



Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational
Mathematics (IMACM)

Preprint BUW-IMACM 26/17

Achraf Zinihi, Matthias Ehrhardt and Moulay Rchid Sidi Ammi

Structure-Preserving Neural ODEs via Nonstandard Finite Difference Discretization

July 12, 2026

<http://www.imacm.uni-wuppertal.de>

Structure-Preserving Neural ODEs via Nonstandard Finite Difference Discretization

Achraf Zinihi^{a,b}, Matthias Ehrhardt^{a,*}, Moulay Rchid Sidi Ammi^b

^a*University of Wuppertal, Applied and Computational Mathematics,
Gaußstrasse 20, 42119 Wuppertal, Germany*

^b*Moulay Ismail University of Meknes, FST Errachidia,
Department of Mathematics, AMNEA Group, Errachidia 52000, Morocco*

Abstract

Although neural ordinary differential equations (NODEs) are a powerful framework for learning continuous-time dynamics, they generally do not preserve essential qualitative properties, such as positivity. We propose a structure-preserving Neural ODE framework based on nonstandard finite difference (NSFD) discretization. The learned dynamics are parameterized by nonnegative production and destruction rates, yielding an explicit, differentiable update that integrates seamlessly into standard automatic differentiation pipelines. We prove that the resulting scheme unconditionally preserves positivity for arbitrary time-step sizes while retaining first-order consistency. We outline an extension based on Patankar-type discretizations that preserves conservation laws exactly. Numerical experiments on an SIR epidemic model show that our approach generates physically meaningful trajectories, remains robust under coarse discretizations, and outperforms conventional NODEs in preserving the qualitative structure of the learned dynamics.

Keywords: Neural ODEs, NSFD schemes, Positivity preservation, Structure-preserving discretization, Scientific machine learning

2020 Mathematics Subject Classification: 65L05, 65L20, 68T07, 92D30

1. Introduction

Neural ODEs [1] parameterize the right-hand side of an initial value problem $\dot{x} = f_\theta(x, t)$ by a neural network f_θ and treat the network as an implicit, continuous-depth model that is fit to data by differentiating through a numerical ODE solver. This formulation is advantageous for learning the latent dynamics of physical and biological systems, including epidemic models, because it preserves the continuous-time structure of the underlying phenomenon rather than treating time as a discrete, recurrent index. However, in these applications, the

*Corresponding author

Email addresses: a.zinihi@edu.umi.ac.ma (Achraf Zinihi), ehrhardt@uni-wuppertal.de (Matthias Ehrhardt), rachidsidiammi@yahoo.fr (Moulay Rchid Sidi Ammi)

state variables usually have a physical meaning, such as compartment sizes, concentrations, or population counts, and they must remain nonnegative. Often, they also satisfy additional invariants, such as a constant total population. A generic f_θ provides no such guarantee: nothing in the training objective prevents the learned flow from producing negative states, particularly under extrapolation or coarse time stepping, and the standard remedies (clipping, penalty terms for constraint violations, or projection back onto the feasible region after each solver step) are post-hoc corrections rather than structural guarantees. Furthermore, these remedies can also introduce non-smoothness or bias into the learned dynamics.

A separate line of work in numerical analysis, dating back to [2] and rigorously developed by [3] and many others (see [4] and the references therein for recent high-order constructions), shows that *nonstandard finite difference* (NSFD) schemes can be designed so that the discrete scheme preserves the positivity, boundedness, and other qualitative features of a continuous dynamical system for *every* step size, not just for a step size below some stability threshold. The central device is a nonlocal treatment of the right-hand side, along with a generalized, model-dependent denominator function that replaces Δt . This approach has been applied extensively to compartmental epidemic models with prescribed (not learned) right-hand sides [5]. In the NODE literature, "structure" usually refers to a soft, statistical constraint, such as an information bottleneck for dimensionality reduction or a linear-in-features form motivated by Koopman operator theory, rather than a hard, provable guarantee on the state itself. By contrast, this work provides a structural constraint of the latter, provable kind.

This work contributes by combining two approaches: we constrain a NODE's vector field to a gain/loss (production/destruction) form and pair it with an NSFD time-stepping rule. This guarantees the positivity of the learned dynamics unconditionally, rather than approximately through penalization. Specifically, we: (i) parameterize the network so that it outputs nonnegative gain and loss rates rather than an unconstrained vector field (Section 2); (ii) construct an NSFD update for this class of networks and prove unconditional positivity for any step size (Section 3); (iii) show that the update is a closed-form, differentiable function of the network output; therefore, it requires no additional computation beyond that used in a standard NODE training loop (Proposition 3); (iv) outline how a total-population invariant can be preserved exactly by transitioning from an independent gain/loss parameterization to a flux-based, Patankar-type scheme [6] (Section 4); and (v) Demonstrate the method on a latent SIR learning problem by comparing it to an unconstrained NODE baseline (Section 5).

2. Gain-Loss Neural ODEs

Let $x(t) \in \mathbb{R}_+^d$ denote the latent state, and suppose the true generative dynamics admit a decomposition into nonnegative gain and loss terms,

$$\dot{x}_i(t) = G_i(x(t), t) - L_i(x(t), t) x_i(t), \quad i = 1, \dots, d, \quad (1)$$

with $G_i \geq 0$ and $L_i \geq 0$ are sufficiently smooth functions on $\mathbb{R}_+^d$. This is not a restriction on the class of dynamics: any locally Lipschitz vector field f satisfying the subtangential condition $f_i(x) \geq 0$ whenever $x_i = 0$ (the standard necessary and sufficient condition for $\mathbb{R}_+^d$

to be positively invariant) can be written in this form, e.g. by taking $L_i \equiv 0$ and $G_i = f_i$ wherever $f_i \geq 0$, and folding negative contributions into $L_i x_i$ elsewhere; the decomposition simply makes the sign structure explicit rather than leaving it implicit in a single network. This decomposition is used to define the positive and negative parts of f_θ . For further details, we refer the reader to [4]. Compartmental epidemic models are a natural fit: in the SIR model of Section 5, new infections are a gain for I and a loss for S , and recovery is a gain for R and a loss for I .

We parameterize G_i and L_i by feed-forward networks with a final softplus (or squared) output activation

$$G_{\theta,i}(x, t) = \text{softplus}(g_{\theta,i}(x, t)) \quad \text{and} \quad L_{\theta,i}(x, t) = \text{softplus}(\ell_{\theta,i}(x, t)),$$

where $g_{\theta,i}, \ell_{\theta,i}: \mathbb{R}^d \times [0, T] \rightarrow \mathbb{R}$ are unconstrained networks. This guarantees $G_{\theta,i} \geq 0$ and $L_{\theta,i} \geq 0$ for every θ , which is the only hypothesis that Theorem 2 below requires. No constraint is imposed on θ during training.

3. An NSFD Update for Gain-Loss Neural ODEs

Given a time grid $t_n = n\Delta t$ and state $x^n \approx x(t_n)$, we propose the following update

$$\frac{x_i^{n+1} - x_i^n}{\phi(\Delta t)} = G_{\theta,i}(x^n, t_n) - L_{\theta,i}(x^n, t_n) x_i^{n+1}, \quad i = 1, \dots, d, \quad (2)$$

where the gain term is evaluated explicitly at x^n , and the loss term is evaluated semi-implicitly, by multiplying the *new* state x_i^{n+1} . The denominator function ϕ replaces Δt and must satisfy the standard NSFD consistency conditions: $\phi(\Delta t) > 0$, $\phi(\Delta t) = \Delta t + O(\Delta t^2)$ as $\Delta t \rightarrow 0^+$, so that (2) reduces to the semi-implicit Euler discretization of (1) in the small-step limit, and it is a first-order consistent scheme. The simplest admissible choice is $\phi(\Delta t) = \Delta t$ itself; more elaborate choices, e.g. $\phi(\Delta t) = (1 - e^{-\lambda\Delta t})/\lambda$, can be used to match a known local rate λ exactly, as discussed in [2, 4, 5].

Since the loss term is linear in x_i^{n+1} , (2) can be solved for x_i^{n+1} in closed form:

$$x_i^{n+1} = \frac{x_i^n + \phi(\Delta t) G_{\theta,i}(x^n, t_n)}{1 + \phi(\Delta t) L_{\theta,i}(x^n, t_n)}, \quad i = 1, \dots, d. \quad (3)$$

Unlike a generic implicit scheme, (3) requires no nonlinear solving at each step: it is an explicit, elementwise rational function of the network outputs at x^n . Consequently, (3) defines a differentiable computational layer that can directly replace a conventional numerical integration layer in NODE training. Since the update consists only of differentiable operations with a strictly positive denominator, it is fully compatible with standard automatic differentiation frameworks.

Proposition 1 (Consistency). *Assume that the exact solution $x(t)$ of (1) is sufficiently smooth on $[0, T]$, and that G_i and L_i are continuously differentiable. Then the NSFD scheme (3) is first-order consistent in time.*

Proof. Let $x(t)$ be the exact solution of (1). By Taylor expansion, we have that

$$x_i(t_{n+1}) = x_i(t_n) + \Delta t \dot{x}_i(t_n) + O(\Delta t^2).$$

Using (1), this becomes

$$x_i(t_{n+1}) = x_i(t_n) + \Delta t \left(G_i(x(t_n), t_n) - L_i(x(t_n), t_n) x_i(t_n) \right) + O(\Delta t^2).$$

Now evaluate the NSFD update (2) along the exact solution. Since $x_i(t_{n+1}) = x_i(t_n) + O(\Delta t)$, and G_i, L_i are smooth, we have

$$L_i(x(t_n), t_n) x_i(t_{n+1}) = L_i(x(t_n), t_n) x_i(t_n) + O(\Delta t).$$

Using $\phi(\Delta t) = \Delta t + O(\Delta t^2)$, the discrete update produces the same expansion up to $O(\Delta t^2)$. Therefore, the local truncation error satisfies $\tau_i^n = O(\Delta t^2)$, which proves first-order consistency in time. $\square$

Thus, the NSFD update (2) has the same formal temporal accuracy as the semi-implicit Euler method and additionally preserves positivity unconditionally.

Theorem 2 (Unconditional Positivity and Boundedness). *Let $x^0 \in \mathbb{R}_+^d$. Then, the sequence generated by (3) satisfies $x^n \in \mathbb{R}_+^d$ for all $n \geq 0$. Furthermore, if the scheme preserves the invariant set $\Omega \subset \mathbb{R}_+^d$, then $x^n \in \Omega$ for all $n \geq 0$, and thus the numerical solution is bounded.*

Proof. By induction on n . The base case $x^0 \in \mathbb{R}_+^d$ holds by assumption. Suppose $x^n \in \mathbb{R}_+^d$. Then for each i , the numerator of (3) is a sum of two nonnegative terms, $x_i^n \geq 0$ and $\phi(\Delta t) G_{\theta,i}(x^n, t_n) \geq 0$, and the denominator is $1 + \phi(\Delta t) L_{\theta,i}(x^n, t_n) \geq 1 > 0$. Hence $x_i^{n+1} \geq 0$ for every i , so $x^{n+1} \in \mathbb{R}_+^d$. If the scheme also preserves a bounded invariant region Ω , then $x^n \in \Omega$ for all $n \geq 0$. Since Ω is bounded, the numerical solution is bounded. $\square$

Note that no bound on the step size Δt is required. The result holds for arbitrarily large step sizes, which is the defining feature of an NSFD scheme as opposed to a standard scheme whose positivity is only conditional on a step-size restriction.

Proposition 3 (Differentiability). *If $g_{\theta,i}, \ell_{\theta,i}$ are differentiable with respect to θ for every x, t , then x_i^{n+1} is also differentiable with respect to θ for every n .*

Proof. Immediate: (3) composes the differentiable softplus map, elementary arithmetic operations with a strictly positive denominator, and $g_{\theta,i}, \ell_{\theta,i}$, all of these are differentiable, and the composition is applied recursively in n . $\square$

Remark 1. Proposition 3 allows (3) to be used as a drop-in replacement for a standard NODE solver step in existing training pipelines. Training proceeds exactly as it would for a discretized NODE, with no need for an adjoint sensitivity equation or an implicit-function-theorem solve, since (3) is already explicit in x^{n+1} .

4. Towards Exact Conservation

Many epidemic models additionally conserve a total quantity, e.g. a closed population $N = S + I + R$. However, the gain/loss parameterization of Section 2 does not enforce this by itself, because $G_{\theta,i}$ and $L_{\theta,i}$ are independent across i . Exact conservation requires that each unit of loss from one compartment appear as a matching unit of gain in another. This structure is studied under the name of *production-destruction systems*, for which positivity and conservation-preserving schemes are known [6]. We outline the corresponding extension here as a natural next step rather than a result we prove in full. We replace the independent networks $G_{\theta,i}$, $L_{\theta,i}$ by a single set of pairwise, nonnegative flux networks $\Phi_{\theta,ij}(x, t) \geq 0$ ($i \neq j$) representing the learned rate of transfer from compartment i to compartment j :

$$G_{\theta,i} = \sum_{j \neq i} \Phi_{\theta,ji} \quad \text{and} \quad L_{\theta,i} x_i = \sum_{j \neq i} \Phi_{\theta,ij}.$$

At the continuous level this decomposition telescopes to exactly

$$\sum_i \dot{x}_i = \sum_i \sum_{j \neq i} \Phi_{\theta,ji} - \sum_i \sum_{j \neq i} \Phi_{\theta,ij} = 0,$$

so the total population is conserved. Simultaneously achieving the *discrete* analogue with unconditional positivity is precisely the problem modified Patankar Runge-Kutta (MPRK) schemes solve for prescribed production-destruction systems, by treating the full off-diagonal flux structure implicitly and solving the resulting (sparse, nonnegative) linear system at each step rather than the diagonal closed form (3). A full positivity-and-conservation proof for the learned flux network Φ_θ , together with the resulting linear-solve training step, is left for future work. We mention this here because it directly targets the stronger " N exactly constant" property that motivated this project, beyond the positivity guarantee already established in Theorem 2.

5. Example and Numerical Results

We illustrate the proposed method on the classical SIR model

$$\dot{S} = -\beta SI, \quad \dot{I} = \beta SI - \gamma I, \quad \dot{R} = \gamma I,$$

which is already in gain/loss form (1) with

$$G_S = 0, \quad L_S = \beta I, \quad G_I = \beta SI, \quad L_I = \gamma, \quad G_R = \gamma I, \quad L_R = 0.$$

We treat the parameters β, γ as unknown and replace them with the gain/loss networks of Section 2. The resulting NSFD update, from (3), is

$$S^{n+1} = \frac{S^n}{1 + \phi(\Delta t) L_{\theta,S}(x^n, t_n)}, \quad I^{n+1} = \frac{I^n + \phi(\Delta t) G_{\theta,I}(x^n, t_n)}{1 + \phi(\Delta t) L_{\theta,I}(x^n, t_n)},$$

and

$$R^{n+1} = R^n + \phi(\Delta t) G_{\theta,R}(x^n, t_n).$$

Remark 2 (A free conservation law for the closed SIR model). Because $\dot{S} + \dot{I} + \dot{R} \equiv 0$, the total population $N = S + I + R$ is conserved by the true dynamics for any β , γ , and $R(t) = N_0 - S(t) - I(t)$. This allows us to avoid the general flux construction of Section 4 here. We only model the (S, I) system with the gain/loss networks $L_{\theta,S}$, $G_{\theta,I}$, $L_{\theta,I}$, as well as the update equation (3), and we recover $R^n := N_0 - S^n - I^n$ algebraically. Section 4’s construction remains necessary when no such algebraic reduction exists (e.g. open populations, or conservation laws not equal to a simple state sum). One caveat: nothing here guarantees $S + I \leq N_0$, so the recovered R^n is not itself covered by Theorem 2, which applies modeled states (S, I) ; we report this quantity honestly in this section rather than assume it away.

Setup. We generated a reference trajectory from the SIR model above with $\beta = 0.4$, $\gamma = 0.1$, $S(0) = 0.99$, $I(0) = 0.01$, $R(0) = 0$ (so $N_0 = 1$), integrated it over $t \in [0, 100]$. We then train two models on the same noisy, subsampled observations of (S, I) at $\Delta t = 1$ (holding out the tail of the trajectory for extrapolation testing):

- *Baseline (B-NODE)*: an unconstrained NODE on the full 3D state (S, I, R) , $\dot{x} = f_\theta(x, t)$ with f_θ a small Multilayer Perceptron (MLP), integrated with explicit Euler and trained by backpropagation through the unrolled steps.
- *NSFD-NODE*: the reduced (S, I) gain/loss network of Remark 2 with the update (3), trained in the same manner (Proposition 3).

Both models use MLPs of the same size (two hidden layers, 32 units, tanh activations, gradient norm clipped to 5), so any difference in behavior is due to the discretization and output parameterization rather than model capacity. We use the same architecture-matched comparison logic to isolate the effect of structuring a NODE’s vector field in [7]. Both models converge to a comparably low training loss ($\sim 10^{-4}$), indicating that the differences in the predictions stem from the discretization strategy rather than differences in fitting accuracy.

Metrics. We evaluated the two models, which were trained for 3000 epochs, without retraining, at three step sizes: a finer-than-training step size ($\Delta t = 0.5$), the training step size ($\Delta t = 1$), and a *coarse-step stress test* ($\Delta t = 5$, five times coarser than the training step). This directly targets Theorem 2, whose guarantee holds for every Δt , not only the one used in training. We report: (i) the minimum state value and the number of negative entries, over the full (S, I, R) trajectory; (ii) the root-mean-square error (RMSE) on (S, I) over the held-out extrapolation window; and (iii) the drift $\max_n |N^n - N_0|$ in total population.

Results. Tables 1 and 2 report the outcome, and Figure 1 shows the trajectories at $\Delta t = 1$. Two things stand out. First, N -conservation is exact for NSFD-NODE at every step size by construction (drift ≈ 0 to floating-point precision) by construction, while the unconstrained baseline drifts substantially (2.9 at $\Delta t \leq 1$, and 18.35 under the coarse-step test) despite matching training loss. Second, positivity: NSFD-NODE’s modeled states (S, I) never go negative, as guaranteed. The baseline substantially violates positivity at every step size, and catastrophically so under the coarse-step test (minimum $-3.19e^{-00}$, RMSE 3.32), despite comparably low training loss. Evidently, low training error does not guarantee of physically consistent unrolled dynamics.

Table 1: Numerical comparison of the unconstrained B-NODE and NSFD-NODE, at a finer-than-training step $\Delta t = 0.5$, the training step size $\Delta t = 1$ and under a coarse-step stress test $\Delta t = 5$.

Step Size	Model	Min. State Value	Negative Entries	RMSE	Max. $ N^n - N^0 $
$\Delta t = 0.5$	B-NODE	$-2.4504 e^{-02}$	72	0.0171	2.9403
	NSFD-NODE	0.0000	0	0.0146	0.0000
$\Delta t = 1$	B-NODE	$-2.4712 e^{-02}$	37	0.0174	2.9409
	NSFD-NODE	0.0000	0	0.0145	0.0000
$\Delta t = 5$	B-NODE	$-3.1906 e^{-00}$	19	3.3236	18.3525
	NSFD-NODE	0.0000	0	0.0249	0.0000

Table 2: Minimum and maximum values of the state variables obtained by the B-NODE and the proposed NSFD-NODE for different time step sizes.

Step Size	S B-NODE		S NSFD-NODE		I B-NODE		I NSFD-NODE		R B-NODE		R NSFD-NODE	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
$\Delta t = 0.5$	0.0043	1.0198	0.0003	0.9900	-0.0245	0.4048	0.0100	0.3913	0.0000	3.9605	0.0010	0.9868
$\Delta t = 1$	0.0043	1.0442	0.0004	0.9900	-0.0247	0.4080	0.0100	0.3993	0.0000	3.9613	0.0007	0.9867
$\Delta t = 5$	-3.1906	6.6857	0.0011	0.9900	-1.3762	5.8326	0.0100	0.4443	-0.7115	11.6093	0.0000	0.9852

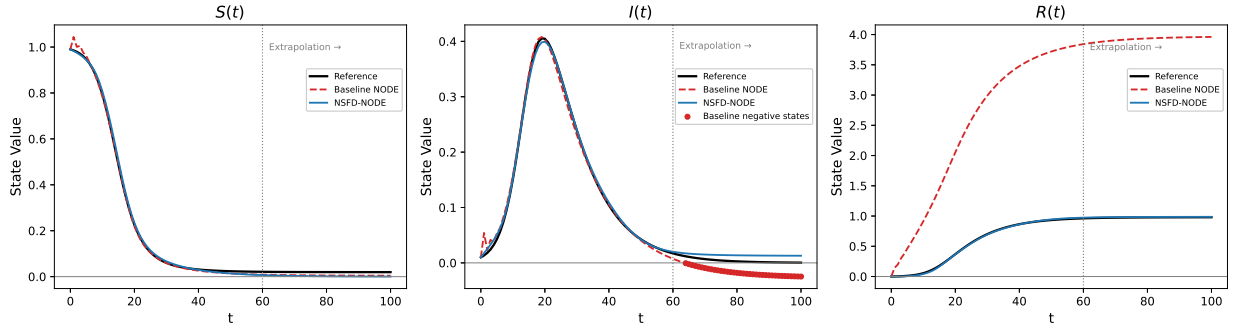


Figure 1: Reference, B-NODE (negative excursions marked), and NSFD-NODE trajectories for $S(t)$, $I(t)$, and $R(t)$ at $\Delta t = 1$. R is recovered algebraically for NSFD-NODE (Remark 2). Dotted line: end of training window.

The main advantage of the scheme is that positivity is a structural consequence of equation (3) rather than a property encouraged by a penalty term. Theorems 2 holds for every θ and every Δt , with no tuning of a penalty weight and no risk of violating the constraint on out-of-distribution inputs. This is precisely where penalty-based approaches tend to fail. However, the cost is architectural; the network must be written in gain/loss form, which is natural for compartmental and production-destruction systems, but restrictive compared to an arbitrary f_θ .

6. Conclusion

We introduced an NSFD discretization for neural ODEs based on a gain/loss parameterization. This update preserves positivity unconditionally, remains bounded, and is differentiable in closed form. These properties make it directly compatible with standard Neural ODE training pipelines. This method addresses a critical limitation of scientific neural ODEs by enforcing physically meaningful nonnegative states via structural discretization instead of a penalty or projection step.

Our work establishes a rigorous connection between structure-preserving numerical analysis and neural ODEs. It demonstrates that qualitative properties can be enforced by design rather than through penalties or projections.

Declarations

Conflict of Interest

The authors declared that they have no conflict of interest.

Data Availability

No data was used for the research described in the article.

References

- [1] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. *Adv. Neural. Inf. Process. Syst.*, 31, 2018.
- [2] R. E. Mickens. *Nonstandard finite difference models of differential equations*. World Scientific, 1994.
- [3] R. Anguelov and J. M. S. Lubuma. Contributions to the mathematics of the nonstandard finite difference method and applications. *Numer. Meth. Part. Diff. Eqs.*, 17(5):518–543, 2001.
- [4] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. A nonstandard finite difference scheme for a nonlinear parabolic equation with p-Laplacian-type diffusion. *arXiv preprint arXiv:2607.00489*, 2026.
- [5] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. A nonstandard finite difference scheme for an SEIQR epidemiological PDE model. *Appl. Math. Comput.*, 520:129953, 2026.
- [6] H. Burchard, E. Deleersnijder, and A. Meister. A high-order conservative patankar-type discretisation for stiff systems of production-destruction equations. *Appl. Numer. Math.*, 47(1):1–30, 2003.
- [7] Jake Buzhardt, C. Ricardo Constante-Amores, and Michael D. Graham. On the relationship between Koopman operator approximations and neural ordinary differential equations for data-driven time-evolution predictions. *Chaos Interdiscip. J. Nonli. Sci.*, 35(4), 2025.