \\[2ex] \dfrac{y_{k+1} - \cos(\Delta t)y_k}{\sin(\Delta t)} = -x_k - (x_{k+1})^2. \end{cases}$$

This scheme has the form (8), with $\mathcal{B}(X_k, X_{k+1}) = B(X_{k+1})$ and no correction term. Eliminating y_k leads to a discretization of (11):

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{\sin^2(\Delta t)} + \frac{2\,[1 - \cos(\Delta t)]\,x_k}{\sin^2(\Delta t)} + x_k^2 = 0. \tag{14}$$

The quantity $\dfrac{2\,[1-\cos(\Delta t)]}{\sin^2(\Delta t)}$ tends to 1 as $\Delta t \to 0$, but we want to have exactly 1 to have an exact computation of the linear term of Equation (11). To this aim, we notice that we can also cast (14) as

$$\frac{x_{k+1}-2x_k+x_{k-1}}{[2\sin(\Delta t/2)]^2}+x_k+\cos^2(\Delta t/2)x_k^2=0. \tag{15}$$

One of the consequences of the conservation law (13) is that any periodic solution oscillates with a constant amplitude. The Mickens scheme (15) has this property. It suffices to note that it is invariant for any transformation which swaps x_{k+1} and x_{k-1}.

However, in the case of a harmonic oscillator, it must also be shown that there is a constant first integral. Moreover it was established in [Mic94] that the discretization of the nonlinear term used in (15) does not make it possible to obtain a constant (discrete) integral. It is then advisable [Mic94] to discretize the non-linear term as

$$b(x) \approx -x_k\left(\frac{x_{k+1}+x_{k-1}}{2}\right)$$

and the new scheme is

$$\frac{x_{k+1}-2x_k+x_{k-1}}{[2\sin(\Delta t/2)]^2}+x_k+\cos^2(\Delta t/2)x_k\frac{x_{k+1}+x_{k-1}}{2}=0. \tag{16}$$

This would necessitate to take

$$\frac{y_{k+1}-\cos(\Delta t)y_k}{\sin(\Delta t)}=-x_k-x_{k+1}\frac{x_{k+2}+x_k}{2}$$

as a discretization for the second equation of the Hamiltonian form. The system is then highly implicit and uses time $t+2\Delta t$ when approximating $B(X(s))$ on the time interval $[t, t+\Delta t]$ in the integral of Equation (3).

4.1.3 Adding a correction term

Let us now see how the NSFD scheme of Section 3.2 reads for Equation (12). Recall that for $n=2$, $R_1(\Delta t, A)\equiv 0$. We also have already computed $\alpha_0(\Delta t)$ and $\alpha_1(\Delta t)$. The scalar NSFD schemes reads

$$\frac{X_{k+1}-\cos(\Delta t)X_k}{\sin(\Delta t)}=AX_k+\mathcal{B}(X_k,X_{k+1})-\tan(\Delta t/2)A^{-1}\mathcal{B}(X_k,X_{k+1}). \tag{17}$$

Up to possible choices for $\mathcal{B}(X_k,X_{k+1})$, and noticing that $-A^{-1}=A$, we find the same scheme as Mickens' but with a correction factor $\tan(\Delta t/2)A\mathcal{B}$. We follow the same steps as in the previous paragraph to obtain a scheme for the initial second order equation.

The scalar NSFD scheme for equation (12) reads

$$\begin{cases}\dfrac{x_{k+1}-\cos(\Delta t)x_k}{\sin(\Delta t)}=y_k+\tan(\Delta t/2)b(x_k,x_{k+1}),\\[2ex] \dfrac{y_{k+1}-\cos(\Delta t)y_k}{\sin(\Delta t)}=-x_k+b(x_k,x_{k+1}).\end{cases}$$

Combining these two equations in the same way than for Mickens' scheme first yields

$$\begin{aligned}\frac{x_{k+1}-2x_k+x_{k-1}}{\sin^2(\Delta t)} = & \frac{2\left[\cos(\Delta t)-1\right]x_k}{\sin^2(\Delta t)} + b(x_{k-1},x_k) \\ & + \frac{\tan(\Delta t/2)}{\sin(\Delta t)}\left[b(x_k,x_{k+1})-\cos(\Delta t)b(x_{k-1},x_k)\right].\end{aligned}$$

With the same transformation that led to (15), we find

$$\frac{x_{k+1}-2x_k+x_{k-1}}{[2\sin(\Delta t/2)]^2}+x_k=\frac{1}{2}[b(x_{k-1},x_k)+b(x_k,x_{k+1})]. \tag{18}$$

The $\cos^2(\Delta t/2)$ coefficient of Equations (15) or (16) has disappeared. Besides to obtain $-x_k\dfrac{x_{k-1}+x_{k+1}}{2}$ in the right-hand side, one has simply to choose $b(x_k,x_{k+1})=-x_kx_{k+1}$, which is a nonlocal discretization of the nonlinearity (and therefore complies to Rule 3) and is only semi-implicit. It only involves the present and the past but not the future, contrarily to what has been observed for Mickens' scheme (16).

In this test case the fixed points are not necessarily of first interest since we look for periodic solutions. However we can notice that the fixed points of the continuous equations are $x=0$ and $x=-1$. These are also the fixed points of our scheme (18) (and also of both Euler methods that are used in the simulations). The Mickens' schemes (15) and (16) both also have $x=0$ as fixed point, but the second one is $-1/\cos^2(\Delta t/2)$ which is different from -1 in general.

4.1.4 Numerical results

We now compare the previous numerical methods for $x_0=0.25$ which lies in the valid interval for initial data, namely $[0,1/2[$. The exact solution is given by $x(t)=x_0+a\,\mathrm{sn}^2(\omega t,m)$ with $a\simeq-0.55$, $\omega\simeq0.53$, and $m\simeq0.33$. Its time evolution over the time interval $[0,35]$ is displayed in Figure 1.

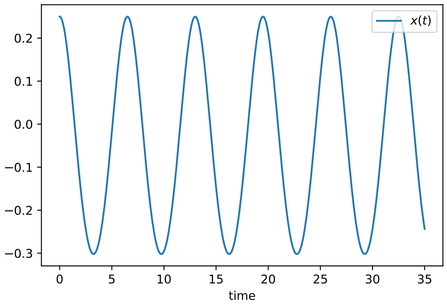


Figure 1: Time evolution of the exact solution of Equation (11) for $x_0=0.25$.

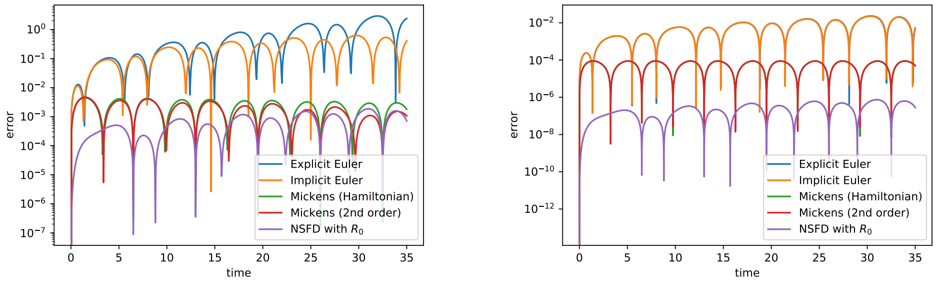
As already mentioned, this solution is periodic in time. For the tests we use different values of Δt and compute solutions *via*

- the explicit Euler scheme,
- the implicit Euler scheme,
- Mickens' scheme (15),
- Mickens' scheme (16),
- the scalar scheme with correction R_0 (18) with $b(x_k, x_{k+1}) = -x_k x_{k+1}$.

The advantage of comparing the methods with a quadratic nonlinearity is that we are able to compute explicitly the iterates for all these methods without adding methods to solve nonlinear systems such as predictor–correctors or fixed points. The time evolution of the relative error between the exact solution x_k^e and the computed solution x_k,

$$E_k = \frac{|x_k - x_k^e|}{x_k^e}, \tag{19}$$

is shown in Figure 2 for two values of the time-step.



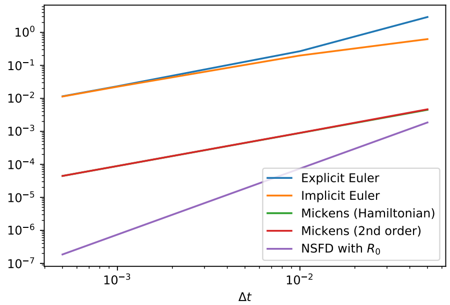
$\Delta t = 0.05$ $\Delta t = 0.001$

Figure 2: Time evolution of the relative errors for $\Delta t = 0.05$ and 0.001 for Equation (11) and $x_0 = 0.25$.

Of course, the three NSFD schemes, (15), (16), and (18), outperform the Euler schemes, in particular they do not show a deterioration of the error as time evolves.

The correction R_0 does indeed improve Mickens' original schemes, gaining more than two errors of magnitude for small Δt. For large Δt, the gain is not so clear, but this is due to the approximation of the nonlinearity, which is the only source of approximation in (18), and which is dominant for $\Delta t = 0.05$.

In Figure 3, we show the convergence of the five methods, plotting $\max_k E_k$ with respect to Δt, and the corresponding numerical orders are given in the adjacent Table. The result is clear in this test case, the additional term R_0 allows to gain one order of convergence.



Explicit Euler	1.18
Implicit Euler	0.88
Mickens (Hamiltonian)	1.00
Mickens (2nd order)	1.01
NSFD with R_0	2.00

Figure 3: Numerical order of the five methods for the quadratic non linear oscillator.

In Figure 4 we illustrate qualitative properties for a relatively large value of Δt. Left, the trajectory of the solution in the (x, y) phase space is plotted. We only plot the trajectories for the methods that use these variables and for which y has not to be reconstructed. The two nonstandard method do preserve periodic trajectories, which as expected Euler methods do not. On the right figure, we plot a numeric analogous of the conserved quantity $E(t)$ given by (13): $\frac{1}{2}y_k^2 + \frac{1}{2}x_k^2 + \frac{1}{3}x_k^3$. We only compare the Mickens' scheme in Hamiltonian form and our scheme. Once more our approach yields better results, and this is even more obvious for smaller values of Δt.

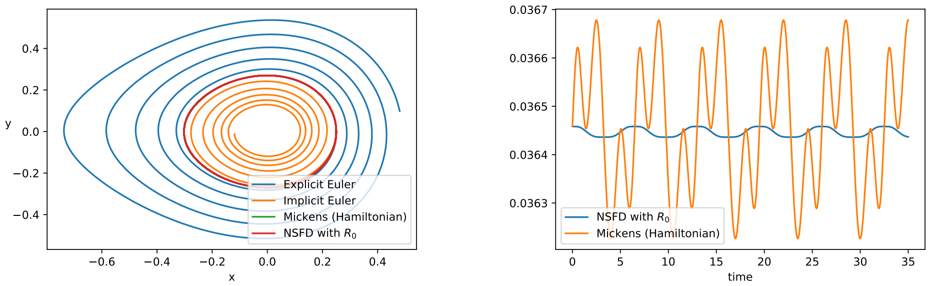


Figure 4: Qualitative properties for $\Delta t = 0.05$. Phase space (left); Time evolution of $E(t)$ (right).

4.2 Impact of R_1

4.2.1 A forest biomass model

To test the impact of R_1 only, we now consider a linear system with dimension greater than 2, namely $n = 3$. Again, according to Remark 1, $T_0 \equiv 0$ in this linear example. We use a simple example found in [GW98] dealing with the evolution of the forest biomass. More precisely, we denote $x(t)$ the biomass decayed into humus, $y(t)$ the biomass of dead trees, and $z(t)$ the biomass of living trees. The corresponding evolution equations are

$$\begin{cases} x'(t) = -x(t) + 3y(t), \\ y'(t) = -3y(t) + 5z(t), \\ z'(t) = -5z(t), \end{cases} \tag{20}$$

with an initial data where there are no dead trees and no humus at $t = 0$, namely $x(0) = 0$, $y(0) = 0$, and $z(0) = z_0$. The corresponding matrix A is

$$A = \begin{pmatrix} -1 & 3 & 0 \\ 0 & -3 & 5 \\ 0 & 0 & -5 \end{pmatrix}, \tag{21}$$

yielding the exact solution

$$\begin{cases} x(t) = \dfrac{15}{8}\left(e^{-t} - 2e^{-3t} + e^{-5t}\right) z_0, \\ y(t) = \dfrac{5}{2}\left(e^{-3t} - e^{-5t}\right) z_0, \\ z(t) = e^{-5t} z_0. \end{cases}$$

4.2.2 Derivation of correction terms

For the matrix A given by Equation (21), we have $A^3 = -15I - 23A - 9A^2$, i.e. $c_0 = -15$, $c_1 = -23$, and $c_2 = -9$. We therefore predict that $\alpha_0(\Delta t) = 1 - \frac{5}{2}\Delta t^3 + \mathcal{O}(\Delta t^4)$, $\alpha_1(\Delta t) = \Delta t - \frac{23}{6}\Delta t^3 + \mathcal{O}(\Delta t^4)$, and $\alpha_2(\Delta t) = \frac{1}{2}\Delta t^2 - \frac{3}{2}\Delta t^3 + \mathcal{O}(\Delta t^4)$. Writing

$$\left(\alpha_0(\Delta t)I + \alpha_1(\Delta t)A + \alpha_2(\Delta t)A^2\right) \begin{pmatrix} 0 \\ 0 \\ z_0 \end{pmatrix} = \begin{pmatrix} x(\Delta t) \\ y(\Delta t) \\ z(\Delta t) \end{pmatrix}$$

yields

$$\begin{aligned} \alpha_0(\Delta t) &= \frac{15}{8}e^{-\Delta t} - \frac{5}{4}e^{-3\Delta t} + \frac{3}{8}e^{-5\Delta t}, \\ \alpha_1(\Delta t) &= e^{-\Delta t} - \frac{3}{2}e^{-3\Delta t} + \frac{1}{2}e^{-5\Delta t}, \\ \alpha_2(\Delta t) &= \frac{1}{8}e^{-\Delta t} - \frac{1}{4}e^{-3\Delta t} + \frac{1}{8}e^{-5\Delta t}. \end{aligned}$$

These values do agree with the predicted expansions at order 3. Since there is no nonlinear part,

$$\begin{aligned} X_{k+1} &= \alpha_0(\Delta t)X_k + \alpha_1(\Delta t)AX_k + \alpha_1(\Delta t)T_1(\Delta t, A, X_k) \\ &= \alpha_0(\Delta t)X_k + \alpha_1(\Delta t)AX_k + \alpha_2(\Delta t)A^2X_k. \end{aligned} \tag{22}$$

4.2.3 Numerical results

For the numerical test case, we compare our method (22), which should be exact since no approximation has been done in its derivation, with the Euler explicit and implicit methods. We also compute

$$X_{k+1} = \gamma_0(\Delta t)X_k + \gamma_1(\Delta t)AX_k + \gamma_2(\Delta t)A^2X_k,$$

where the γ_j are the order 3 approximations of α_j, namely $\gamma_0(\Delta t) = 1-\frac{5}{2}\Delta t^3$, $\gamma_1(\Delta t) = \Delta t - \frac{23}{6}\Delta t^3$, and $\gamma_2(\Delta t) = \frac{1}{2}\Delta t^2 - \frac{3}{2}\Delta t^3$. Finally we derive a NSFD scheme on the above principles but for each equation separately, leading to

$$\begin{cases} \dfrac{x_{k+1} - x_k}{1 - e^{-\Delta t}} = -x_k + 3y_k, \\ \dfrac{y_{k+1} - y_k}{(1 - e^{-3\Delta t})/3} = -3y_k + 5z_k, \\ \dfrac{z_{k+1} - z_k}{(1 - e^{-5\Delta t})/5} = -5z_k. \end{cases} \tag{23}$$

The exact solution is computed over the time interval $[0, 10]$, corresponding to ten years of time evolution, and the result is displayed on Figure 5.

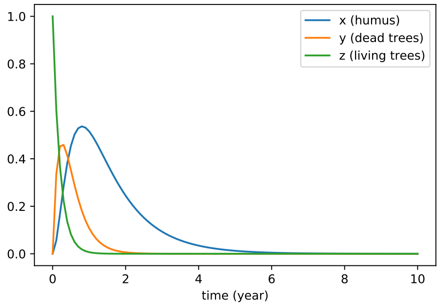
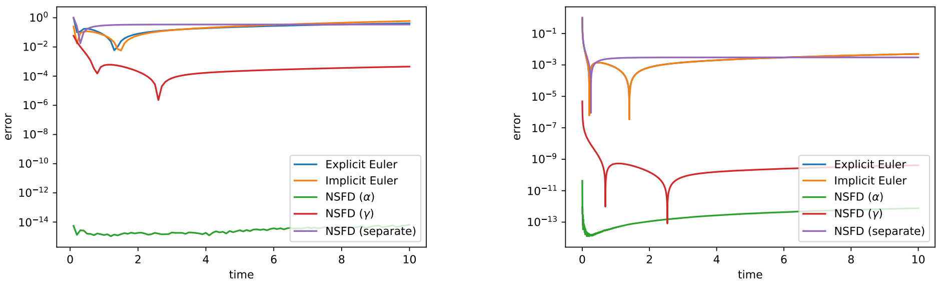


Figure 5: Time evolution of the exact solution of Equation (20) for $z_0 = 1$.

The biomass decayed into humus (corresponding to x) has the slowest time evolution and the errors accumulated on x are greater than on the other variables. This is why we will show the relative errors computed as (19) on this variable. Figure 6 shows the relative errors for the five different methods.

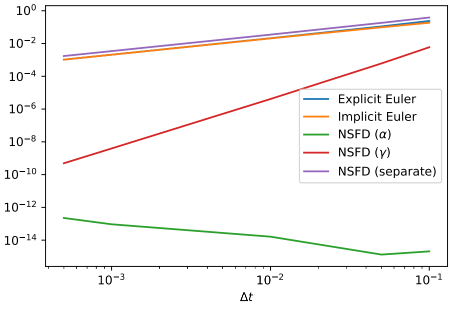


$\Delta t = 0.1$ $\Delta t = 0.001$

Figure 6: Time evolution of the relative errors for $\Delta t = 0.1$ and 0.001 for the forest biomass model.

As expected, our method is exact. The order 3 method also behaves very well. It has the major advantage to be derived only with the knowledge of the coefficient of the characteristic polynomial of matrix A which is much easier to compute than the α_j. The performance of the traditional NSFD method (23) is comparable to that of the explicit and implicit Euler methods.

In Figure 7, we show the convergence of the five methods, plotting $\max_k E_k$ with respect to Δt and the corresponding numerical orders. The NSFD method with the R_1 correction and exact (α) coefficients is exact and the computation of the order is not applicable (n.a.). When approximated with γ coefficients, it is of order 3, which is already a great enhancement compared to the other methods (even the NSFD method that treats equations separately) which are of order one.



Explicit Euler	1.02
Implicit Euler	0.98
NSFD (α)	n.a.
NSFD (γ)	3.07
NSFD (separate)	1.02

Figure 7: Numerical order of the five methods for the biomass model.

4.3 Impact of R_0 and R_1: A forest biomass model with constant force

To test the impact of both R_0 and R_1, we consider the forest biomass model (20), in which we introduce a constant forcing by planting trees. This corresponds to add a constant z_f in the right-hand side of the last equation, modeling the time evolution of living trees. Hence the system reads

$$\begin{cases} x'(t) = -x(t) + 3y(t), \\ y'(t) = -3y(t) + 5z(t), \\ z'(t) = -5z(t) + z_f, \end{cases} \tag{24}$$

with initial conditions $x(0) = 0$, $y(0) = 0$, and $z(0) = z_0$.

The analytical solution is given by

$$\begin{cases} x(t) = \frac{15}{8}\left(e^{-t} - 2e^{-3t} + e^{-5t}\right) z_0 + \frac{1}{8}\left(8 - 15e^{-t} + 10e^{-3t} - 3e^{-5t}\right) z_f, \\ y(t) = \frac{5}{2}\left(e^{-3t} - e^{-5t}\right) z_0 + \frac{1}{6}\left(2 - 5e^{-3t} + 3e^{-5t}\right) z_f, \\ z(t) = e^{-5t}(z_0 - \frac{1}{5}z_f) + \frac{1}{5}z_f. \end{cases}$$

We display in Figure 8 the time evolution of this analytical solution for $z_0 = 1$ and $z_f = 0.5$. We observe in particular the theoretical long time limits, z_f, $z_f/3$, and $z_f/5$ for x, y, and z respectively.

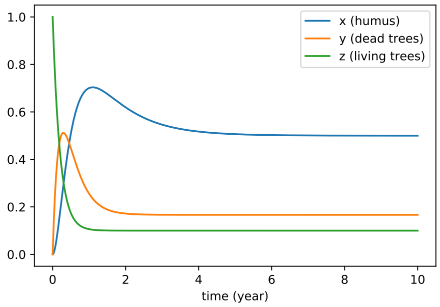


Figure 8: Time evolution of the exact solution of System (24) for $z_0 = 1$ and $z_f = 0.5$.

The NSFD scheme (8) reads

$$\begin{aligned} X_{k+1} = \alpha_0(\Delta t)X_k + \alpha_1(\Delta t)\left[AX_k + \mathcal{B}\right] + \alpha_2(\Delta t)A\left[AX_k + \mathcal{B}\right] \\ + (\alpha_0(\Delta t) - 1)A^{-1}\mathcal{B}, \end{aligned} \tag{25}$$

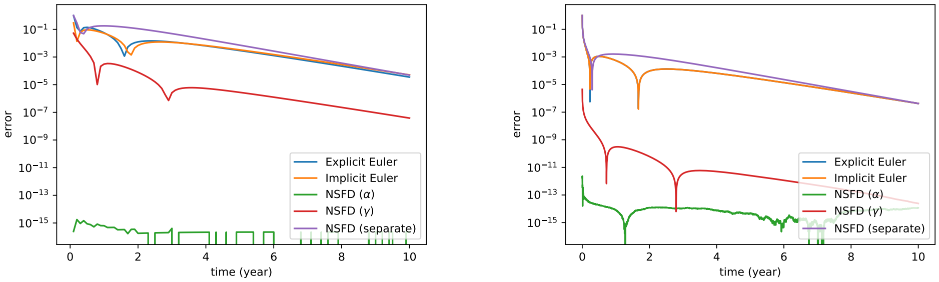
where the matrix A and coefficients α_j are the same as in (22), but now we have a constant nonlinearity $\mathcal{B}$

$$\mathcal{B} = \begin{pmatrix} 0 \\ 0 \\ z_f \end{pmatrix}. \tag{26}$$

This test case enables to study the impact of the correction term for the nonlinear part without any approximation on the nonlinearity itself.

We anew compare this method, with the explicit and implicit Euler schemes, the scheme where the coefficients α_j are replaced by the corresponding γ_j, and the traditional NSFD scheme where we replace the last equation in System (23) by

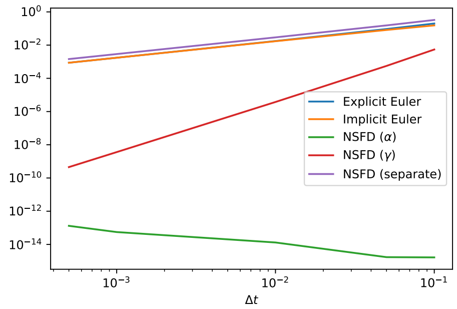
$$\frac{z_{k+1} - z_k}{(1 - e^{-5\Delta t})/5} = -5z_k + z_f.$$



$\Delta t = 0.1$ $\Delta t = 0.001$

Figure 9: Time evolution of the relative errors for $\Delta t = 0.1$ and 0.001 for the forest biomass model with constant forcing.

The time evolution of the relative error between the analytical exact solution and the approximated solutions is shown in Figure 9 for different values of the time step. Our method is exact and behaves very well for any time step. Replacing the α_j's by their third order approximation also yields good results, while the traditional Mickens' NSFD scheme is comparable to the explicit and implicit Euler methods. We illustrate this also on Figure 10. The numerical order at 10^{-2} precision are the same as in the previous test case.



Explicit Euler	1.02
Implicit Euler	0.98
NSFD (α)	n.a.
NSFD (γ)	3.07
NSFD (separate)	1.02

Figure 10: Numerical order of the five methods for the biomass model with constant forcing.

4.4 Non autonomous systems: A forest biomass model with a seasonal plantation

To continue to explore method errors, we now modify the biomass model (20) to have both corrections R_0 and R_1 and this time a nonlinearity that models seasonal plantations, and which amounts to performing a sinusoidal forcing

$$\begin{cases} x'(t) = -x(t) + 3y(t), \\ y'(t) = -3y(t) + 5z(t), \\ z'(t) = -5z(t) + z_f\left[1 + \cos(\omega t)\right], \end{cases} \tag{27}$$

with initial conditions $x(0) = 0$, $y(0) = 0$, and $z(0) = z_0$. Such a time dependent forcing will have to be approximated in the numerical schemes.

This example does not match the description of Section 3 and, to begin with, Equation (2) because it is non autonomous. To include this type of equations we should define $\mathcal{B}(t, X_k, X_{k+1}, \Delta t)$, and the rest of the discussion would be still valid.

The exact analytical solution of the new system is

$$\left\{\begin{array}{rl} x(t) = & \dfrac{15}{8}\left(e^{-t} - 2e^{-3t} + e^{-5t}\right) z_0 + \dfrac{1}{8}\left(8 - 15e^{-t} + 10e^{-3t} - 3e^{-5t}\right) z_f \\ & +15\dfrac{3(5-3\omega^2)\cos(\omega t) + \omega(23-\omega^2)\sin(\omega t)}{(1+\omega^2)(9+\omega^2)(25+\omega^2)} z_f \\ & +\dfrac{15}{8}\left(\dfrac{-e^{-t}}{1+\omega^2} + \dfrac{6e^{-3t}}{9+\omega^2} + \dfrac{-5e^{-5t}}{25+\omega^2}\right) z_f, \\ y(t) = & \dfrac{5}{2}\left(e^{-3t} - e^{-5t}\right) z_0 + \dfrac{1}{6}\left(2 - 5e^{-3t} + 3e^{-5t}\right) z_f \\ & +5\dfrac{(15-\omega^2)\cos(\omega t) + 8\omega\sin(\omega t)}{(9+\omega^2)(25+\omega^2)} z_f + \dfrac{5}{2}\left(\dfrac{-3e^{-3t}}{9+\omega^2} + \dfrac{5e^{-5t}}{25+\omega^2}\right) z_f, \\ z(t) = & e^{-5t} z_0 + \dfrac{1}{5}\left(1 - e^{-5t}\right) z_f + \dfrac{5\cos(\omega t) + \omega\sin(\omega t)}{25+\omega^2} z_f + \dfrac{-5e^{-5t}}{25+\omega^2} z_f. \end{array}\right.$$

This solution is displayed in Figure 11. We choose $z_0 = 1$ and $z_f = 0.5$ to have the same mean limits as in the previous simulations. We also choose $\omega = 2\pi$ to have a one year period for the forcing.

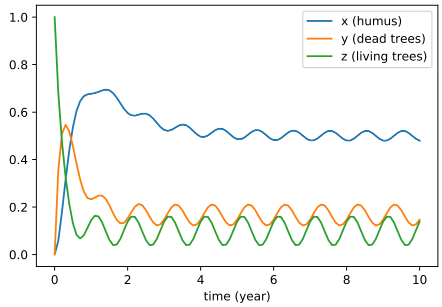


Figure 11: Time evolution of the exact solution of System (27) for $z_0 = 1$, $z_f = 0.5$ and $\omega = 2\pi$.

The numerical schemes we compare are exactly the same as before, except for the treatment of B:

$$B(t) = \begin{pmatrix} 0 \\ 0 \\ z_f\left[1 + \cos(\omega t)\right] \end{pmatrix},$$

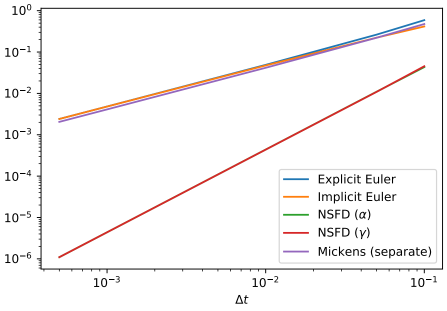
which is now time-dependent and for which we have to choose an approximation. For the computation of X_{k+1} from X_k, five approximations have been used and compared if relevant, namely

$$\mathcal{B}_{\text{left}} = B(t_k),\ \mathcal{B}_{\text{right}} = B(t_{k+1}),\ \mathcal{B}_{\text{middle}} = B((t_k + t_{k+1})/2),$$

$$\mathcal{B}_{\text{half}} = (B(t_k) + B(t_{k+1}))/2,\ \mathcal{B}_{\text{mean}} = \int_{t_k}^{t_{k+1}} B(t)dt.$$

The explicit and implicit Euler methods clearly use $\mathcal{B}_{\text{left}}$ and $\mathcal{B}_{\text{right}}$ respectively, but the question is open for the other numerical methods.

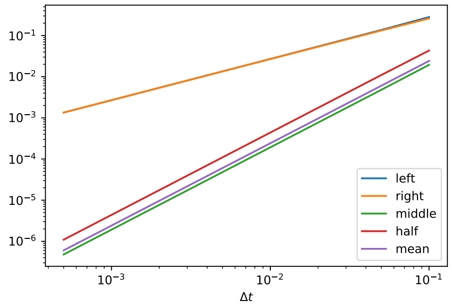
In a first row of numerical tests we compare the errors when $\mathcal{B}$ is approximated by $\mathcal{B}_{\text{half}}$. We choose this because it is the form which (besides the explicit one) is the easiest to extend when nonlinearities involving X_k are concerned. Figure 12 shows the errors for the five studied schemes. Again our method and its third order approximation outperform the three other schemes and yield second order schemes, which corresponds to the error in approximating the nonlinearity.



Explicit Euler	1.03
Implicit Euler	0.98
NSFD (α)	2.00
NSFD (γ)	2.00
NSFD (separate)	1.02

Figure 12: Numerical order of the various method for $\mathcal{B} = \mathcal{B}_{\text{half}}$.

We can discuss a little further by comparing the use of $\mathcal{B}_{\text{left}}$, $\mathcal{B}_{\text{right}}$, $\mathcal{B}_{\text{middle}}$, $\mathcal{B}_{\text{half}}$, and $\mathcal{B}_{\text{mean}}$ for $\Delta t = 0.001$. The numerical results are displayed in Figure 13 for the NSFD scheme with α coefficients. The computation with $\mathcal{B}_{\text{half}}$ yields the worst results among the other methods but the difference is not significant enough to be worth when dealing with more complex nonlinearities or time dependent forcings.



left	1.01
right	0.99
middle	2.00
half	2.00
mean	2.00

Figure 13: Numerical order of the NSFD method for various approximations of B.

All the computations in this numerical test section have been performed using Python notebooks, that can be found at the following address: https://gricad-gitlab.univ-grenoble-alpes.fr/bidegarb/nsfd_systems.

5 Discussion

5.1 Extended rules for systems

We have defined two new rules for NSFD schemes for systems of ODEs. These rules stem from a careful derivation when splitting the equation into a linear and a nonlinear part. The only approximations are made on the nonlinear part.

In a first step a matrix formulation is given, leading to a generalization of the second rule (Rule 2'), which addresses the treatment of the first derivative. The usual scalar functions ϕ and ψ, involved in the denominator and the numerator respectively, are then replaced by matrix valued functions. The system is treated as a whole, contrarily to what can usually been done where each equation is taken into account more or less separately. An example of this separate treatment is illustrated by (23).

The matrix formulation is an exponential integrator, and deriving a scalar version of this scheme allows to avoid the possible difficulties in computing the matrix exponentials. This leads to usual scalar coefficients in the discretization of the first order derivative, but they are the same for all the equations, and to correction terms in the right-hand side, which are described by Rule 3'.

In the examples we have separated the effect of the two correction terms on purpose. But of course they are designed also to work together. If the system is linear, or the nonlinearity is a constant forcing term, no approximation is made at any stage of the derivation and the obtained scheme is exact. In the case of a constant forcing term and for at least three coupled equations the two correction terms are nonzero.

5.2 Deriving the scalar coefficients

The derivation of the scalar coefficient is tedious. The examples shown here are quite simple since they deal with very few equations. In our second example, we computed $\exp(\Delta t A)$ formally and wrote equation (6), which led to solve a three-dimensional linear system in the α_j. Computing $\exp(\Delta t A)$ formally needs to know the eigenvalues and eigenvectors.

Replacing this formal derivation by a numerical determination of the α_j, computing $\exp(\Delta t A)$ numerically and solving the resulting systems numerically can destroy the quality of the method. We have experienced ourselves that even not being careful with the computation of the (scalar) exponentials in the construction of the α_j in Section 4.2 leads to destroy the fine equilibrium that leads to the expansions in Proposition 1 and to a not better scheme than the explicit Euler scheme!

If the formal computation is not possible, we strongly recommend to replace the α_j by their n-th order approximation as done in Section 4.2 with γ_j coefficients. This approximation has the advantage to only use the knowledge of the coefficients of the characteristic polynomial. This polynomial is easier to compute than the α_j. It is indeed the first step in the computation of the α_j. Taking γ_j simply consists in using the truncated series $S_{n-1}(\Delta t A) = \sum_{j=0}^{n-1} \frac{\Delta t^j}{j!} A^j$ instead of the matrix exponential. For a linear system with $n = 5$, this is equivalent to use the classical order 4 Runge–Kutta method. For other system dimensions, we also have a Runge–Kutta-like method, but with an order that is adapted to n.

5.3 Singular linear part

In the previous discussion, we have used A^{-1} and implicitly have supposed that A was non-singular. If A is singular, the nonlinearity B can be written as $B = AC + K$ where K belongs to the kernel of A. Then

$$\int_0^{\Delta t} e^{(\Delta t - s)A} ds B = (e^{\Delta t A} - I)C = (e^{\Delta t A} - I)A^+ B,$$

where A^+ is the generalized inverse of A. This allows to generalize our approach in the singular case.

6 Conclusion

Having considered the NSFD method as a special class of exponential integrators, we have been able to revisit Mickens' rules to apply them to systems of ODEs. When these systems are linear, the method is exact. In the Hamiltonian nonlinear case, it consists in adding to Mickens' schemes a correction term, that has been shown to improve the accuracy.

Acknowledgements

The authors wish to thank Prof J. M.-S. Lubuma for his enlightening comments on this work.

References

[AL01] R. Anguelov and J. M.-S. Lubuma, *Contributions to the Mathematics of the Nonstandard Finite Difference Method and Applications.* Numerical Methods for Partial Differential Equations, **17**(5), 518–543 (2001).

[Cie11] J.L. Cieśliński, *On the exact discretization of the classical harmonic oscillator equation.* Journal of Difference Equations and Applications, **17**(11), 1673–1694 (2011).

[Cie13] J.L. Cieśliński, *Locally exact modifications of numerical schemes.* Computers & Mathematics with Applications, **65**(12), 1920–1938 (2013).

[CR10] J.L. Cieśliński and B. Ratkiewicz, *Improving the accuracy of the discrete gradient method in the one-dimensional case.* Physical Review E, **81**(1), 016704:1–6 (2010).

[CR11] J.L. Cieśliński and B. Ratkiewicz, *Energy-preserving numerical schemes of high accuracy for one-dimensional Hamiltonian systems.* Journal of Physics A: Mathematical and Theoretical, **44**(15), 155206:1–14 (2011).

[GW98] Grant B. Gustafson and Calvin H. Wilcox, *Analytical and computational methods of advanced engineering mathematics*, Springer (1998).

[HL99] M. Hochbruck and C. Lubich, *A Gautschi-type method for oscillatory second-order differential equations.* Numerische Mathematik, **83** 403–426 (1999).

[HO10] M. Hochbruck and A. Ostermann, *Exponential integrators.* Acta Numerica, **19**, 209–286 (2010).

[Hu06] H. Hu, *Exact solution of a quadratic nonlinear oscillator.* Journal of Sound and Vibration **295**, 450–457 (2006).

[Mic94] R. E. Mickens, *Nonstandard finite difference models of differential equations.* World scientific (1994).

[Mic00] R. E. Mickens, *Nonstandard finite difference schemes.* In R. E. Mickens (ed), Applications of Nonstandard Finite Difference Schemes, World scientific, pp. 1–54 (2000).

[MOR05] R. E. Mickens, K. Oyedeji, and S. Rucker, *Exact finite difference scheme for second-order, linear ODEs having constant coefficients.* Journal of Sound and Vibration, **287**(4–5), 1052–1056 (2005).

[MR94] R. E. Mickens and I. Ramadhani, *Finite-difference schemes having correct linear stability properties for all step-sizes III.* Computers & Mathematics with Applications, **27**(4), 77–84 (1994).

[MV03] C. Moler and C. Van Loan, *Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty–Five Years Later.* SIAM Review, **45**(1), 3–49 (2003).

[Pat16] K. C. Patidar, *Nonstandard finite difference methods: recent trends and further developments.* Journal of Difference Equations and Applications, **22**(6), 817–849 (2016).

[Pot82a] R. B. Potts, *Best difference equation approximation to Duffing's equation.* The ANZIAM Journal, **23**(4), 349–356 (1982).

[Pot82b] R. B. Potts, *Differential and difference equations.* The Americal Mathematical Monthly, **89**(6), 402–407 (1982).

[QT18] D. Quang A and H. Manh Tuan, *Exact finite difference schemes for three-dimensional linear systems with constant coefficient.* Vietnam Journal of Mathematics, **46**, 471–492 (2018).

[RM13] L.-I. W. Roeger and R. E. Mickens, *Exact finite difference scheme for linear differential equation with constant coefficients.* Journal of Difference Equations and Applications, **19**(10), 1663–1670 (2013).

[Roe08] L.-I. W. Roeger, *Exact finite-difference schemes for two-dimensional linear systems with constant coefficients.* Journal of Computational and Applied Mathematics, **219**(1), 102–109 (2008).

[SBF18] M. E. Songolo and B. Bidégaray-Fesquet, *Nonstandard finite-difference schemes for the two-level Bloch model.* International Journal of Modeling, Simulation and Scientific Computing, **9**(4), 1850033:1-23 (2018).

[SBF21] M. E. Songolo and B. Bidégaray-Fesquet, *Strang splitting schemes for N-level Bloch models.* To appear in International Journal of Modeling, Simulation, and Scientific Computing.