Advances in Continuous and Discrete Models

RESEARCH Open Access

Forward invariant set preservation in discrete dynamical systems and numerical schemes for ODEs: application in biosciences

Roumen Anguelov[1,2*] and Jean M.-S. Lubuma[3]

*Correspondence:
roumen.anguelov@up.ac.za
[1]Department of Mathematics and Applied Mathematics, University of Pretoria, Pretoria, South Africa
[2]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
Full list of author information is available at the end of the article

Abstract

We present two results on the analysis of discrete dynamical systems and finite difference discretizations of continuous dynamical systems, which preserve their dynamics and essential properties. The first result provides a sufficient condition for forward invariance of a set under discrete dynamical systems of specific type, namely time-reversible ones. The condition involves only the boundary of the set. It is a discrete analog of the widely used tangent condition for continuous systems (*viz.* the vector field points either inwards or is tangent to the boundary of the set). The second result is nonstandard finite difference (NSFD) scheme for dynamical systems defined by systems of ordinary differential equations. The NSFD scheme preserves the hyperbolic equilibria of the continuous system as well as their stability. Further, the scheme is time reversible and, through the first result, inherits from the continuous model the forward invariance of the domain. We show that the scheme is of second order, thereby solving a pending problem on the construction of higher-order nonstandard schemes without spurious solutions. It is shown that the new scheme applies directly for mass action-based models of biological and chemical processes. The application of these results, including some numerical simulations for invariant sets, is exemplified on a general Susceptible-Infective-Recovered/Removed (SIR)-type epidemiological model, which may have arbitrary large number of infective or recovered/removed compartments.

Mathematics Subject Classification: 34C45; 37C79; 65L99; 92D30; 92E99

Keywords: Tangent condition; Invariant sets; Time reversible schemes; Mass action principle; Epidemiological model; SIR; Finite difference method

1 Introduction

Dynamical systems are pervasive in the modelling of naturally occurring phenomena [25]. While there are many results of the theory of continuous dynamical systems that permit to model the features of real-life processes, capturing them with discrete dynamical systems can be challenging. One typical and fundamental example is the forward invariance of a closed set D contained in open set $\Omega \subset \mathbb{R}^n$ with respect to the following system of

autonomous ordinary differential equations:

$$\frac{dx}{dt} \equiv \dot{x} = f(x) \quad \text{in } \Omega, \tag{1}$$

in which the independent variable is the time $t \geq 0$ and the function f is smooth enough. For convenience, we recall that the said forward invariant property means that

$$\mathcal{S}(t)x_0 \in D \quad \text{at all time } t \in [0,\infty) \text{ and for any } x_0 \in D, \tag{2}$$

where $\mathcal{S}(t)$ denotes the solution operator and we assume that there exists a unique global solution $S(t)x_0$ for any initial condition $x(0) = x_0 \in D$.

A short account on the invariance problem is given in [27]. The basic hypothesis for the invariance property (2) is a tangent condition that roughly states that at a point x on the boundary ∂D of D, i.e., $x \in \partial D$, the vector $f(x)$ is either tangent to D or points into the interior, $\overset{\circ}{D}$, of D. In 1942, Nagumo [21] formulated the tangent condition in the following form, using the distance $d(x, D)$ from a point x to the set D:

$$\lim_{h\to 0+} \frac{1}{h} d\big(x + hf(x), D\big) = 0 \quad \text{for } x \in \text{closure}(D). \tag{3}$$

It took nearly three decades for the invariance problem to be revitalized, starting with the seminal works of Brezis [8], who used the definition (3), and that of Bony [7], who came up with the following subtler formulation of the tangent condition based on the inner product, $\langle x, y\rangle$, in $\mathbb{R}^n$:

$$\big\langle \nu(x), f(x)\big\rangle \leq 0 \quad \text{for } x \in \partial D, \tag{4}$$

where $\nu(x)$ is any outer normal to D at x (in the sense of Bony). Let us recall that a vector $\nu(x) \neq 0$ is called outer normal vector to D at $x \in \partial D$ if the open ball with center $x + \nu(x)$ and radius $|\nu(x)|$ has no intersection with $\overset{\circ}{D}$. The straight line through x perpendicular to $\nu(x)$ is called a tangent. This general definition of normal vector and tangent does not require any smoothness of the boundary ∂D, which is rather convenient in applications. Let us note that the normal vector $\nu(x)$, similarly the associated tangent, need not exist and, if existing, need not be unique. However, $\nu(x)$ at the point $x \in \partial D$ reduces to the classical outer normal if the boundary is smooth, so the tangent exists in the classical sense at x.

In the late sixties, there was a great deal of solving the invariance problem (and /or related issues of existence of solution of the initial value problem) under either formulation of the tangent condition (3) or (4), as seen from the contributions in [10, 13, 23], where the works of Bony and Brezis [7, 8] were revisited, compared, and extended. Since then, the tangent condition has abundantly been used, particularly in mathematical biology, to show the forward invariant structure of the positive cone $\mathbb{R}^n_+$ with respect to ordinary differential equation models and even with respect to reaction diffusion models (see for instance [24, 28]). However, to the authors' best knowledge, the use of the tangent condition to establish the forward invariant nature of sets under discrete dynamical systems has not been explored.

The purpose of this work is primarily to fill this gap. To this end we consider a discrete dynamical system

$$x_{k+1} = F(h, x_k) \equiv F(h)(x_k), \tag{5}$$

where in the map, $F \equiv F(h,x) \equiv F(h)(x) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, the parameter $h > 0$ represents the time step size between two consecutive states x_k and x_{k+1}. This map is assumed to be as smooth as needed in the two arguments. The specific class of discrete dynamical systems we will focus on is defined as follows [12]:

Definition 1 A discrete dynamical system (5) is called symmetric or (time) reversible if

$$x_k = F(-h, x_{k+1}). \tag{6}$$

More generally, the discrete dynamical system is said to be reversible for all x whenever $y = F(h,x)$ implies that $x = F(-h,y)$, which means that $F(-h,\cdot)$ is the inverse $F(h,\cdot)$.

We introduce a discrete analog of the tangent condition that in essence states the following: *there exists* $\overline{h} > 0$ *such that for all* $x \in \partial D$ *and* $h \in (0, \overline{h}]$, *the vector* $F(-h,x) - x$ *does not point inside* $\overset{\circ}{D}$. Under this condition, we show that the set D is invariant under the map $F(h,.)$ for every h in $(0, \overline{h}]$.

A natural question of interest is to link the above-stated discrete tangent condition to continuous dynamical systems. This brings us to the second purpose of this paper, which we achieve by considering the sequence (x_k) obtained recursively through Eq. (5) as approximations at the discrete times $t_k = hk$, $h > 0$, $k = 0, 1, 2, \ldots$ of the solutions $x(t) = S(t)x_0$ of the continuous system (1). In this case, it is essential to assume that

$$F \text{ is at least continuous on } x \text{ and continuously differentiable on } h, \tag{7}$$

and that it satisfies the consistency requirements,

$$F(0,x) = x, \qquad \frac{dF(0,x)}{dh} = f(x). \tag{8}$$

Our specific aim is that the numerical method (5) is reliable in the three important directions below. The numerical method must satisfy the discrete tangent condition whenever the tangent condition holds for the continuous system, thereby showing that the scheme preserves the forward invariant structure. Furthermore, the numerical method must be elementarily stable and second-order convergent.

Regarding the latter targeted property, we stress that the construction of higher-order nonstandard finite difference (NSFD) schemes that are dynamically consistent with the underlying features of the continuous differential equations, particularly for those with transient dynamics, is a pending problem. Several authors have attempted to address this problem. These include the nonstandard versions of the classical θ-method, with $\theta = 1/2$, developed in [1–3, 16], where the positivity of the discrete solution is achieved by also using Mickens' rule on a nontrivial denominator function for the discrete derivative [18, 19]. In the same vein, we mention the recent work [15], valid for a scalar differential equation,

where the second-order accuracy and elementary stability are achieved by a modified non-standard θ method, $0 \leq \theta \leq 1$, in which the nontrivial denominator function varies at each iteration.

The rest of the paper is organized as follows. In the next section, we state precisely the discrete tangent condition in two equivalent forms and establish the forward invariance property under time-reversible flows. We devote Sect. 3 to the construction of a NSFD scheme that is time-reversible, and show that, in essence, this property makes our scheme of second-order accuracy and elementarily stable. In the same section, it is further shown via the discrete tangent condition that the NSFD scheme preserves the forward invariance structure of the continuous dynamical system. The case of mass action models of biological and chemical processes is considered in detail in Sect. 4, followed by Sect. 5 on a general SIR-type epidemiological model with arbitrary number of infective or recoverred/removed compartments. Concluding remarks with possible extensions of this work are given in Sect. 6.

2 Domain preserving property of reversible schemes

The theory for continuous dynamical systems of the form (1) provides theorems proving that a set in $\mathbb{R}^n$ is forward invariant through conditions only on the boundary of the set, namely the tangent condition as stated in (3) or (4), see for instance [27]. In this section we derive discrete counterparts of the tangent condition and resulting theorems for the forward invariance of sets under reversible maps.

Let the closed set D be a subset of Ω and be related to its interior, $\overset{\circ}{D}$, and boundary, ∂D, as follows:

$$D = \text{closure}(\overset{\circ}{D}) = \overset{\circ}{D} \cup \partial D. \tag{9}$$

Theorem 2 *Let the discrete dynamical system* (5) *be reversible for all* $h \in (0, \overline{h}]$ *and let* (7)–(8) *hold. Then the set D is invariant under* $F(h, \cdot)$ *for every* $h \in (0, \overline{h})$ *if and only if,*

$$F(-h, \partial D) \cap \overset{\circ}{D} = \emptyset, \quad h \in (0, \overline{h}]. \tag{10}$$

Proof To show that Condition (10) is sufficient for the invariance of D, let us consider $x \in \overset{\circ}{D}$. Assume that there exists $h \in (0, \overline{h}]$ such that $F(h, x) \notin \overset{\circ}{D}$. Since $F(h, x)$ is continuous on h, there exists $\hat{h} \in (0, h]$ such that $y = F(\hat{h}, x) \in \partial D$. Equivalently, $x = F(-\hat{h}, y)$, which contradicts the Condition (10). Therefore, $F(h, x) \in \overset{\circ}{D}$. Considering that x is an arbitrary element of $\overset{\circ}{D}$, we have $F(h, \overset{\circ}{D}) \subseteq \overset{\circ}{D}$. Using (9) and the continuity of F on x, we obtain $F(h, D) \subseteq D$, $h \in (0, \overline{h}]$.

Conversely, let D be invariant under $F(h, \cdot)$ for every $h \in (0, \overline{h}]$, i.e.,

$$F(h, D) \subseteq D \quad \text{for every } h \in (0, \overline{h}], \tag{11}$$

and assume that Condition (10) is violated, that is, there exists $h \in (0, \overline{h}]$ and $x \in F(-h, \partial D) \cap \overset{\circ}{D}$. Then, using the fact that $F(-h, \cdot)$ is the inverse of $F(h, \cdot)$ we have

$$y = F(h, x) \in F\big[h, F(-h, \partial D)\big] \cap F(h, \overset{\circ}{D}) = \partial D \cap F(h, \overset{\circ}{D}). \tag{12}$$

From the continuity of $F(-h,\cdot)$, it follows that the set $F(h,\overset{\circ}{D})$ is open. Therefore, there exists $\varepsilon > 0$ such that the open ball $V_\varepsilon y$ with radius ε and centered at y satisfies $V_\varepsilon(y) \subseteq F(h,\overset{\circ}{D})$. Since y is on the boundary of D, the open ball $V_\varepsilon(y)$ contains points which are not in D. More precisely, there exists $z \notin D$ such that

$$z \in V_\varepsilon(y) \subseteq F(h,\dot{D}).$$

This is a contradiction to (11), which proves that Condition (10) is necessary for the stated invariance of D. □

Condition (10) is not easy to verify. If the set D is convex, we replace below Condition (10) by an equivalent condition, requiring that the vector $F(-h,x) - x$ be outward directed or tangential to D at $x \in \partial D$. The latter condition is our proposed discrete analog of Bony [7] tangent condition for continuous dynamical systems, based on Bony's definition of outer normal vector $\nu(x)$, stated in (4).

Let D be convex. Then some comments on outer normal vectors are in order. First, let us recall that a supporting hyperplane to D at a point $x \in \partial D$ is a hyperplane containing the point x, while the set D is contained in one of the closed halfspaces determined by the hyperplane. The so-called *supporting hyperplane theorem* [17] states that for D, a closed convex set with nonempty interior, there exists a supporting hyperplane at every $x \in \partial D$. It is easy to see that the vector $\nu(x)$, which is normal to the supporting hyperplane at x and points in the halfspace not containing D, is an outer normal vector to D at x in Bony's sense given in the introduction. Under the stated assumptions, it follows from the hyperplane separation theorem for convex sets [17] that one can associate a supporting hyperplane with every outer normal vector, a fact that is intuitively clear. To fix the notation, the supporting hyperplane at $x \in \partial D$ with an outer normal vector $\nu(x)$ is the set

$$\mathcal{H}^0_{\nu(x)} := \left\{ y \in \mathbb{R}^n : \langle y - x, \nu(x) \rangle = 0 \right\},$$

which is the common boundary of the following two closed halfspaces:

$$\mathcal{H}^+_{\nu(x)} = \left\{ y \in \mathbb{R}^n : \langle y - x, \nu(x) \rangle \geq 0 \right\},$$
$$\mathcal{H}^-_{\nu(x)} = \left\{ y \in \mathbb{R}^n : \langle y - x, \nu(x) \rangle \leq 0 \right\}.$$

Then, the fact that $\mathcal{H}^0_{\nu(x)}$ is a supporting hyperplane of D at $x \in \partial D$ is expressed by the inclusion

$$D \subseteq \mathcal{H}^-_{\nu(x)},$$

or, equivalently,

$$\overset{\circ}{D} \subseteq \overset{\circ}{\mathcal{H}}{}^-_{\nu(x)}. \tag{13}$$

Theorem 3 *Let a discrete dynamical system (5) be reversible for all $h \in (0,\overline{h}]$ and let D be a closed and convex subset of Ω satisfying (9). Then the set D is invariant under $F(h,\cdot)$ for*

every $h \in (0, \overline{h}]$ *if, and only if, for every* $h \in (0, \overline{h}]$ *and* $x \in \partial D$*, there exists* $y \in \partial D$ *and an outer normal vector* $\nu(y)$ *such that*

$$\langle \nu(y), F(-h,x) - y \rangle \geq 0. \tag{14}$$

Proof We show that the discrete tangent condition (14) implies Condition (10). Indeed, let $h \in (0, \overline{h}]$ and $x \in \partial D$. Using the convexity of D, the inequality (14) shows that

$$F(-h,x) \in \mathcal{H}^+_{\nu(y)}. \tag{15}$$

Using (13), we obtain

$$\overset{\circ}{D} \cap \mathcal{H}^+_{\nu(y)} \subseteq \overset{\circ}{\mathcal{H}^-_{\nu(y)}} \cap \mathcal{H}^+_{\nu(y)} = \emptyset. \tag{16}$$

Then (15) and (16) imply that $F(-h,x) \notin \overset{\circ}{D}$. Considering that $h \in (0, \overline{h}]$ and $x \in \partial D$ are arbitrary, Condition (10) follows.

To show the converse implication (10) $\Longrightarrow$ (14), let $h \in (0, \overline{h}]$ and $x \in \partial D$. Condition (10) shows that $F(-h,x) \notin \overset{\circ}{D}$. From the hyperplane separation theorem of convex sets it follows that there exists $y \in \partial D$ and an outer normal vector $\nu(y)$ such that $F(-h,x) \in \mathcal{H}^+_{\nu(y)}$, which implies (14). □

The difficulty raised about the practical use of Condition (10) is equally present in its equivalent formulation (14). In applications, it is easier to use the latter condition with $y = x$, i.e., Eq. (17) below. This turns out to be our discrete tangent condition, which, as stated in the next theorem, is a sufficient condition under which the forward invariance property with respect to discrete dynamical systems is achieved.

Theorem 4 *Let a discrete dynamical system* (5) *be reversible for all* $h \in (0, \overline{h}]$ *and let* D *be a closed and convex subset of* Ω *satisfying* (9)*. Then the set* D *is invariant under* $F(h, \cdot)$ *for every* $h \in (0, \overline{h}]$ *if*

$$\langle \nu(x), F(-h,x) - x \rangle \geq 0, \ \ \textit{for } h \in (0, \overline{h}],\ x \in \partial D,\ \ \nu(x) \textit{ - outer normal to } D \textit{ at } x. \tag{17}$$

3 Reversible nonstandard finite difference schemes

In the construction of nonstandard finite difference schemes, Mickens [18, 19] made an important observation, namely, that the discrete models of differential equations have a larger parameter space than the corresponding differential equations. Adding to this Mickens' Rule 1, which says that the orders of the discrete derivatives must be exactly equal to the orders of the corresponding derivatives of the differential equations, it is convenient in the setting of Mickens' nonlocal approximation Rule 3, to view the right-hand side of the system (1) as a restriction, on the diagonal $z = y$, of a certain function of two variables $\varphi \equiv \varphi(y,z)$ such that $f(x) = \varphi(x,x)$ and hence the system takes the form

$$\dot{x} = \varphi(x,x), \quad x \in \Omega \subseteq \mathbb{R}^n. \tag{18}$$

We consider the numerical method,

$$\frac{x_{k+1}-x_k}{h}=\frac{1}{2}\big(\varphi(x_{k+1},x_k)+\varphi(x_k,x_{k+1})\big), \tag{19}$$

which is a NSFD scheme [4], on which only the nonlocal approximation of the right-hand side of (18) is used, the rule on the complex denominator function of the derivatives being excluded. The NSFD scheme (19) is implicit. The existence and uniqueness of solution of the respective system of algebraic equation is an issue which requires attention on its own. To make this NSFD scheme a generalized dynamical system on Ω [25], we will assume here that

$$\begin{aligned}&\text{(i)}\quad \text{given } x_k\in\Omega, \text{ equation (19) has a unique solution } x_{k+1} \text{ in } \Omega;\\ &\text{(ii)}\quad \text{different values of } x_k \text{ result in different solutions for } x_{k+1}.\end{aligned} \tag{20}$$

Finding general conditions for (20) to hold is challenging, as is the case for general equations of the form $G(x,y)=0$ solved by using for instance the implicit function theorem, sophisticated fixed-point iterations, *etc.* However, in particular settings, as the ones that arise in applications and are considered in the sequel, such conditions are relatively easy to formulate.

Theorem 5 *Assume that* (20) *holds. Then the NSFD scheme* (19) *satisfies the following properties*:

a) *It is time-reversible*;
b) *It is elementarily stable in the sense that*
   b1) *Its fixed points are precisely the equilibrium points of the system* (18);
   b2) *It replicates the local stability of all hyperbolic equilibrium points of* (18);
c) *It is a second-order scheme for the Equation* (18).

Proof The Assumption (20) implies that the NSFD scheme (19) can be written in the explicit form (5), where the function F satisfies the equation

$$F(h,x)-x=\frac{h}{2}\big[\varphi\big(F(h,x),x\big)+\varphi\big(x,F(h,x)\big)\big]. \tag{21}$$

a) If in the Equation (19) we interchange x_k and x_{k+1} and replace h by $-h$, the equation remains the same. Therefore, (6) holds, which means that the NSFD scheme is time-reversible.

b1) It is clear that $\overline{x}$ is a fixed point of (19) if, and only if, $\varphi(\overline{x},\overline{x})=0$, *i.e.* $\overline{x}$ is an equilibrium point of (18).

b2) Let us use the explicit form (5) of (19). From (21), the Jacobian matrix (in the x variable) of F satisfies

$$\begin{aligned}\frac{\partial F(h,x)}{\partial x}-I=\frac{h}{2}\Bigg(&\frac{\partial\varphi}{\partial y}\big(F(h,x),x\big)\frac{\partial F(h,x)}{\partial x}+\frac{\partial\varphi}{\partial z}\big(F(h,x),x\big)\\&+\frac{\partial\varphi}{\partial y}\big(x,F(h,x)\big)+\frac{\partial\varphi}{\partial z}\big(x,F(h,x)\big)\frac{\partial F(h,x)}{\partial x}\Bigg),\end{aligned} \tag{22}$$

where I denotes here and after the $n \times n$ identity matrix. Let $\overline{x}$ be an arbitrary fixed point of (19). At $x = \overline{x}$, using $F(h,\overline{x}) = \overline{x}$, Equation (22) simplifies to

$$\left(I - \frac{h}{2}\left(\frac{\partial \varphi}{\partial y}(\overline{x},\overline{x}) + \frac{\partial \varphi}{\partial z}(\overline{x},\overline{x})\right)\right)\frac{\partial F(h,\overline{x})}{\partial x} = I + \frac{h}{2}\left(\frac{\partial \varphi}{\partial y}(\overline{x},\overline{x}) + \frac{\partial \varphi}{\partial z}(\overline{x},\overline{x})\right),$$

or, equivalently,

$$\left(I - \frac{h}{2}\frac{\partial f(\overline{x})}{\partial x}\right)\frac{\partial F(h,\overline{x})}{\partial x} = I + \frac{h}{2}\frac{\partial f(\overline{x})}{\partial x}. \tag{23}$$

Let $\lambda \equiv \Re(\lambda) + \imath\Im(\lambda)$ be an eigenvalue of $\frac{\partial f(\overline{x})}{\partial x}$ with associated left eigenvector v. Then multiplying both sides of (23) on the left by v, we obtain

$$\left(1 - \frac{h}{2}\lambda\right)v\frac{\partial F(h,\overline{x})}{\partial x} = \left(1 + \frac{h}{2}\lambda\right)v.$$

It follows from (20) that the matrix $\frac{\partial F(h,\overline{x})}{\partial x}$ is not singular. Then, due to the obvious impossibility for $1 - \frac{h}{2}\lambda$ and $1 + \frac{h}{2}\lambda$ to be both zero, neither of them is. Hence

$$v\frac{\partial F(h,\overline{x})}{\partial x} = \frac{1 + \frac{h}{2}\lambda}{1 - \frac{h}{2}\lambda}v.$$

Therefore, to every eigenvalue λ of $\frac{\partial f(\overline{x})}{\partial x}$ with associated left eigenvector v corresponds an eigenvalue $\mu = \frac{1+\frac{h}{2}\lambda}{1-\frac{h}{2}\lambda}$ of the matrix $\frac{\partial F(h,\overline{x})}{\partial x}$ with the same left eigenvector v. After some technical manipulation, we obtain

$$\begin{aligned} |\mu|^2 &= \frac{(1 + \frac{h}{2}\Re(\lambda))^2 + \frac{h^2}{4}\Im^2(\lambda)}{(1 - \frac{h}{2}\Re(\lambda))^2 + \frac{h^2}{4}\Im^2(\lambda)} \\ &= \frac{1 + \frac{h^2}{4}|\lambda|^2 + h\Re(\lambda)}{1 + \frac{h^2}{4}|\lambda|^2 - h\Re(\lambda)}. \end{aligned} \tag{24}$$

It follows from (24) that

$$\Re(\lambda) < 0 \quad \Longleftrightarrow \quad |\mu| < 1,$$

which implies that a hyperbolic equilibrium point $\overline{x}$ of (18) is asymptotically stable if and only if it is asymptotically stable as a fixed-point of (19).

c) Let $x \in \Omega$ and let $u(t) = \mathcal{S}(t)x$ denote the solution of (1) with $u(0) = x$. Denote

$$\xi = \frac{1}{2}\big(F(h)(x) + x\big).$$

Taylor expansion about $h = 0$ yields

$$\xi = \frac{1}{2}\left(x + F(0,x) + h\frac{dF}{dh}(0,x) + O\big(h^2\big)\right),$$

so that

$$\xi = u\left(\frac{h}{2}\right) + O\left(h^2\right), \tag{25}$$

by the consistency conditions (8). Then, for the local truncation error,

$$E(h) := u(h) - F(h,x) = u(h) - x - \frac{h}{2}\varphi\big(F(h)(x),x\big) - \frac{h}{2}\varphi\big(x,F(h)(x)\big),$$

of the NSFD scheme (5) and (21), Taylor expansions of u about $\frac{h}{2}$ and of φ about $(y,z) = (\xi,\xi)$ yield

$$\begin{aligned} E(h) &= h\dot{u}\left(\frac{h}{2}\right) - \frac{h}{2}\left(\varphi(\xi,\xi) + \frac{\partial\varphi(\xi,\xi)}{\partial y}\big(F(h)(x) - \xi\big) + \frac{\partial\varphi(\xi,\xi)}{\partial z}(x-\xi)\right) \\ &\quad - \frac{h}{2}\left(\varphi(\xi,\xi) + \frac{\partial\varphi(\xi,\xi)}{\partial y}(x-\xi) + \frac{\partial\varphi(\xi,\xi)}{\partial z}\big(F(h)(x) - \xi\big)\right) + O\left(h^3\right) \\ &= h\left(f\left(u\left(\frac{h}{2}\right)\right) - f(\xi)\right) + O\left(h^3\right). \end{aligned}$$

Then performing a further Taylor expansion and using (25), we obtain $E(h) = O(h^3)$ from which second-order accuracy follows by the standard theory for one-step numerical methods. □

Remark 6 Since the scheme (19) is symmetric, the statement c) in Theorem 5 can be derived from the general theory of symmetric schemes [12]. We note that properties in items a) and b) are of qualitative nature and do not follow from the standard numerical analysis theory, involving zero-stability, consistency, and therefore convergence.

4 Application to mass action-type models

Modelling of biological and chemical processes by applying the mass action principle for representing the interaction of the involved species typically results in a system of the form (18) in Ω, a convex and compact subset of $\mathbb{R}^n_+$, where the function φ is linear in both its arguments. Then φ can be represented as

$$\varphi(x,y) = P(x)y + A(x+y) + b = Q(y)x + A(x+y) + b, \tag{26}$$

where P and Q are $n \times n$ matrix functions of x and y, respectively. The matrix A and the vector b are constant. In this particular setting, the numerical method (19) can be written in the form

$$\frac{x_{k+1} - x_k}{h} = \frac{1}{2}\big(P(x_k)x_{k+1} + Q(x_k)x_{k+1}\big) + A(x_{k+1} + x_k) + b, \tag{27}$$

or, equivalently,

$$\left(I - \frac{h}{2}\big(P(x_k) + Q(x_k)\big) - hA\right)x_{k+1} = (I + hA)x_k + hb. \tag{28}$$

Furthermore, and also expressing its reversibility, the equation (27) can be written in the form

$$\left(I + \frac{h}{2}\big(P(x_{k+1}) + Q(x_{k+1})\big) + hA\right)x_k = (I - hA)x_{k+1} - hb. \tag{29}$$

Considering that Ω is compact, there exists $\overline{h}$ such that the matrices

$$\begin{aligned} &M(h,x) := I - \frac{h}{2}\big(P(x) + Q(x)\big) - hA \ \&\ M(-h,x) := I + \frac{h}{2}\big(P(x) + Q(x)\big) + hA \\ &\text{are both diagonally dominant matrices for all } h \in (0,\overline{h}] \text{ and } x \in \Omega. \end{aligned} \tag{30}$$

Then, these matrices are non-singular and (28) can be written explicitly as

$$x_{k+1} = F(h, x_k), \tag{31}$$

where

$$F(h,x) = \big(M(h,x)\big)^{-1}\big((I + hA)x + hb\big). \tag{32}$$

Further, we have

$$F^{-1}(h,x) = F(-h,x) = \big(M(-h,x)\big)^{-1}\big((I - hA)x - hb\big). \tag{33}$$

Theorem 3 is a useful tool for determining the invariance of Ω under $F(h,\cdot)$. Using the specific form of the maps (32), we derive an explicit representation of $F(-h,x) - x$. Indeed, using (33) and the expression of $M(h,x)$ in (30), we have

$$\begin{aligned} M(-h,x)\big(F(-h,x) - x\big) &= (I - hA)x - hb - \left(I + \frac{h}{2}\big(P(x) + Q(x)\big) + hA\right)x \\ &= \frac{h}{2}\big(P(x)x + 2Ax\big) + \frac{h}{2}(Q(x)x + 2Ax \\ &= -\frac{h}{2}\varphi(x,x) - \frac{h}{2}\varphi(x,x) \\ &= -hf(x). \end{aligned} \tag{34}$$

Therefore,

$$F(-h,x) - x = -h\big(M(-h,x)\big)^{-1}f(x), \tag{35}$$

and also, by Taylor expansion,

$$F(-h,x) - x = -hf(x) + h^2\int_0^1 \frac{d^2F(-th,x)}{dt^2}(1-t)\,dt.$$

Then, the relation (17) can be written equivalently as

$$\big\langle \nu(x), \big(M(-h,x)\big)^{-1}f(x)\big\rangle \le 0, \quad x \in \partial\Omega, \tag{36}$$

or

$$\langle \nu(x), -hf(x)\rangle + h^2 \int_0^1 \left\langle \nu(x), \frac{d^2F(-th,x)}{dt^2}(1-t)\right\rangle dt \geq 0, \quad x \in \partial\Omega. \tag{37}$$

Assume that the continuous tangent condition (4) holds with strict inequality < 0 for all $x \in \partial\Omega$ and for the dynamical system (1) or (18) in which the function $f(x) = \varphi(x,x)$ is given by (26). To be more precise, we assume that there exists a constant upper bound $c < 0$ in the said strict inequality, i.e., $\leq c < 0$ for all $x \in \partial\Omega$. Then, it follows from (37) that there exists $\overline{h} > 0$ such that the relation (36) holds in the following uniform manner, which is the discrete tangent condition (17):

$$\langle \nu(x), \big(M(-h,x)\big)^{-1} f(x)\rangle \leq 0 \quad \text{for } x \in \partial\Omega, \forall h \in (0,\overline{h}]. \tag{38}$$

On the contrary, if (4) holds in the limit case with the equality '$= 0$', the discrete tangent condition (38) must be checked directly, because the previous argument is no longer true.

Note that the scheme (27) is constructed in the general form (19). It satisfies Assumption (20) with the unique solution being given in the explicit form (31) for $h \in (0,\overline{h}]$. The invariance of Ω follows from Theorem 3 combined with the discrete tangent condition (38), which, as shown above, is inherited from the continuous tangent condition in the strict case and is checked in the limit case. We have thus proved the following theorem:

Theorem 7 *Assume that the tangent condition* (4) *is satisfied for the continuous model* (1), (18), *and* (26). *Then the NSFD scheme* (27) *inherits the forward invariance on* Ω *in the sense explained above. Furthermore, the scheme* (27) *satisfies the properties a)–c) in Theorem* 5.

Remark 8 In limit situations which are not conclusive (e.g., stability of the disease-free equilibrium when the basic reproduction number $\mathcal{R}_0 = 1$ [2, 14]), it is standard in mathematical analysis to deal with such cases differently. This is the essence of the requirement that the discrete tangent condition be checked directly in the limit case when $\langle \nu(x), f(x)\rangle = 0$. As a matter of fact, it often occurs in practical applications that the vector $\nu(x)$ is a left eigenvector of $M(-h,x)$ with a positive eigenvalue λ. If this is the case, then the discrete tangent condition (38) follows directly from the tangent condition (4) of the continuous system. Indeed,

$$\langle \nu(x), \big(M(-h,x)\big)^{-1} f(x)\rangle = \nu(x)^T \big(M(-h,x)\big)^{-1} f(x) = \lambda^{-1}\langle \nu(x), f(x)\rangle \leq 0.$$

Remark 9 The NSFD scheme (19) is an implicit method. When applied to mass action models, and due to the nonlocal approximation of nonlinear terms, it is interesting to observe that the method can be written in the form (28). Therefore, every step requires only solving a linear system—a computational effort comparable to the explicit methods.

5 Example: a general epidemiological model

In this section, we show how the theory and tools developed in Sects. 2, 3, and 4 can be put together in deriving second-order scheme with the properties a)–b) in Theorem 5. While the theoretical existence of $\overline{h}$ is important, deriving constructively the value of $\overline{h}$ is critical for any specific application. We show how the value of $\overline{h}$ can be computed for the

considered model. Although the numerical approach proposed in this paper is exemplified on the model considered in this section, its realm of application is wider.

5.1 The model

We consider a general compartmental model of a contagious disease in a single population. Let the model have n compartments with their sizes given by the coordinates of a vector $x = (x^1, x^2, \dots, x^n)^T \in \mathbb{R}^n$, where we fix $x^1 = S$—the compartment of susceptible individuals. We assume that the infection rate is given via mass-action term of the form

$$\sum_{j=2}^{n} \beta_j S x^j.$$

If the jth compartment contains infective individuals, then $\beta_j > 0$. If the individuals in the jth compartment are not infective, then $\beta_j = 0$. Examples of infective compartments are: symptomatic infective, asymptomatic infective, recovered, but still infective, partially isolated infective, etc. Examples of not infective, other than the susceptible, are: exposed (latent period), recovered with temporary immunity, recovered with permanent immunity, fully isolated (receiving treatment or not), dead, vaccinated (one or more types of vaccines), and others. The newly infected individuals are distributed in the compartments $x^2, \dots, x^n$. Let $g_{ij} S x^j$ be the growth rate of the ith compartment due to the susceptible infected by the individuals in the jth compartment. Clearly, we have $g_{ij} \geq 0$, $i, j = 2, \dots, n$, and

$$\beta_j = g_{2j} + \cdots + g_{nj}. \tag{39}$$

Denote $g_{1j} = -\beta_j$, $j = 2, \dots, n$, and $g_{i1} = 0$, $i = 1, 2, \dots, n$. Then using the matrix $G = (g_{ij})$, the mass action is represented by the vector $x^1 Gx$. The remaining transfer rates between compartments are assumed to be linear and represented by a matrix C, which is a Metzler matrix, that is, the nondiagonal entries of C are nonnegative. Further, a constant recruitment rate is given by a vector $b = (b^1, b^2, \dots, b^n)^T \geq 0$, where b^i is the recruitment rate in the ith compartment. The recruitment in the compartment of susceptible is always positive, that is $b^1 > 0$, but other compartments may also have positive recruitment rates. Then, the compartmental epidemiological model is represented as a system of differential equations

$$\frac{dx}{dt} = f(x) := Gx e_1^T x + Cx + b, \tag{40}$$

where e_1 is the first vector in the canonical basis $\{e_1, e_2, \dots, e_n\}$ in $\mathbb{R}^n$.

There are a large number of models in mathematical epidemiology that can be represented in the generic form (40). These include the SEIR-types of models, [14], as well as models with more strains, [9, 26], models with multiple types of infective and treated individuals (SEITR model), [22], models including vaccination (SVEIR model), [11].

The system (40) is of the form (18) with

$$\varphi(x, y) = Gx e_1^T y + \frac{1}{2} C(x + y) + b,$$

which in turn is of the form (26), where

$$P(x) = Gxe_1^T, \qquad Q(y) = e_1^T yG = y_1 G, \qquad A = \frac{1}{2}C,$$

and vector b as given. Assuming that the compartments take jointly into account the whole population, by adding all equations we obtain that the total population $N = x^1 + x^2 + \cdots + x^n$ satisfies the conservation law,

$$\frac{dN}{dt} = \Lambda - \mu N, \tag{41}$$

where $\Lambda = b^1 + b^2 + \cdots + b^n$ and μ is the disease-independent removal/death rate. It is easy to see that this property implies that $\nu = (1, 1, \ldots, 1)^T$ is a left eigenvector of C and we have $\nu^T C = -\mu\nu^T$. Since C is a Metzler matrix, this shows that all diagonal entries of C are negative. More precisely, $c_{ii} \leq -\mu < 0$. Let us mention that due to the property (39), we have $\nu^T G = (0, 0, \ldots, 0)^T$. Hence, the mass action kinetics do not affect the demography.

It is easy to see, using the tangent condition (4), that the model (40) defines a forward dynamical system on

$$\Omega = \left\{ x \in \mathbb{R}^n_+ : \sum_{i=1}^n x^i \leq \frac{\Lambda}{\mu} \right\}.$$

Indeed, Ω is bounded by the coordinate planes and the plane

$$\sum_{i=1}^n x^i = \frac{\Lambda}{\mu}. \tag{42}$$

The respective outer normal vectors are

$$-e_1, -e_2, \ldots, -e_n, \nu.$$

Using the Metzler property of C, we have

$$\left\langle -e_i, f(x) \right\rangle \Big|_{x^i=0} = -b^i - x^1 \sum_{j \neq i} g_{ij} x^j - \sum_{j \neq i} c_{ij} x^k \leq 0, \quad i = 2, \ldots, n,$$

$$\left\langle -e_1, f(x) \right\rangle \Big|_{x^1=0} = -b^1 - \sum_{j=2}^n c_{1j} x^j < 0.$$

Further, on the plane (42), we have

$$\left\langle \nu, f(x) \right\rangle = \sum_{i=1}^n f_i(x) = \Lambda - \mu \sum_{i=1}^n x_i = 0, \tag{43}$$

which completes the proof of the claim that Ω is forward invariant for (40).

5.2 Discretization and its properties

We consider the numerical method (27) for the model (40) on Ω. The matrices $M(h,x)$ and $M(-h,x)$, see (30), are both strictly diagonally dominant if and only if

$$1-\frac{h}{2}\left(x^1\sum_{j=1}^{n}|g_{ij}|+\sum_{j=1}^{n}|g_{ij}|x^j+\sum_{j=1}^{n}|c_{ij}|\right)$$
$$=1-\frac{h}{2}\left(\sum_{j=1}^{n}|g_{ij}|\left(x^1+x^k\right)+\sum_{j=1}^{n}|c_{ij}|\right)>0,\quad i=1,\dots,n.$$

Therefore, if

$$h\le\tilde{h}:=\frac{2}{\frac{\Lambda}{\mu}\|G\|_\infty+\|C\|_\infty},\tag{44}$$

then both matrices $M(h,x)$ and $M(-h,x)$ are strictly diagonally dominant for all $x\in\Omega$ and $h\in(0,\tilde{h}]$. Hence the scheme can be written in the explicit form (31).

Our aim is to apply Theorem 7. Here, we will derive the invariance of Ω under (31) directly since one of the tangent conditions is satisfied as an equality, see (43). In this way we also derive a computable value of $\overline{h}$.

We deal first with the plane (42). Considering that $v^TP=v^TQ=(0,\dots,0)^T$ and that $v^TC=-\mu C$, we obtain that $v^TM(-h,x)=(1-\frac{h}{2}\mu)v^T$. Since C is a Metzler matrix, we have that $\mu=-\sum_{k=1}^{n}c_{ik}\le|c_{ii}|\le\|C\|_\infty$, $i=1,\dots,n$. Therefore, if h satisfies (44), v is a left eigenvalue of $M(-h,x)$ with a positive eigenvalue $1-\frac{h}{2}\mu$. Then the tangent condition (38) holds on the entire plane (42) due to Remark 8.

Next, we consider the boundary of Ω on the coordinate plane $x^1=S=0$. It follows from (35) that the vector $y=x-F(-h,x)$ is the solution $y=(y^1,\dots,y^n)$ of the linear system

$$M(-h,x)|_{S=0}\,y=hf(x)_{S=0},$$

or, equivalently,

$$\left(I+\frac{h}{2}\left(P(x)+C\right)\right)y=h(Cx+b).\tag{45}$$

For the tangent condition to hold, we need to show that

$$\left\langle -e_1,F(-h,x)-x\right\rangle=-F^1(-h,x)=y^1\ge0.$$

Using that $x\in\Omega$ and that y is a solution of the linear system (45), we obtain

$$\|y\|_\infty=\left\|h(Cx+b)-\frac{h}{2}\left(Gxe_1^T+C\right)y\right\|_\infty$$
$$\le h\left(\|C\|_\infty\frac{\Lambda}{\mu}+\|b\|_\infty\right)+\frac{h}{2}\left(\frac{\Lambda}{\mu}\|G\|_\infty+\|C\|_\infty\right)\|y\|_\infty.\tag{46}$$

Assume that (44) holds. Then, (46) yields

$$\begin{aligned}\|y\|_\infty &\le h\frac{\|C\|_\infty\frac{\Lambda}{\mu}+\|b\|_\infty}{1-\frac{h}{2}(\frac{\Lambda}{\mu}\|G\|_\infty+\|C\|_\infty)}\\ &\le 2h\left(\frac{\Lambda}{\mu}\|C\|_\infty+\|b\|_\infty\right).\end{aligned} \tag{47}$$

From the first equation in the system (45), we have

$$\begin{aligned}\left(1+\frac{h}{2}\left(-\sum_{j=2}^{n}g_{1j}x^j+c_{11}\right)\right)y^1 &= h\left(b_1+\sum_{j=2}^{n}c_{1j}x^j\right)-\frac{h}{2}\sum_{j=2}^{n}c_{1j}y^j\\ &\ge \frac{h}{2}\left(2b_1-\|y\|_\infty\sum_{j=2}^{n}c_{1j}\right)\\ &\ge h\left(b_1-h\left(\|C\|_\infty\frac{\Lambda}{\mu}+\|b\|_\infty\right)\sum_{j=2}^{n}c_{1j}\right).\end{aligned}$$

Taking into account that the coefficient of y^1 is positive, y^1 is nonnegative provided

$$h\le\tilde{\tilde{h}}:=\frac{b_1}{(\|C\|_\infty\frac{\Lambda}{\mu}+\|b\|_\infty)\sum_{j=2}^{n}c_{1j}}.$$

Let

$$\overline{h}=\min\{\tilde{h},\tilde{\tilde{h}}\}.$$

Then we have established so far that the tangent condition (38) holds on the entire plane (42) and on the boundary of Ω on the coordinate plane $x^1=0$ for all $h\in(0,\overline{h}]$.

The fact that $F^i(h,x)\ge 0$ for $i=2,\dots,n$, $x\in\Omega$, $h\in(0,\overline{h}]$ can be established in a more direct way. Let the $(n-1)\times(n-1)$ matrix $\tilde{M}(h,x)$ be obtained from $M(h,x)$ given in (30) by deleting the first row and the first column. It is easy to see that $\tilde{M}(h,x)=I-\frac{h}{2}(x^1\tilde{G}+\tilde{C})$, where $\tilde{G}$ and $\tilde{C}$ are obtained from matrices G and C by deleting their first row and the first column, respectively. We note that the non-diagonal entries $\tilde{M}(h,x)$ are non-positive. Further, $\tilde{M}(h,x)$ inherits from $M(x,h)$ its diagonal dominance. Hence, $\tilde{M}(h,x)$ is an M-matrix, which implies that $(\tilde{M}(h,x))^{-1}\ge 0$. Denoting $\tilde{b}=(b^2,\dots,b^n)^T\ge 0$, $\tilde{x}=(x^2,\dots,x^n)^T$, it follows from (32) that

$$\begin{pmatrix}F^2(h,x)\\F^3(h,x)\\\cdots\\F^n(h,x)\end{pmatrix}=\left(\tilde{M}(h,x)\right)^{-1}\left(\left(I+\frac{h}{2}\tilde{C}\right)\tilde{x}+\tilde{b}+F^1(h,x)\tilde{G}\tilde{x}\right)\ge 0, \tag{48}$$

for $x\in\Omega$ and $F^1(h,x)\ge 0$. Let us recall that, while we established the tangent property for the whole plane (42), we established the tangent property only for that part of the coordinate plate $x^1=0$, which is boundary of Ω. Hence, this result by itself does not imply $F^1(h,x)\ge 0$. We establish the nonnegativity of $F(h,x)$ as follows. The tangent condition on

the boundary of Ω on $x^1 = 0$, implies that as h increases from 0 to $\overline{h}$, $F(h,x)$ may not leave Ω through this part of the boundary. Let $x^1 > 0$. Then $F^1(h,x) > 0$ for h small enough. It follows from (48) that, as h increases, while $F^1(h,x) > 0$, we have $F(h,x) \geq 0$. Hence, $F(h,x)$ may leave Ω only through the boundary on the plane $x^1 = 0$. Due to the tangent condition, this may not happen for $h \in (0,\overline{h}]$. Therefore, $F(h,x) \in \Omega$ for all $x \in \Omega$ and $h \in (0,\overline{h}]$.

Thus, from Theorem 7, we obtain that the NSFD scheme for the model (40) is reversible, it defines a discrete dynamical system on the domain Ω of the continuous dynamical system (40), it is elementarily stable and second-order accurate.

Remark 10 The model (40) assumes only one compartment of susceptible individuals. In principle, the same method of proof works for a vector of susceptible compartments (for instance, as in [20]). We present this simple version of the model to avoid technicalities, which may obscure the main ideas demonstrated on this example. Further, we note that if there there is a disease induced mortality, (41) is satisfied as an inequality. In order to have (41) as equation, we can transform the model by introducing a compartment or compartments for deaths due to the disease with a removal rate equal to the natural death rate. This transformation increases the dimensions of the domain of the model, but allows for the case of disease induced mortality to be accommodated within the framework of (27) applied to the model (40).

The proof of the properties of NSFD method (27) applied to the model (40) uses the specific demographic equation (41). Let us note that demographic dynamics are not a restriction to applying the NSFD method (27). However, different demographic equations would require different approaches to proving the tangent condition for the domain of the model.

Essential for the application of the NSFD method (27) is the fact that the force of infection is represented through the mass action principle and is given in the form $\sum_{j=2}^{n} \beta_j x_j$. We may note that a model with standard incidence force of infection in the form $\sum_{j=2}^{n} \beta_j \frac{x_j}{N}$, is easily converted to mass action model for $z = \frac{x}{N}$, [9, 14]. Hence, the NSFD method (27) can be applied. However, the transformation changes the model and the domain, so that the proof of the tangent condition need to be adapted accordingly. Clearly, if the model cannot be written in a mass action form, the NSFD method (19) can be applied under the Assumption (20), but the method (27) is not applicable.

Remark 11 It follows from Theorem 7 that the NSFD scheme preserves all hyperbolic equilibria of the model as well as their stability properties, without any prior knowledge of the values of these equilibria or even their existence. In this way, the discetizations by (27) provide means of discovery of asymptotic dynamics of the model. This is solely done by using the nonlocal discretizations of nonlinear terms; a way to bring in our NSFD scheme Mikens' rule on nontrivial denominator function is discussed in [5]. A natural question is if this result cannot be extended to invariant sets other than equilibria. There are nonstandard numerical methods preserving, for specific models, specific invariant sets such as first integrals considered for instance by the authors [4]. However, constructing schemes and developing the associated theory, for preserving invariant sets in general is an open problem. Nevertheless, it is interesting that the method (27) applied to (40) preserves two important invariant sets on the boundary of the domain, as discussed in the next subsection.

5.3 Preserving invariant sets

The model (40) has two forward invariant sets on the boundary of the domain Ω: the disease-free manifold and the constant population manifold. We show that both are preserved by the numerical method (27).

The disease free manifold is given by

$$DFM = \{x \in \Omega : \langle e_j, x\rangle = 0, j = \mathcal{I}\}, \tag{49}$$

where $\mathcal{I}$ is the set of the indexes of the compartments promoting the disease—infective or transferring into infective. If the disease-free manifold exists, then the model (40) has the following properties

$$\langle e_j, b\rangle = 0, \quad j \in \mathcal{I}, \tag{50}$$

$$\text{if} \quad \langle e_j, x\rangle = 0, \quad j \in \mathcal{I}, \quad \text{then} \quad Gx = 0 \quad \text{and} \quad \langle (e_j, Cx\rangle = 0, \quad j \in \mathcal{I}. \tag{51}$$

Let $x \in DFM$. Then,

$$F(h,x) = \left(I - \frac{h}{2}\left(x^1 G + C\right)\right)^{-1}\left(\left(I + \frac{h}{2}C\right)x + hb\right).$$

Using (50)–(51), it is easy to see that

$$\left\langle e_j, \left(I + \frac{h}{2}C\right)x + hb\right\rangle = 0, \quad j \in \mathcal{I},$$

and, consequently, by mathematical induction,

$$\left\langle e_j, \left(x^1 G + C\right)^m \left(\left(I + \frac{h}{2}C\right)x + hb\right)\right\rangle = 0, \quad j \in \mathcal{I}, m = 1, 2, \ldots. \tag{52}$$

Since $h \leq \tilde{h}$, we have

$$\left(I - \frac{h}{2}\left(x^1 G + C\right)\right)^{-1} = \sum_{m=0}^{\infty} \frac{h^m}{2^m}\left(x^1 G + C\right)^m. \tag{53}$$

It follows from (52) and (53) that $\langle e_j, F(h,x)\rangle = 0, j \in \mathcal{I}$. Considering that it has been proved already that $F(h,x) \in \Omega$, we obtain that $F(h,x) \in DFM$, that is, DFM is forward invariant under $F(h,\cdot)$, $h \in (0, \overline{h}]$.

The constant population manifold is given by

$$\left\{x \in \Omega : \langle v, x\rangle = \frac{\Lambda}{\mu}\right\}. \tag{54}$$

Considering the specific form of the matrix $M(h,x)$ for (40), multiplying (28) on the left by v, we obtain

$$\left(1 + \frac{1}{2}\mu h\right)\langle v, x_{k+1}\rangle = \left(1 - \frac{1}{2}\mu h\right)\langle v, x_k\rangle + h\Lambda, \tag{55}$$

or, equivalently,

$$\langle v, F(h,x)\rangle = \frac{(1-\frac{1}{2}\mu h)\langle v,x\rangle + h\Lambda}{1+\frac{1}{2}\mu h}.$$

Direct substitution shows that if $\langle v,x\rangle = \frac{\Lambda}{\mu}$ then $\langle v, F(h,x)\rangle = \frac{\Lambda}{\mu}$. Therefore, (54) is forward invariant under $F(h,\cdot)$, $h \in (0,\overline{h}]$. Note that with the notation $N_k = \langle v, x_k\rangle = x_k^1 + x_k^2 + \cdots + x_k^n$, Equation (55) can be written as

$$\frac{N_{k+1}-N_k}{h} = -\frac{\mu}{2}(N_k + N_{k+1}),$$

which is a second-order finite difference method for the Equation (41).

5.4 Some illustrative numerical simulations

In Sect. 5.2 and Sect. 5.3, we established that the NSFD scheme (28) applied to the model (40) preserves some properties of (40) considered as a dynamical system. These include: the domain Ω, the hyperbolic equilibria with their stability, the disease-free manifold and the constant population manifold. These are properties of qualitative nature and cannot be derived through the techniques of the standard numerical analysis, involving consistence, stability and therefore convergence. To illustrate this point, we consider examples where well-known methods do not preserve the mentioned properties for a very simple version of (40), the SIR model:

$$\begin{aligned}
\frac{dS}{dt} &= \Lambda - \gamma SI - \mu S,\\
\frac{dI}{dt} &= \gamma SI - (\delta+\mu)I,\\
\frac{dR}{dt} &= (\delta+\mu)I.
\end{aligned} \tag{56}$$

For the simulations, we use (i) midpoint method, which is a second-order method, and (ii) the fourth-order Runge–Kutta method as provided by the function `ode45` in Matlab. Figures 1 and 2 provide phase diagrams of some numerical solutions in the $S \times I$ plane. The boundary of the domain is marked by a solid magenta line. The red star denotes the disease-free equilibrium, which is a stable equilibrium of the model (56) for the stated values of the parameters. Both simulations illustrate that numerical solutions produced by the stated methods may leave the domain of the model. Further, the trajectories by the Midpoint method presented in Fig. 1 converge (correctly) to the stable equilibrium, but some of them approach it from the exterior of the domain. The numerical solution presented in Fig. 2 eventually oscillates with nondecreasing amplitude as it approaches the stable equilibrium. Note that the Matlab function `ode45` automatically selects optimal step size. Hence, h is not specified.

For comparison, we present on Fig. 3 the numerical solutions by the method (27) with $h = 0.1$ using the same data as in Fig. 1 and in Fig. 2. For the model (56), we have $\overline{h} = \tilde{h} = \frac{2}{\gamma\frac{\Lambda}{\mu}+\nu+\delta}$. Hence, for both sets of values of parameters we have $h = 0.1 < \overline{h}$. Therefore, it follows from the discussion in Sect. 5.2 that the method (27) satisfies the properties stated in Theorem 7. One can also observe on the graphs in Fig. 3 that both the domain

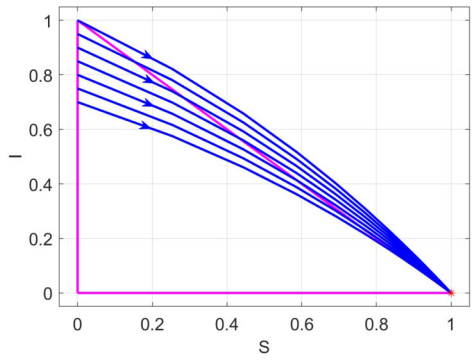

Figure 1 Phase diagram of the numerical solution of (56) with $\Lambda = \mu = 3$, $\gamma = 1$, $\delta = 15$ by the Midpoint method with $h = 0.1$

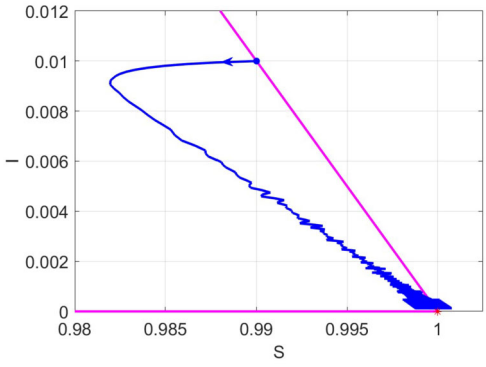

Figure 2 Phase diagram of the numerical solution of (56) with parameters $\Lambda = \mu = 1$, $\gamma = 2$, $\delta = 1$ by the Matlab function `ode45`

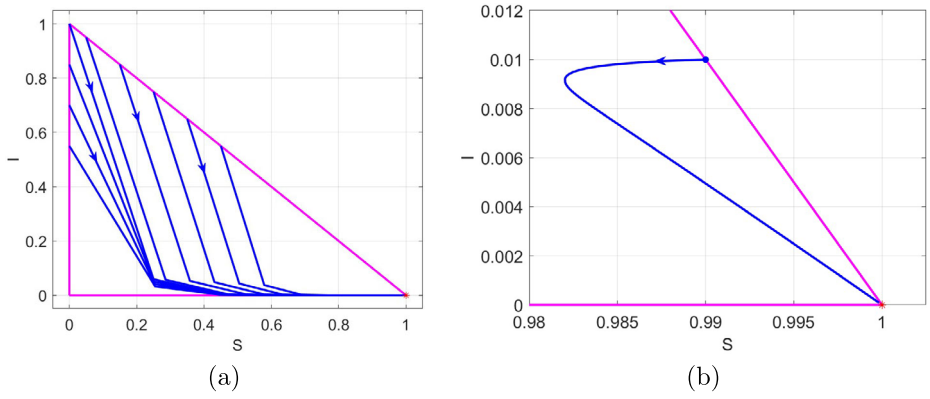

Figure 3 Phase diagrams of the numerical solutions by the NSFD method (27) of (56) with $h = 0.1$ and parameters as in (a) Fig. 1 and as in (b) Fig. 2

and the stability of the equilibrium of the model (56) are preserved. Further, we may note that the values of the parameters used in Fig. 1 are associated with a disease, which is low contagious and with a fast removal rate. Hence, all solutions very quickly approach the disease-free manifold, which is the primary driver of the long-term dynamics. This

property is correctly represented by the trajectories on Fig. 3(a), but not by the trajectories on Fig. 1, including those within the domain.

6 Conclusion

The literature on forward invariant sets in $\mathbb{R}^n$ with respect to continuous dynamical systems defined by ordinary differential equations is rich (see, for instance, the short account in [27]). However, the invariance problem is not sufficiently investigated for discrete dynamical systems. Our findings in this work are threefold. Firstly, we formulated a discrete analog of the tangent condition by Bony [7], under which we established the forward invariance of sets with respect to time-reversible discrete dynamical systems. Secondly, we constructed a new NSFD scheme to approximate the solutions of continuous dynamical systems. Apart from its elementary stability, this new scheme is innovative and reliable in that: (a) it is time reversible, (b) it inherits in a discrete manner the tangent condition and the associated forward invariant property from the continuous system, and (c) it has second-order accuracy, thereby addressing the challenge of designing high-order numerical schemes that cannot exhibit spurious/ghost solutions or other elementary instability that do not correspond to the feature of the continuous equations. Thirdly, we showed that our construction and findings apply directly for mass action-type models in biological and chemical processes, specifically the preservation by the NSFD scheme of forward invariance of the biologically feasible domain and other dynamics of the continuous system.

Our future research includes the following:

(1) Formulating a discrete analog of the tangent condition (3) by Nagumo [21] in the case when the latter is not equivalent to Bony's condition (4). Conditions for their equivalence are stated in [27].
(2) Investigate the preservation by our time-reversible NSFD scheme of the global asymptotic stability of the disease-free equilibrium of continuous epidemic models when the basic reproduction number is less than 1. (See [2] for an approach, which avoids the use of the Lyapunov function for discrete dynamical systems.)
(3) Extending this study to dynamical systems with non-hyperbolic equilibrium points, an approach considered in [6] as far as the stability for one-dimensional models is concerned.
(4) Though a standard incidence-based epidemic model can be transformed into a mass action-based one (with different dynamics) to which the NSFD scheme proposed in this paper can be applied, as mentioned in Remark 10, investigating the reverse process in order to recover the preserved dynamics of the initial model is an issue of interest.
(5) Other possible generalizations are: (i) NSFD scheme (27) applied to extended model (40) with vector of susceptible compartments and/or disease-induced death rate, (ii) models with demographics different from (40), and (iii) models where the general method (19) can be applied but not the method (27).

Acknowledgements

Apart from the acknowledgements for funding mentioned above, R. Anguelov acknowledges the hospitality of the University of the Witwatersrand, where this work was finalized during his research visit. Both authors are greatful to the two anonymous reviewers for their sound comments, remarks and suggestions, which greatly contributed to the improvement of this paper.

Funding
R Anguelov was partially supported by the DSI/NRF SARChI Chair on Mathematical Models and Methods in Bioengineering and Biosciences at the University of Pretoria. J Lubuma is grateful to the NRF for financial support under the Competitive Programme for Rated Researchers (CPRR). He also acknowledges the support of the University of the Witwatersrand under the Science Faculty Start-up Funds for Research.

Availability of data and materials
Not applicable.

Declarations

Competing interests
The authors declare no competing interests.

Author contributions
Equal contribution by the authors. All authors read and approved the final manuscript.

Author details
[1]Department of Mathematics and Applied Mathematics, University of Pretoria, Pretoria, South Africa. [2]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria. [3]School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa.

Received: 16 March 2023 Accepted: 6 September 2023 Published online: 22 September 2023

References

1. Anguelov, R., Berge, T., Chapwanya, M., Djoko, J.K., Kama, P., Lubuma, J.M.-S., Terefe, Y.: Nonstandard finite difference method revisited and application to the Ebola virus disease dynamics transmission. J. Differ. Equ. Appl. **26**(6), 818–854 (2020)
2. Anguelov, R., Dumont, Y., Lubuma, J.M.-S., Shillor, M.: Dynamically consistent nonstandard finite difference schemes for epidemiological models. J. Comput. Appl. Math. **255**, 161–182 (2014)
3. Anguelov, R., Kama, P., Lubuma, J.M.-S.: On nonstandard finite difference models of reaction-diffusion equations. J. Comput. Appl. Math. **175**, 11–29 (2005)
4. Anguelov, R., Lubuma, J.M.-S.: Contributions to the mathematics of the nonstandard finite difference method and applications. Numer. Methods Partial Differ. Equ. **17**, 518–543 (2001)
5. Anguelov, R., Lubuma, J.M.-S.: A second-order nonstandard finite difference scheme and application to model of biological and chemical processes. In: Gumel, A.B. (ed.) Mathematical and Computational Modeling of Phenomena Arising in Population Biology and Nonlinear Oscillations. AMS Contemporary Mathematics (in press)
6. Anguelov, R., Lubuma, J.M.-S., Shillor, M.: Topological dynamic consistency of nonstandard finite difference schemes for dynamical systems. J. Differ. Equ. Appl. **17**(12), 1769–1791 (2011)
7. Bony, J.M.: Principe du maximum, inégalité de Harnack et unicité du problème de Cauchy pour les opérateurs elliptiques dégénérés. Ann. Inst. Fourier **19**(1), 277–304 (1969)
8. Brezis, H.: On a characterization of flow-invariant sets. Commun. Pure Appl. Math. **XXIII**, 261–263 (1970)
9. Castillo-Chavez, C., Hethcote, H.W., Andreasen, V., Levin, S.A., Liu, W.M.: Epidemiological models with age structure, proportionate mixing, and cross-immunity. J. Math. Biol. **27**, 233–258 (1989)
10. Crandall, M.G.: A generalization of Peano's existence theorem and flow invariance. Proc. Am. Math. Soc. **36**(1), 151–155 (1972)
11. Gumel, A.B., Connel McCluskey, C., Watmough, J.: An sveir model for assessing potential impact of an imperfect anti-SARS vaccine. Math. Biosci. Eng. **3**(3), 485–512 (2006)
12. Hairer, E., Lubich, C., Wanner, G.: Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations. Springer, Berlin (2006)
13. Hartman, P.: On invariant sets and on a theorem of Wazewski. Proc. Am. Math. Soc. **32**(2), 511–520 (1972)
14. Hethcote, H.W.: The mathematics of infectious disease. SIAM Rev. **42**, 599–653 (2000)
15. Kojouharov, H.V., Roy, S., Gupta, M., Alalhareth, F., Slezak, J.M.: A second-order modified nonstandard theta method for autonomous differential equations. Appl. Math. Lett. **112**, 106775 (2021)
16. Lubuma, J.M.-S., Roux, A.: An improved theta method for systems of ordinary differential equations. J. Differ. Equ. Appl. **9**, 1023–1035 (2003)
17. Luenberger, D.G.: Optimization by Vector Space Methods. Wiley, New York (1969)
18. Mickens, R.E.: Nonstandard Finite Difference Models of Differential Equations. World Scientific, Singapore (1994)
19. Mickens, R.E.: Nonstandard Finite Difference Schemes: Methodology and Applications. World Scientific, Singapore (2021)
20. Moya, E.D., Rodrigues, D.S., Pietrus, A., Severo, A.M.: A mathematical model for HIV/AIDS under pre-exposure and post-exposure prophylaxis. Biomathematics **11**, 2208319 (2022). https://doi.org/10.55630/j.biomath.2022.08.3191/28
21. Nagumo, M.: Über die Lage der Integralkurven gewöhnlicher Differentialgleichungen. In: Proceedings of the Physico-Mathematical Society of Japan, 3rd Series, vol. 24, pp. 551–559 (1942)
22. Otunuga, O.M., Ogunsolu, M.O.: Qualitative analysis of a stochastic SEITR epidemic model with multiple stages of infection and treatment. Infect. Dis. Model. **5**, 61–90 (2020)
23. Redheffer, R.M.: The theorems of Bony and Brezis on flow-invariant sets. Am. Math. Mon. **79**(7), 740–747 (1972)
24. Smith, H.L.: Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems. Am. Math. Soc., Providence (1995)
25. Stuart, A.M., Humphries, A.R.: Dynamical Systems and Numerical Analysis. Cambridge University Press, New York (1998)

26. van den Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**, 29–48 (2002)
27. Walter, W.: Ordinary Differential Equations. Springer, New York (1998)
28. Yamazaki, K., Wang, X.: Global well-posedness and asymptotic behaviour of solutions to a reaction-convective-diffusion cholera epidemic model. Discrete Contin. Dyn. Syst., Ser. B **21**, 1297–1316 (2016)

Publisher's Note