Efficient relaxation scheme for the SIR and related compartmental models

Vo Anh Khoa
*Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, USA and Department of Mathematics, Florida A&M University, Tallahassee, FL 32307, USA**

Pham Minh Quan
Department of Computer Science, Florida State University, Tallahassee, FL 32306, USA

Ja'Niyah Allen and Kbenesh W. Blayneh
*Department of Mathematics, Florida A&M University, Tallahassee, FL 32307, USA**

In this paper, we introduce a novel numerical approach for approximating the Susceptible-Infectious-Recovered (SIR) model in epidemiology. Our method enhances the existing linearization procedure by incorporating a suitable relaxation term to tackle the transcendental equation of nonlinear type. Developed within the continuous framework, our relaxation method is explicit and easy to implement, relying on a sequence of linear differential equations. This approach yields accurate approximations in both discrete and analytical forms. Through rigorous analysis, we prove that, with an appropriate choice of the relaxation parameter, our numerical scheme is non-negativity-preserving; moreover, it is strongly convergent to the true solution. We also extend the applicability of our relaxation method to handle some variations of the traditional SIR model. Finally, we present numerical examples using simulated data to demonstrate the effectiveness of our proposed method.

Keywords: SIR models, relaxation, global convergence, infectious diseases

I. INTRODUCTION

A. Background

In recent years, the world has witnessed the devastating impact of infectious diseases on a global scale. From the rapid spread of COVID-19 to the resurgence of long-standing ailments like measles and influenza, understanding the dynamics of epidemics has become crucial for protecting public health. To gain a deeper understanding of these intricate phenomena, scientists have increasingly relied on mathematical modeling as an influential tool for unraveling the complex mechanisms governing disease transmission. Among the various models, the Susceptible-Infectious-Recovered (SIR) model has emerged as a fundamental framework, providing valuable insights into epidemic dynamics; cf. e.g. [7, 12, 16] for its applications to modeling the influenza, Ebola and COVID-19.

The SIR model, initially proposed in the early 20th century by Kermack and McKendrick [11], has since been refined and adapted to address contemporary challenges. This model effectively captures the fundamental dynamics of epidemics by dividing a population into three distinct compartments: susceptible individuals, infectious individuals, and removed individuals. By considering the interactions between these compartments, the SIR model takes into account various factors such as transmission rates, removal rates, and the depletion of susceptible individuals over time. The "removals" in this context encompass individuals who are isolated, deceased, or have recovered and gained immunity. Additionally, the model assumes that individuals who have gained immunity or recovered enter a new category that is not susceptible to the disease.

Consider a homogeneously mixed group of individuals of total size $N \gg 1$. Let $t \in (0, T)$ be the time variable with $T > 0$ being the final time of observations. We take into account the following functions:

$$\begin{aligned} S(t) &= \text{number of susceptibles at time } t, \\ I(t) &= \text{number of infectives at time } t, \\ R(t) &= \text{number of removals at time } t. \end{aligned}$$

Initiated, again, by Kermack and McKendrick in 1927 [11], the evolutionary dynamics of these individuals can be modeled through the following system of ordinary differential equations (ODEs):

$$\begin{cases} I'(t) = \beta S(t) I(t) - \gamma I(t), \\ S'(t) = -\beta S(t) I(t), \\ R'(t) = \gamma I(t). \end{cases} \tag{1}$$

Here, the following assumptions are considered.

(A1) The total population size is always $N > 0$, meaning that $S(t) + I(t) + R(t) = N$ for all t.

(A2) We know the infection rate $\beta > 0$ from the infection process, and the removal rate $\gamma > 0$ from the removal process.

* anhkhoa.vo@ttu.edu, vakhoa.hcmus@gmail.com

(A3) The initial conditions are $S(0) = n > 1$, $I(0) = a \geq 1$ and $R(0) = 0$.

The explicit solution of the SIR model, despite its basic structure, is widely known to be unattainable due to the exponential nonlinearity of the transcendental equation governing removals. Consequently, numerous numerical methods have been proposed to address this fundamental model. The Taylor expansion method, initially employed by Kermack and McKendrick in 1927, approximates the exponential term, leading to an approximate analytic solution. This technique was utilized to simulate the plague epidemic of 1905-1906 in Bombay, India, and has since become a tutorial resource for students at both undergraduate and graduate levels; cf. [4].

B. Historical remarks

Over the years, many different methods have been studied to solve the SIR model. Piyawong et al. [23] introduced an unconditionally convergent scheme that captures the long-term behavior of the SIR model, offering improved numerical stability compared to conventional explicit finite difference methods. Mickens [20] and, recently, Conte et al. [6] proposed and analyzed stable nonstandard finite difference methods that effectively preserve the positivity of the SIR solutions. Semi-analytical methods, such as the Adomian decomposition approach [19], have also been proposed, alongside other methods cited therein, to derive approximate analytical solutions. Additionally, the solutions of the SIR model can be expressed in terms of the Lambert function, as demonstrated in the publications by [2, 24]. Furthermore, an alternative approach involving parametric analysis has been employed to obtain analytical solutions; see e.g. [9]. While most of these approaches are local approximations, a recent global semi-analytical approach utilizing the Padé approximation has been presented in [5].

C. A glance of our relaxation scheme

While the above-mentioned approximation methods have demonstrated numerical effectiveness, their convergence theories have received limited investigation. Some publications discussing discrete methods focus solely on stability analysis, leaving the global convergence analysis unexplored. In this study, we propose a novel numerical approach that guarantees global convergence. Our approach employs a relaxation procedure, derived from the conventional linearization technique, to approximate the SIR model in a continuous setting. Unlike the classical version, our modified procedure introduces a relaxation term. In the existing literature on partial differential equations (e.g., [14, 15, 21, 31] and references cited therein), this relaxation term mitigates the local convergence issues encountered by conventional linearization techniques such as Newton's method. Consequently, it permits to choose an arbitrary starting point while guaranteeing the global convergence. Within our specific context, the relaxation term facilitates capturing the non-negativity of solutions while preserving global convergence. It is worth mentioning that these two features appear to be important in the designation of many numerical schemes; cf. e.g. [5, 6, 10, 14, 15, 17, 20, 21, 31, 32] for different problems taken into account.

By relying only on the dependence of the relaxation constant on the removal rate, our approach accurately captures the long-term behavior of the system. Furthermore, our explicit and easy-to-implement approximate scheme is governed by a sequence of linear differential equations. The desired approximate solution can be obtained discretely or analytically based on individual preference.

D. Organization of the paper

Our paper is four-fold. In section II, we begin by revisiting the transcendental equation for removals and discussing the essential properties of the SIR model. The latter part of this section focuses on introducing the proposed relaxation scheme and establishing its theoretical foundations. We prove that this scheme is globally strongly convergent and preserves non-negativity. Additionally, we derive an error estimate in C^0. In section III, we extend the applicability of our proposed scheme to some variants of the SIR model. Then, to validate the effectiveness of our method, numerical examples are presented in section IV. Finally, we provide some concluding remark in section V, and the appendix contains the proofs of our central theorems.

II. RELAXATION PROCEDURE

A. Transcendental equation revisited

It is well known that the SIR model can be solved from a transcendental differential equation. Here, we revisit how to get such an equation to complete our analysis of the proposed scheme below. Let $\mu = \beta/\gamma$ be the reciprocal relative removal rate. By the second and third equations of system (1), we have

$$\frac{dS}{dR} = -\mu S,$$

which leads to $\ln(S) = -\mu R + c$. Therefore, we arrive at

$$S(t) = e^{c-\mu R(t)}. \tag{2}$$

To find c, we set $t = 0$ in (2) and use (A3). Indeed, by $n = S(0) = e^{c-\mu R(0)} = e^c$, we get $c = \ln(n)$ and thus, deduce that

$$S(t) = ne^{-\mu R(t)}. \tag{3}$$

Combining this and (A1), we derive the following nonlinear differential equation for $R(t)$:

$$R'(t) = \gamma\left(N - ne^{-\mu R(t)} - R(t)\right). \tag{4}$$

Remark. To this end, the notation C^m is used to denote the space of functions with m continuous derivatives. For the particular C^0 space, it is the space of continuous functions on $[0, T]$ with the standard max norm.

Theorem 1. *The differential equation (4) admits a unique C^1 non-negative solution $R(t)$. Moreover, the existence and uniqueness in C^1 of positive $S(t)$ and $I(t)$ to the SIR system (1) follow.*

Proof. The positivity of $S(t)$ and $I(t)$ is guaranteed by the first and second equations of system (1), i.e.

$$S(t) = n\exp\left(-\beta\int_0^t I(s)\,ds\right),$$
$$I(t) = a\exp\left(\int_0^t (\beta S(s) - \gamma)\,ds\right).$$

Then, by the third equation of (1), the non-negativity of $R(t)$ follows.

Cf. [30, Theorem 3.2], since the right hand side of (4) is globally Lipschitzian, the equation admits a unique local C^1 solution. Moreover, in view of the fact that $\left|\gamma\left(N - ne^{-\mu R(t)} - R(t)\right)\right| \le \gamma|R(t)| + \gamma(N + n)$ for any $t \ge 0$, the obtained solution is global as a by-product of [30, Theorem 3.9]. Observe that the right hand side of (3) is decreasing in the argument of $R(t)$. Thereby, the existence and uniqueness of $S(t)$ follows. We also get the existence and uniqueness of $I(t)$ in view of the fact that the total population is conserved; cf. (A1).

Hence, we complete the proof of the theorem. □

Observe that if one can approximate $R(t)$ well in (4), then $S(t)$ and $I(t)$ will be well approximated via (3) and (1), respectively. Define $g(r) = \gamma n e^{-\mu r} + \gamma r$ for $r \in \mathbb{R}$. We can rewrite (4) as

$$R'(t) = \gamma N - g(R(t)).$$

Remark. By the first and second equations of the SIR system (1), we get

$$\frac{dI}{dS} = -1 + \frac{1}{\mu S}.$$

Therefore, using $S(0) = n$ and $I(0) = a$, we find that $I(t) - a = -S(t) + n + \frac{1}{\mu}\ln(S(t)) - \frac{1}{\mu}\ln(n)$. Equivalently, we deduce that

$$I(t) = \frac{1}{\mu}\ln(S(t)) - S(t) + a + n - \frac{1}{\mu}\ln(n).$$

Since function $f(S) = \frac{1}{\mu}\ln(S) - S$ for $S > 0$ attains its maximum at $S = \mu^{-1}$, we can estimate the so-called amplitude, which is the maximum value of I, in the following manner.

$$I_{\max} = -\frac{1}{\mu}\ln(\mu) - \frac{1}{\mu} + a + n - \frac{1}{\mu}\ln(n). \tag{5}$$

B. Derivation and analysis of the numerical scheme

Let $\{R_k\}_{k=0}^{\infty}$ be a time-dependent sequence satisfying, for $k = 1, 2, 3, \ldots$,

$$R_k'(t) + MR_k(t) = \gamma N - g(R_{k-1}(t)) + MR_{k-1}(t). \tag{6}$$

The sequence $\{R_k\}_{k=0}^{\infty}$ aims to approximate $R(t)$ in (4) in the sense that R_k will be close to R as $k \to \infty$ uniformly in time. The accompanying initial condition for equation (6) is $R_k(0) = 0$ for any $k \ge 0$. Since our approximate model performs as an iterative scheme, we need a starting point, $R_0(t)$. Here, we choose $R_0(t) = 0$ based on the initial condition of $R(t)$ (cf. (A3)), which is the best information given to the sought $R(t)$.

Also, in (6), we introduce a k-independent constant $M \ge \gamma > 0$ for the so-called relaxation process. This relaxation term plays a very important role. It allows us to prove the non-negativity of the relaxation scheme, while many numerical approaches, including the regular linearization method, do not have or cannot prove this feature. Herewith, the regular linearization method we meant is the scheme $\{R_k\}_{k=0}^{\infty}$ in (6) with either $M = 0$ or only the exponential term in $g(r)$ being linearized.

Formulated below is the theorem showing that the scheme $\{R_k\}_{k=0}^{\infty}$ preserves the non-negativity of the removals over the relaxation process.

Theorem 2. *The sequence $\{R_k\}_{k=0}^{\infty}$ is a non-negativity-preserving scheme. Moreover, it holds true that for all $M \ge \gamma$,*

$$0 \le g(R_k) \le \gamma n + MR_k \quad \textit{for any } k \ge 0 \textit{ and } t \ge 0.$$

Proof. We prove this theorem by induction. The statement holds true for $k = 1$. Indeed, since $R_0(t) = 0$, the equation for $R_1(t)$ reads as

$$R_1'(t) + MR_1(t) = \gamma N - \gamma n \geq 0.$$

Therefore, we get

$$\begin{aligned} R_1(t) &= \frac{\gamma(N-n)}{M}\left(1 - e^{-Mt}\right) \geq 0, \\ g(R_1) &= \gamma n e^{-\mu R_1(t)} + \gamma R_1(t) \leq \gamma n + MR_1(t). \end{aligned}$$

Next, assume that the statement holds true for $k = k_0$. We prove that it also holds true for $k = k_0 + 1$. By (6), we have

$$\begin{aligned} &R_{k_0+1}'(t) + MR_{k_0+1}(t) \\ &\quad = \gamma N - g(R_{k_0}(t)) + MR_{k_0}(t) \geq 0. \end{aligned}$$

Thus, we obtain $R_{k_0+1}(t) \geq e^{-Mt} R_{k_0+1}(0) \geq 0$. As a by product, we can estimate that

$$\begin{aligned} g(R_{k_0+1}) &= \gamma n e^{-\mu R_{k_0+1}(t)} + \gamma R_{k_0+1}(t) \\ &\leq \gamma n + MR_{k_0+1}(t). \end{aligned}$$

Hence, we complete the proof of the theorem. □

In the following, we formulate the strong convergence result for the scheme $\{R_k\}_{k=0}^{\infty}$. For ease of presentation, proof of this result is deliberately placed in the Appendix. It is worth mentioning that proof of the strong convergence of the scheme relies so much on the strict estimation of g'. Such an estimation can merely be obtained by the aid of the non-negativity of the scheme.

Theorem 3. *The sequence $\{R_k\}_{k=0}^{\infty}$ defined in (6) is strongly convergent in C^0 toward the true solution $R(t)$ to Equation (4). In particular, we can find a number $C = C(T, M, \gamma, n, \mu) > 0$ independent of k such that the following error estimate holds true:*

$$\max_{0 \leq t \leq T} |R_k(t) - R(t)|^2 \leq \frac{C^k}{k!} \max_{0 \leq t \leq T} |R(t)|^2.$$

Our theoretical finding below shows that when $n\mu < 1$, the scheme $\{R_k(t)\}_{k=0}^{\infty}$ converges faster than the case $n\mu \geq 1$. The proof of the following corollary is also found in the Appendix.

Corollary 4. *Assume that $n\mu < 1$. We can find a constant $c \in (0,1)$ independent of k such that the following error estimate holds true:*

$$\max_{0 \leq t \leq T} |R_k(t) - R(t)| \leq c^k \max_{0 \leq t \leq T} |R(t)|. \quad (7)$$

As readily expected, for every step k, we obtain a non-homogeneous differential equation that can be effectively approximated in the discrete framework. Mimicking the proof of Theorem 3 and using Theorem 2, we have

$$\begin{aligned} &R_k'(t) + MR_k(t) \\ &\quad \leq \gamma N - \gamma n + (M + \gamma n\mu - \gamma) R_{k-1}(t). \end{aligned} \quad (8)$$

Indeed, by (39), for $p(r) = g(r) - Mr$,

$$\begin{aligned} p(0) - p(R_{k-1}) &\leq |p(R_{k-1}) - p(0)| \leq |p'| |R_{k-1}| \\ &\leq (M + \gamma n\mu - \gamma) R_{k-1}, \end{aligned}$$

which leads to

$$-g(R_{k-1}) + MR_{k-1} \leq (M + \gamma n\mu - \gamma) R_{k-1} - \gamma n.$$

Thus, by (8) it follows that

$$\begin{aligned} &R_k(t) \\ &\leq e^{-Mt} \int_0^t e^{Ms} \left(\gamma a + (M + \gamma n\mu - \gamma) R_{k-1}(s)\right) ds \\ &\leq t\gamma a + (M + \gamma n\mu - \gamma) \int_0^t R_{k-1}(s)\, ds. \end{aligned}$$

By induction and by the choice $R_0(t) = 0$, we can show that

$$|R_k(t)| \leq \gamma a \sum_{i=1}^{k} (M + \gamma n\mu - \gamma)^{i-1} \frac{t^i}{i!}.$$

Therefore, if we choose $\gamma \leq M \leq \gamma + \frac{1}{T}$, then $|R_k(t)| \leq T\gamma a \left(e^{\gamma n\mu T} - 1\right)$. Note that this bound is independent of k. Thus, by Theorem 1, we get $R_k \in C^1$ for any k with, cf. (8), Theorem 2 and the choice $M \leq \gamma + \frac{1}{T}$,

$$\begin{aligned} |R_k'(t)| &\leq \gamma a + (M + \gamma n\mu - \gamma) T\gamma a \left(e^{\gamma n\mu T} - 1\right) \\ &\leq \gamma a \left(1 + (\gamma n\mu T + 1)\left(e^{\gamma n\mu T} - 1\right)\right) \\ &\leq \gamma a e^{2\gamma n\mu T}. \end{aligned}$$

Furthermore, by differentiating both sides of (6) with respect to time, we can demonstrate that $R_k \in C^2$ for any k. Indeed,

$$\begin{aligned} &R_k''(t) + MR_k'(t) \\ &\quad = \left(M - \gamma + \gamma n\mu e^{-\mu R_{k-1}(t)}\right) R_{k-1}'(t) \\ &\quad \leq (M - \gamma + \gamma n\mu) \left|R_{k-1}'(t)\right|. \end{aligned}$$

This yields that $|R_k''(t)| \leq \gamma a \left(\gamma n\mu + \frac{2}{T} + \gamma\right) e^{2\gamma n\mu T}$, which is a k-independent upper bound. This ensures the Euler method's global error by leveraging its existing convergence theory. For completeness, we present below the discrete solution to our proposed scheme.

Consider the time increment $\Delta t = T/P$ for $P \geq 2$ being a fixed integer. Then, we set the mesh-point in time by $t_p = p\Delta t$ for $0 \leq p \leq P$. We seek $R_k^p \approx R_k(t_p)$ as a discrete solution to equation (6). By the standard Euler method, R_k^p is determined by the following equation:

$$\begin{aligned} &R_k^p + \Delta t M R_k^p \\ &\quad = R_k^{p-1} + \Delta t \left(\gamma N - g\left(R_{k-1}^p\right) + M R_{k-1}^p\right). \end{aligned} \tag{9}$$

By this way, the global error of the Euler method is attained in the sense that for every k, there exists a constant $\tilde{C} > 0$ such that

$$\max_{0 \leq p \leq P} |R_k^p - R_k(t_p)| \leq \tilde{C} \Delta t. \tag{10}$$

We accentuate that by the above analysis of R_k, i.e. $R_k \in C^2$ for any k, the constant $\tilde{C}$ is independent of k. Thus, by Theorem 3, we can estimate the distance between the discrete (approximate) solution R_k^p and the true solution R at each mesh-point,

$$\max_{0 \leq p \leq P} |R_k^p - R(t_p)| \leq \tilde{C} \Delta t + \sqrt{\frac{C^k}{k!}} \max_{0 \leq t \leq T} |R(t)|. \tag{11}$$

Remark. We have the following remarks:

- After obtaining the approximator R_k^p for $R(t_p)$, we can compute $S(t_p)$ using (3). Then, the approximate solution for $I(t_p)$ can be determined using (A1), specifically $I(t_p) = N - S(t_p) - R(t_p)$.

- Both $\tilde{C}$ and C in (11) are independent of P and k. As a by product, our discrete relaxation scheme $\{R_k^p\}_{k=0}^{\infty}$, as defined in (9), is globally strongly convergent in C^0. Similar to the proof of Theorem 2, we can demonstrate that R_k^p is non-negativity-preserving. It is important to note that many previous approximations, such as the method of series expansions [5, 19], parametrization method [9, 25] and finite difference method [23, 29], did not adequately address the preservation of non-negativity/positivity. Furthermore, certain recent positive numerical schemes fail to provide an error bound, as observed in the publications [6, 13].

- While the application of the Euler method is initiated, we can further employ higher-order numerical methods to produce a faster convergent solver for the linear differential equation of R_k. Among these, the Runge-Kutta method stands out as the most favorable choice, offering a convergence rate of order $q \geq 2$. Building upon the analysis of the Euler method above, we can prove that all derivatives of the right hand side of (6) exist up to order q and $R_k \in C^q$ for any k. Therefore, we can show that R_k^p globally converges to R_k with a rate of $\mathcal{O}(\Delta t^q)$; cf. e.g. [8, Theorem 3.4] for the existing theory on the global convergence of the Runge-Kutta method, generalizing (10). Note here that $\mathcal{O}(x)$ is the conventional Landau symbol.

- Similar to (11), the convergence of the Runge-Kutta method remains unaffected by k. Nevertheless, it is important to emphasize that this convergence is heavily contingent upon the upper bounds of the involved derivatives. Considering the boundedness of R_k, R_k' and R_k'' established above, it becomes evident that these bounds tend to increase as the order rises. Consequently, it is crucial to exclusively employ variants of the Runge-Kutta method with appropriately high orders. This perspective holds true when applying the Runge-Kutta method directly to the differential equation (4).

III. EXTENSIONS TO OTHER SIR MODELS

In this section, we briefly discuss the applicability of the relaxation method to other population models of SIR type. In particular, we show below how the proposed approach can be adapted to approximate the SIRD (Susceptible-Infectious-Recovered-Deceased) and SIRX models; cf. [1, 3, 18] for an overview of these models.

SIRD model

The SIRD model extends the SIR model by distinguishing between recovered and deceased individuals. In this framework, the removals in the SIR model no longer encompass the number of infected individuals who have passed away. To account for mortality, a mortality rate $\sigma > 0$ is introduced, representing the rate at which infected individuals succumb to death. Consequently, the death rate per unit of time is calculated as the product of the mortality rate and the number of infected individuals. Additionally, as the number of deceased individuals is excluded from the removals, the rate of change of infections over time is adjusted to reflect the loss caused by mortality. Mathematically, the

SIRD model reads as

$$\begin{cases} I'(t) = \beta S(t) I(t) - (\gamma + \sigma) I(t), \\ S'(t) = -\beta S(t) I(t), \\ R'(t) = \gamma I(t), \\ D'(t) = \sigma I(t), \end{cases} \tag{12}$$

where $D(t)$ stands for the number of deceased people (after infection) at time t. Assume accordingly that initially, there is no deceased individual. Consequently, in (12), we apply the following assumptions, which are modified versions of **(A1)**, **(A2)**, and **(A3)**.

(B1) The total population size is always conserved with $N > 0$, meaning that $S(t) + I(t) + R(t) + D(t) = N$ for all t.

(B2) We know the infection rate $\beta > 0$ from the infection process, the removal rate $\gamma > 0$ from the removal process, and the death rate $\sigma > 0$ from the mortality process.

(B3) The initial conditions are $S(0) = n > 1$, $I(0) = a \geq 1$, $R(0) = 0$ and $D(0) = 0$.

Remark. By the first and second equations of the SIRD system (12), we see that

$$\frac{dI}{dS} = -1 + \frac{\gamma + \sigma}{\beta S}.$$

Similar to the classic SIR model (1), we can thus formulate the so-called amplitude in the following fashion:

$$I_{\max} = \frac{\gamma + \sigma}{\beta} \ln\left(\frac{\gamma + \sigma}{\beta}\right) - \frac{\gamma + \sigma}{\beta} + a + n - \frac{\gamma + \sigma}{\beta} \ln(n), \tag{13}$$

when S reaches $\frac{\gamma+\sigma}{\beta}$.

Remark. The SIRD model (12) resembles the SIRX model without containment rate. In the SIRX model, an additional class called "X" was introduced to account for the impact of social or individual behavioral changes during quarantine. Individuals in this class, referred to as symptomatic quarantined individuals, no longer contribute to the transmission of the infection. Instead of σ, the SIRX model without containment rate consideres $\kappa > 0$ that represents the rate at which infected individuals are removed due to quarantine measures. The SIRX model with the containment rate is not the scope of our paper since the associated transcendental system does not take the same form of (6). Indeed, the transcendental system governing the full SIRX model is of an integro-differential equation.

Now, we detail the transcendental equation for $R(t)$ and the application of the relaxation scheme. From the second and third equations of system (12), we deduce that

$$S(t) = ne^{-\mu R(t)}, \tag{14}$$

where we have recalled the reciprocal relative removal constant $\mu = \beta/\gamma$. Using the same way, the third and last equations of system (12) give

$$D(t) = \frac{\sigma}{\gamma} R(t), \tag{15}$$

by virtue of $R(0) = D(0) = 0$ (cf. (B3)). Then, plugging (14), (15) and (B1) into the third equation of (12), we obtain the following differential equation for $R(t)$:

$$R'(t) = \gamma \left[N - ne^{-\mu R(t)} - \left(1 + \frac{\sigma}{\gamma}\right) R(t)\right]. \tag{16}$$

Henceforth, our relaxation scheme in this case becomes

$$R_k'(t) + \overline{M} R_k(t) = \gamma N - \overline{g}(R_{k-1}(t)) + \overline{M} R_{k-1}(t), \tag{17}$$

where $\overline{g}(r) = \gamma n e^{-\mu r} + (\gamma + \sigma) r$. Similar to the SIR model, here we rely on (B3) to choose $R_k(0) = 0$ for any $k \geq 0$ as the initial condition and $R_0(t) = 0$ as the starting point.

By choosing $\overline{M} \geq \gamma + \sigma$, our sequence $\{R_k\}_{k=0}^{\infty}$ (defined in (17)) is non-negativity-preserving and globally strongly convergent to R of the transcendental equation (16). These findings are analogous to our central Theorems 2 and 3, and therefore, we omit their formulations. Besides, Theorem 1 is applied to (16), guaranteeing the global existence and uniqueness of the C^1 solutions to the SIRD model (12). Indeed, the non-negativity of the solutions to (12) is obtained, following the fact that

$$S(t) = n \exp\left(-\beta \int_0^t I(s)\, ds\right),$$
$$I(t) = a \exp\left(\int_0^t (\beta S(s) - \gamma - \sigma)\, ds\right).$$

Moreover, the global existence and uniqueness of these solutions are guaranteed because the right-hand side of (16) is globally Lipschitzian, and it satisfies that $\left|\gamma N - \gamma n e^{-\mu R(t)} - (\gamma + \sigma) R(t)\right| \leq (\gamma + \sigma) |R(t)| + \gamma (N + n)$ for any $t \geq 0$.

SIR model with background mortality

The SIR model, along with its variants SIRD and SIRX, assumes a constant population size. These models, known as epidemiological SIR-type models

without vital dynamics, are limited in their representation of population changes; see (A1) and (B1). The SIR model with vital dynamics addresses this limitation by incorporating birth and death rates to account for population size fluctuations.

In the present work, we explore that the transcendental system governing the SIR model with background mortality takes the form of (6). With σ being the death rate, the population experiences changes over time. Here, individuals from all compartments can exit through deaths, allowing for a more realistic representation of population dynamics. Mathematically, the SIR model with background mortality can be expressed as follows:

$$\begin{cases} I'(t) = \beta S(t) I(t) - \gamma I(t) - \sigma I(t), \\ S'(t) = -\beta S(t) I(t) - \sigma S(t), \\ R'(t) = \gamma I(t) - \sigma R(t). \end{cases} \tag{18}$$

In this perspective, we make use of the following assumptions.

(C1) The total population size is dependent of t, i.e. $N = N(t) > 0$. It can be computed that $N(t) = S(t) + I(t) + R(t) = e^{-\sigma t} N_0$ for some fixed $N_0 > 0$.

(C2) We know the infection rate $\beta > 0$ from the infection process, the removal rate $\gamma > 0$ from the removal process, and the death rate $\sigma > 0$ from the mortality process.

(C3) The initial conditions are $S(0) = n > 1$, $I(0) = a \geq 1$ and $R(0) = 0$. This implies that $N_0 = n + a$.

Similar to the classical SIR model, we seek the transcendental equation for R prior to the application of our proposed numerical scheme. When doing so, we define $\overline{R}(t) = e^{\sigma t} R(t)$ and $\overline{S}(t) = e^{\sigma t} S(t)$. By the second and third equations of system (18), we find that

$$\overline{S}'(t) = -\beta \overline{S}(t) I(t), \tag{19}$$

$$\overline{R}'(t) = \gamma e^{\sigma t} I(t). \tag{20}$$

Therefore, we deduce that

$$\frac{d\overline{S}}{d\overline{R}} = -\mu e^{-\sigma t} \overline{S}, \tag{21}$$

or equivalently, $\ln\left(\overline{S}\right) = -\mu e^{-\sigma t} \overline{R} + \tilde{c}(t)$. Herewith, we have recalled the reciprocal relative removal rate $\mu = \beta/\gamma$. Henceforth, we have

$$\overline{S}(t) = e^{\tilde{c}(t) - \mu e^{-\sigma t} \overline{R}(t)}. \tag{22}$$

Since $\overline{S}(0) = n$ and $\overline{R}(0) = R(0) = 0$ by (D3), we find that $\tilde{c}(0) = \ln(n)$. Moreover, by taking the derivative in time of (22), we arrive at

$$\overline{S}'(t) = e^{\tilde{c}(t) - \mu e^{-\sigma t} \overline{R}(t)} \left[\tilde{c}'(t) - \mu e^{-\sigma t} \left(-\sigma + \overline{R}'(t)\right)\right]. \tag{23}$$

Dividing both sides of (23) by $\overline{R}'(t)$, we find that

$$\frac{\overline{S}'(t)}{\overline{R}'(t)} = \frac{e^{\tilde{c}(t) - \mu e^{-\sigma t} \overline{R}(t)} \left[\tilde{c}'(t) - \mu e^{-\sigma t} \left(-\sigma + \overline{R}'(t)\right)\right]}{\overline{R}'(t)}.$$

Then combining this with (21) and (22), we derive the following differential equation for $\tilde{c}(t)$:

$$e^{\tilde{c}(t) - \mu e^{-\sigma t} \overline{R}(t)} \left[\tilde{c}'(t) - \mu e^{-\sigma t} \left(-\sigma + \overline{R}'(t)\right)\right] = -\mu e^{-\sigma t} e^{\tilde{c}(t) - \mu e^{-\sigma t} \overline{R}(t)} \overline{R}'(t),$$

or equivalently, $\tilde{c}'(t) = -\mu\sigma e^{-\sigma t}$. Thus, we obtain $\tilde{c}(t) = \ln(n) + \mu\left(e^{-\sigma t} - 1\right)$ and

$$\overline{S}(t) = n e^{\mu\left(e^{-\sigma t} - 1\right)} e^{-\mu e^{-\sigma t} \overline{R}(t)}.$$

Together with the back-substitution $e^{-\sigma t} \overline{R}(t) = R(t)$, we thereby get $\overline{S}(t) = n e^{\mu\left(e^{-\sigma t} - 1\right)} e^{-\mu R(t)}$. Now, we note that by (C1) and (C3), $e^{\sigma t} I(t) = N_0 - \overline{S}(t) - \overline{R}(t)$ holds true for any t. Plugging this into (20) and using the fact that $\overline{S}(t) = n e^{\mu\left(e^{-\sigma t} - 1\right)} e^{-\mu R(t)} = n e^{\mu\left(e^{-\sigma t} - 1\right)} e^{-\mu e^{-\sigma t} \overline{R}(t)}$, we derive the transcendental equation for $\overline{R}$ as follows:

$$\overline{R}'(t) = \gamma \left[N_0 - n e^{\mu\left(e^{-\sigma t} - 1\right)} e^{-\mu e^{-\sigma t} \overline{R}(t)} - \overline{R}(t)\right]. \tag{24}$$

By setting $\hat{g}(r) = \gamma n e^{\mu\left(e^{-\sigma t} - 1\right)} e^{-\mu e^{-\sigma t} r} + \gamma r$, our relaxation scheme for the SIR model with background mortality is structured by

$$\overline{R}_k'(t) + \hat{M} \overline{R}_k(t) = \gamma N_0 - \hat{g}\left(\overline{R}_{k-1}(t)\right) + \hat{M} \overline{R}_{k-1}(t). \tag{25}$$

Similar to the above-mentioned SIR-based models, we choose $\overline{R}_k(0) = 0$ for any $k \geq 0$ as the initial condition and $\overline{R}_0(t) = 0$ as the starting point, based on the fact that $\overline{R}(0) = R(0) = 0$. Also, here we take $\hat{M} \geq \gamma$ to ensure the non-negativity preservation and global strong convergence of the sequence $\left\{\overline{R}_k\right\}_{k=0}^{\infty}$ (defined in (25)) to the sought $\overline{R}$ of the transcendental equation (24). As another analog of Theorems 2 and 3, we omit details of the formulations of the theoretical results for the sequence $\left\{\overline{R}_k\right\}_{k=0}^{\infty}$. It is also worth mentioning that Theorem 1 remains true in this case, providing the global existence and uniqueness of C^1 solutions to the SIR model with mortality (18). Indeed, the solutions to

(18) are non-negative, since it holds true that

$$\overline{S}(t) = n \exp\left(-\beta \int_0^t I(s)\, ds\right),$$
$$I(t) = a \exp\left(\int_0^t (\beta S(s) - \gamma - \sigma)\, ds\right),$$

where by the derivation of (19), we know that $S(t) = e^{-\sigma t}\overline{S}(t)$. Furthermore, the global existence and uniqueness results are obtained because the right-hand side of (24) is globally Lipschitzian, and it is bounded from above by $\gamma |R(t)| + \gamma (N_0 + n)$ for any $t \geq 0$.

IV. NUMERICAL EXPERIMENTS

In this section, we verify the numerical performance of the proposed relaxation method. Initially, we employ various approaches to solve the SIR model (1) for the purpose of comparison. These include our method (6), as well as the standard methods: approximate analytic solution, regular linearization procedure, and conventional explicit Euler method. It is important to note that since the conventional explicit Euler method is considered in this comparison, we also apply the Euler method to our relaxation scheme, as outlined in (9), as well as to the regular linearization procedure.

Additionally, it is worth mentioning that the approximate analytic solution for $R(t)$ (referred to as R_{a}) can be found in [3, 4, 11]. In particular, it is of the following form:

$$R_{\mathrm{a}}(t) = \frac{1}{n\mu^2}\left[n\mu - 1 + \eta \tanh\left(\frac{\gamma\eta t}{2} - \psi\right)\right], \quad (26)$$

where

$$\eta = \left[2n\mu^2 (N - n) + (n\mu - 1)^2\right]^{1/2},$$
$$\psi = \tanh^{-1}\left[\frac{1}{\eta}(n\mu - 1)\right].$$

The approximate analytic solution mentioned above corresponds to the solution of the Riccati equation. However, it is applicable only when μR is sufficiently small. Furthermore, the conventional explicit Euler method is expressed as follows:

$$R^p = R^{p-1} + \Delta t \left(\gamma N - g\left(R^{p-1}\right)\right), \quad (27)$$

where $R^p \approx R(t_p)$ is the discrete solution to the nonlinear differential equation (4) for $t_p = p\Delta t$ being the mesh-point in time.

In the second test, we utilize the widely used Runge-Kutta RK4 method to solve the SIR model (6) by applying it to our relaxation scheme (1). We then compare its performance when using the Euler-relaxation method (9) and when directly applying the Runge-Kutta RK4 method to (4). For sake of clarity, we provide the formulation of the RK4 method for solving a differential equation of a general type $R'(t) = F(t, R(t))$:

$$R^p = R^{p-1} + \frac{1}{6}K_1\left(t_{p-1}, R^{p-1}\right) + \frac{1}{3}K_2\left(t_{p-1}, R^{p-1}\right) + \frac{1}{3}K_3\left(t_{p-1}, R^{p-1}\right) + \frac{1}{6}K_4\left(t_{p-1}, R^{p-1}\right), \quad (28)$$

where we have denoted the intermediate values by

$$K_1\left(t_{p-1}, R^{p-1}\right) = \Delta t F\left(t_{p-1}, R^{p-1}\right), \quad (29)$$
$$K_2\left(t_{p-1}, R^{p-1}\right) = \Delta t F\left(t_{p-1} + \frac{\Delta t}{2}, R^{p-1} + \frac{K_1}{2}\right), \quad (30)$$
$$K_3\left(t_{p-1}, R^{p-1}\right) = \Delta t F\left(t_{p-1} + \frac{\Delta t}{2}, R^{p-1} + \frac{K_2}{2}\right), \quad (31)$$
$$K_4\left(t_{p-1}, R^{p-1}\right) = \Delta t F\left(t_{p-1} + \Delta t, R^{p-1} + K_3\right). \quad (32)$$

When using our method, it is important to note that for each iteration k, we solve the linear differential equation $R_k'(t) = F(t, R_k(t))$, where $F(t, R_k(t)) = -MR_k(t) + \gamma N - g(R_{k-1}(t)) + MR_{k-1}(t)$. Notice that in this equation, the presence of the midpoint $t_{p-1} + \frac{\Delta t}{2}$, applied to $R_{k-1}(t)$ obtained from the previous step, leads to the following linear approximation:

$$R_{k-1}\left(t_{p-1} + \frac{\Delta t}{2}\right) = \frac{1}{2}\left[R_{k-1}(t_{p-1}) + R_{k-1}(t_{p-1} + \Delta t)\right]. \quad (33)$$

Denote this approximation by $R_{k-1}^{p-0.5} = R_{k-1}\left(t_{p-1} + \frac{\Delta t}{2}\right)$. Thereby, we seek R_k^p satisfying (28) in which the

intermediate values are given by

$$K_1 = \Delta t \left(-M R_k^{p-1} + \gamma N - g\left(R_{k-1}^{p-1}\right) + M R_{k-1}^{p-1} \right), \tag{34}$$

$$K_2 = \Delta t \left[-M \left(R_k^{p-1} + \frac{K_1}{2} \right) + \gamma N - g\left(R_{k-1}^{p-0.5}\right) + M R_{k-1}^{p-0.5} \right], \tag{35}$$

$$K_3 = \Delta t \left[-M \left(R_k^{p-1} + \frac{K_2}{2} \right) + \gamma N - g\left(R_{k-1}^{p-0.5}\right) + M R_{k-1}^{p-0.5} \right], \tag{36}$$

$$K_4 = \Delta t \left[-M \left(R_k^{p-1} + K_3 \right) + \gamma N - g\left(R_{k-1}^{p}\right) + M R_{k-1}^{p} \right]. \tag{37}$$

On the other hand, when directly applying the Runge-Kutta RK4 method to the nonlinear differential equation (4), we have $F(t, R(t)) = \gamma\left(N - ne^{-\mu R(t)} - R(t)\right)$.

In the third test, we present the numerical performance of the proposed method in solving the SIR-based models discussed in section III. In particular, our focus in this test is on

1. the scheme $\{R_k\}_{k=0}^{\infty}$, defined in (17), for the SIRD model (12). In this model, the relaxation parameter satisfies $\overline{M} \geq \gamma + \sigma$.

2. the scheme $\{R_k\}_{k=0}^{\infty}$ computed from $\left\{\overline{R}_k\right\}_{k=0}^{\infty}$ (defined in (25)) for the SIR model with background mortality (18). In this case, we condition that $\hat{M} \geq \gamma$.

To evaluate the accuracy of the relaxation scheme, we assess the proximity of the approximation when approaching the maximum value of I. It is important to recall that explicit expressions for $I_{\max}$ have been derived for each specific case. The reader is referred to (5) for the classical SIRD model, and (13) for the SIRD model. For the SIR model with background mortality, since the maximum value of I cannot be found explicitly, we run the simulation with several values of P and K to verify the numerical stability. When increasing these parameters, we also identify the numerical amplitude and peak day to see the performance of our relaxation method in the Euler and RK4 frameworks.

Method	
#1	(6) coupled with the Euler method
#2	(6) with $M = 0$
#3	Analytic solution (26)
#4	Conventional Euler method (27)
#5	(6) coupled with the RK4 method
#6	Conventional RK4 method (28)–(37)

Table I: Numerical methods examined in section IV.

Method #	1	1	2	2	3	4	4
# of time step P	100	1000	100	1000	None	100	1000
# of iteration K	5	50	2	4	None	None	None
Amplitude $I_{\max}$	797	800	800	800	755	793	800
Peak day	25	23	138	35	114	32	25

Table II: Values of the computed amplitude $I_{\max}$ obtained from different methods and the corresponding peak days. Method #1: our Euler-relaxation method (9). Method #2: the regular linearization method, i.e. our proposed method (6) but with $M = 0$. Method #3: the approximate analytic solution R_{a} formulated in (26). Method #4: the conventional explicit Euler method (27) applied directly to the nonlinear differential equation (4). By (5), the true amplitude $I_{\max,\mathrm{true}}$ is 800 in this scenario.

Test 1

In this test, we compare our Euler-relaxation approach with the approximate analytic solution (26), the regular linearization procedure (which arises when the relaxation parameter vanishes), and the direct explicit Euler method (27). Alongside assessing numerical stability, we evaluate the performance of these methods based on the amplitude $I_{\max}$ presented in (5) and the peak day.

Here, we consider a population sample of $N = 1000$ for the SIR model (1) over the course of one year ($T = 365$). Initially, we assume that there are $a = 2$ infected people in this sample, leaving $n = 998$ individuals susceptible to infection. Furthermore, we choose a removal rate of $\gamma = 0.02$ and an infection rate of $\beta = 0.0004$. With these choices, we obtain a reciprocal relative removal rate of $\mu = \beta/\gamma = 0.02$, indicating that $n\mu = 19.96 > 1$. Additionally, for our relaxation process, we set $M = 0.02$.

Our numerical results for Test 1 are presented

in Table II. Based on the maximum amplitude ($I_{\max}$), our proposed method within the Euler context (method #1) outperforms the approximations obtained from methods #2–4. The first two columns of Table II demonstrate the numerical stability of our proposed method, particularly when dealing with relatively small values of P and K. Remarkably, when $P = 100$ and $K = 5$, our method yields an $I_{\max}$ value of 797, which is very close to the true value of 800 as shown in (5). In contrast, the $I_{\max}$ obtained from the approximate analytic solution (method #3) shows a significant deviation. We also observe that the amplitude $I_{\max}$ obtained from method #3 remains unaffected regardless of the choice of P.

A comparison between methods #1 and #2 reveals that while the regular linearization technique can provide a satisfactory estimate of $I_{\max}$ (800 when considering $P = 100$ and $K = 2$), method #2 suffers from severe numerical instability as illustrated in the second row of Figure 1, particularly when increasing K to obtain a more accurate graphical representation. Note that to help visualize the instability better, we deliberately use a log-scale in the vertical axis.

Furthermore, when P and K are relatively small, our proposed method shows a slight improvement over the conventional Euler method (method #4) within the same Euler context. At a coarse grid level, method #4 yields a relative error of 0.875%, while our method achieves a lower relative error of 0.375%. Upon increasing P to 1000, both methods #1 (with an increased $K = 50$) and #4 demonstrate comparable accuracy in terms of amplitude and graphical representation, as depicted in the first and last rows of Figure 1.

Our numerical investigation reveals that the true peak value ($I_{\max,\text{true}}$) is attained on the 24th day by employing sufficiently large values of P (over 3000) in both reliable methods #1 and #4. Comparing the peak days, it becomes evident from the last row of Table II that our relaxation method outperforms methods #2 and #3. While our method and method #4 achieve similar accuracy in terms of graphical simulation and amplitude, our proposed method detects the peak day earlier and with greater reliability. Specifically, considering small P and K, our relaxation method identifies the peak outbreak on the 25th day, which closely aligns with the true peak (24th), in contrast to the peak day of 32nd obtained from method #4. For larger P, our method predicts an earlier peak occurrence (day 23rd), which proves advantageous in practical scenarios compared to the peak day of 25th obtained from method #4. The ability to predict the peak event of a disease earlier is of practical significance for decision-makers, enabling them to implement and sustain timely public health measures and interventions aimed at mitigating the disease risk.

Test 2

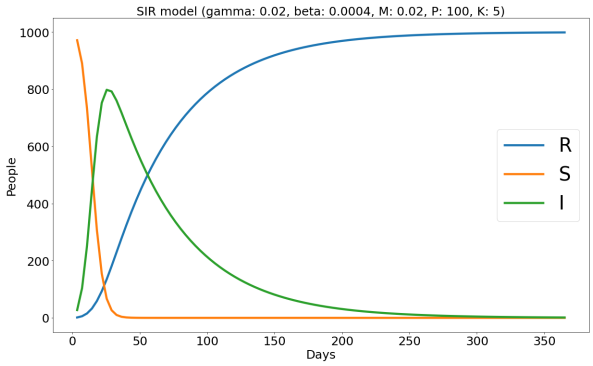
Our second test focuses on the numerical comparison between two approaches: applying the well-known Runge-Kutta RK4 method (referred to as method #5) to our relaxation scheme (6) and applying it directly to the nonlinear differential equation (4) (denoted as method #6). Additionally, we compare the convergence speed of method #5 with method #1, referred above to as the Euler-relaxation method (9).

As RK4 is a fourth-order method, we deliberately choose a large population size of $N = 97.47 \times 10^6$ and a transmission rate of $\beta = 3 \times 10^{-9}$. Assuming the initial infected population is $a = 11$, and the removal rate remains constant at $\gamma = 0.05$ throughout the entire six-month period ($T = 180$), we can calculate that the simulated disease reaches its peak at $I_{\max,\text{true}} = 51367769$; cf. (5). Moreover, based on numerical observations with a sufficiently large value of P (>2000), we find that this peak is reached on the 73rd day.

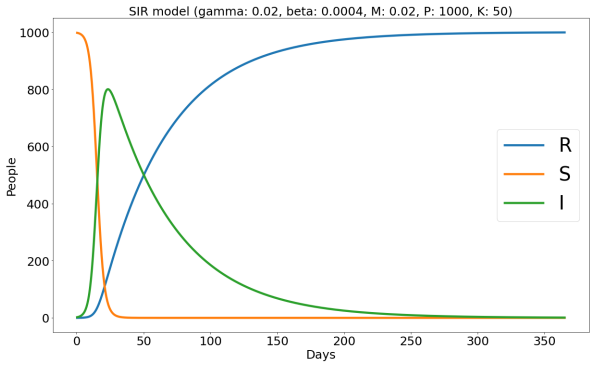
Our numerical results are tabulated in Table III, accompanied by corresponding graphical illustrations in Figure 2. We see that within the same RK4 framework, our proposed relaxation method (method #5) outperforms the direct approach. When the number of time steps is small ($P = 50$), method #5 with $K = 20$ yields an amplitude $I_{\max}$ of 51295165 with a relative accuracy of 0.14%, while method #6 achieves 0.81%. Both methods capture the peak day (72) well compared to the true value of 73. Note in this case that we choose $K = 20$, a larger value than in Test 1, due to the larger population under consideration. Cf. Theorem 3, the choice of K does affect our error estimation which heavily depends on the total size of the removal population.

We also see that when increasing P to 2000, our proposed method #5 with an increased $K = 50$ precisely achieves the true amplitude, $I_{\max,\text{true}} = 51367769$, while the direct RK4 method produces a very close approximation of 51367765. Both methods also identify the peak day as the 73rd day.

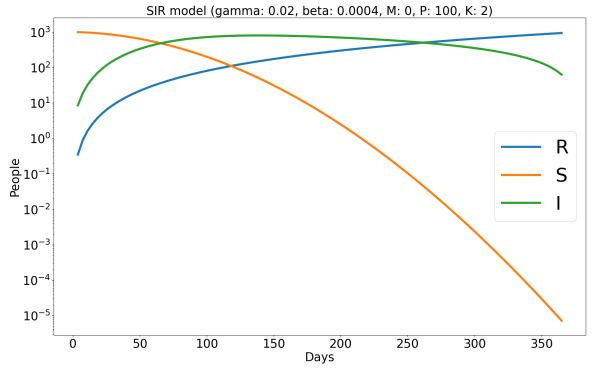
Furthermore, we compare our relaxation method to the Euler and Runge-Kutta frameworks. In terms of amplitude, although method #1 initially provides a better value of 51341234 with an accuracy



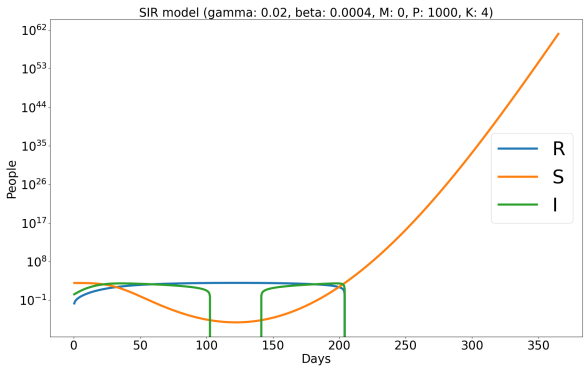
(a) Method #1 ($P = 100, K = 5$)



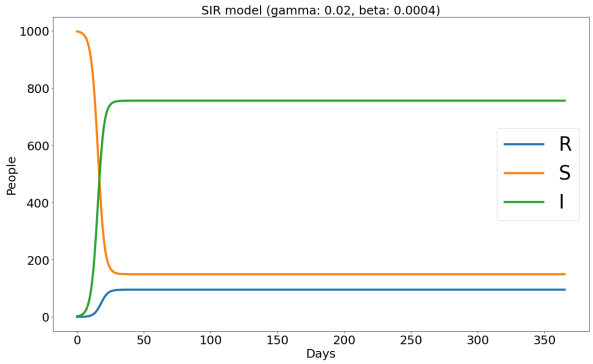
(b) Method #1 ($P = 1000, K = 50$)



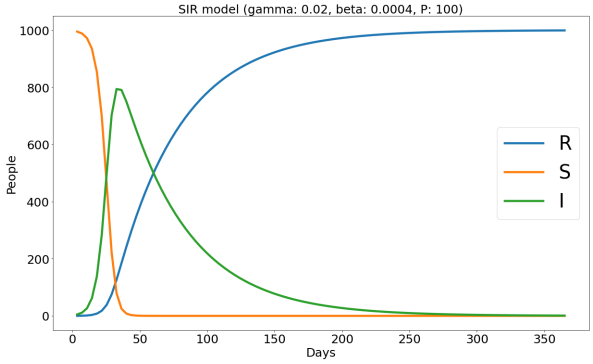
(c) Method #2 ($P = 100, K = 2$)



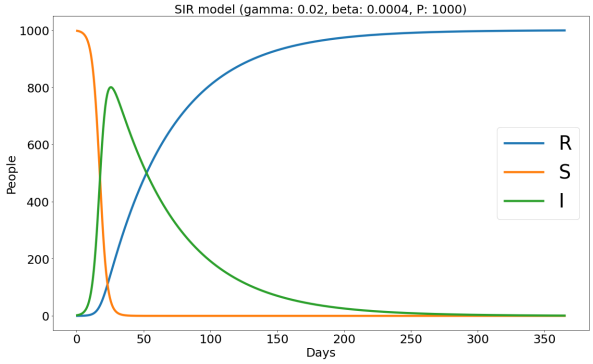
(d) Method #2 ($P = 1000, K = 4$)



(e) Method #3



(f) Method #4 ($P = 100$)



(g) Method #4 ($P = 1000$)

Figure 1: Graphical illustrations of Test 1. Row 1: Euler-relaxation method. Row 2: regular linearization method. Row 3: approximate analytic method. Row 4: direct Euler method. Note that to help visualize the instability better, a log-scale in the vertical axis is used in Row 2.

of 0.05%, it fails to accurately detect the peak day, significantly deviating from the true value of 73 (predicting 54 instead). Increasing P to 2000 improves the amplitude to 51367573, but it still performs worse than the direct RK4 method's amplitude of 51367765. Herewith, method #1 achieves an improved peak day of 72. Additionally, based on the simulation of method #1, we observe that to reach the true amplitude ($I_{\max,\text{true}} = 51367769$) and the true peak day of 73, at least $P = 19000$ and $K = 100$ are required. Henceforth, our relaxation method in the RK4 framework, as readily expected, outperforms itself in the Euler framework.

Remark. It is essential to select $M = \gamma$ as the optimal choice, as this helps to minimize the left-hand side of (42). To verify this, we run method #5 using a coarse grid of $P = 50, K = 20$ for different values of $M \in \{0.1, 0.2, 0.5\}$. We observe that when M runs far away from $\gamma = 0.05$, the corresponding amplitude $I_{\max}$ becomes less accurate; compared to our numerical results tabulated in Table III. We particularly report that the amplitude $I_{\max}$ drops from 51245377 with $M = 0.1$ to 41107779 with $M = 0.2$, and further to 29834 with $M = 0.5$.

Since the RK4 framework converges very quickly, for a fine grid with $P = 2000, K = 50$, increasing M to 0.5 does not affect the value of the amplitude $I_{\max}$. However, this increase in M shows an adverse impact on identifying the peak day. For $M = 0.5$, the amplitude remains accurate at 51367768, close to the true value of 51367769. Yet, the peak day shifts to day 78th, which is 5 days later than the actual peak day (73rd). When M is increased to 1–20 times larger than the original choice of 0.05–the approximation to the true value deteriorates significantly. In particular, for $M = 1$, the amplitude is reduced to 564059, and the peak day shifts to day 122nd.

Test 3

As previously mentioned, in our last experiment, we aim to broaden the scope of the proposed relaxation method by applying it to various SIR-type models: specifically, the SIRD model and the SIR model with background mortality. These models share the same input parameters as those used in Test 1, where we set $N = 1000$, $n = 998$, $T = 365$, $\gamma = 0.02$, $\beta = 0.0004$. In the SIRD model (12), we choose a death rate of $\sigma = 0.01$, which implies a choice of the relaxation parameter $\overline{M} \geq \gamma + \sigma = 0.03$. In the SIR model with background mortality (18), we use the background death of $\sigma = 0.001$ and select $\hat{M} \geq \gamma = 0.02$.

Our numerical findings for the SIRD model are detailed in Figure 3. We specifically investigate the scenario where $\overline{M} = 0.015$, thereby contravening the relaxation condition ($\overline{M} \geq 0.03$). Consistent with our theorem concerning non-negativity preservation, we observe that the relaxed solution with $\overline{M} = 0.015$, obtained from both the Euler and RK4 frameworks, fails to maintain non-negativity over time. This is evident in the first column of Figure 3. When $\overline{M} = 0.03$ (a case that adheres to the condition), we also note that the RK4-relaxation outperforms the Euler-relaxation approach. Given the $I_{\max,\text{true}} = 730$ (as formulated in (13)), and with a coarse mesh of $P = 200$ and $K = 10$, we observe that the RK4-relaxation produces an amplitude identical to the true value. In contrast, the Euler-relaxation yields a value of 729. It is also worth mentioning that the accurate amplitude is attained when applying the Euler-relaxation with $P = 800$ and $K = 10$.

Our numerical results pertaining to the SIR model with background mortality are presented in Figure 4. It is evident that both the Euler and RK4 relaxation methods show numerical stability and non-negativity preservation as we increase the values of (P, K) from $(100, 5)$ to $(1000, 50)$.

Consistent with our prior tests, the RK4-relaxation method continues to outperform the Euler-relaxation method. Leveraging this numerical stability, we run the RK4-relaxation method using large values of P and K to determine the numerical amplitude and peak day. Our findings reveal a numerical amplitude of 777, peaking on the 24th day. Within the RK4 framework, achieving this numerical amplitude and peak day requires approximately $P = 300$ and $K = 20$. In contrast, the Euler framework demands a minimum of $P = 1700$ and $K = 20$ for similar outcomes.

Method #	5	5	6	6	1	1
# of time step P	50	2000	50	2000	50	2000
# of iteration K	20	50	None	None	20	50
Amplitude $I_{\max}$	51295165	51367769	50948480	51367765	51341234	51367573
Peak day	72	73	72	73	54	72

Table III: Values of the computed amplitude $I_{\max}$ obtained from different methods and the corresponding peak days. Method #5: our RK4-relaxation method (28) applied with (33)–(37). Method #6: the conventional RK4 method (28)–(32) applied directly to the nonlinear differential equation (4). Method #1: our Euler-relaxation method (9). By (5), the true amplitude $I_{\max,\text{true}}$ is 51367769 in this scenario.

V. CONCLUDING REMARKS

This work presents a novel numerical approach for solving the SIR model in population dynamics. While various approximation methods have been proposed for this classical model, the analysis of their convergence has been limited and challenging. Our approach introduces the relaxation procedure to approximate the continuous model. By carefully selecting the relaxation parameter, we achieve global strong convergence of the scheme and effectively preserve non-negativity. The proposed scheme is explicit and straightforward to implement, enabling us to obtain the approximate solution at either the discrete or analytical level. Additionally, we showcase the applicability of our scheme to numerous variants of the SIR model.

In our future work, we will develop a globally strongly convergent higher-order scheme based on the current relaxation method. Additionally, we plan to apply this method to more complex SIR-based models, involving multiple compartments and dimensions, as explored in recent works such as [22, 26, 27] and references therein. Furthermore, we will attempt to integrate the method with other techniques that have been designed to solve fractional systems, as studied in [33–35] and dissipative systems, as presented in [28].

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

V. A. Khoa: Supervision, Conceptualization, Formal analysis, Writing–review & editing. **P. M. Quan:** Software, Visualization. **J. Allen:** Formal analysis, Visualization. **K. W. Blayneh:** Formal analysis, Writing–review & editing.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

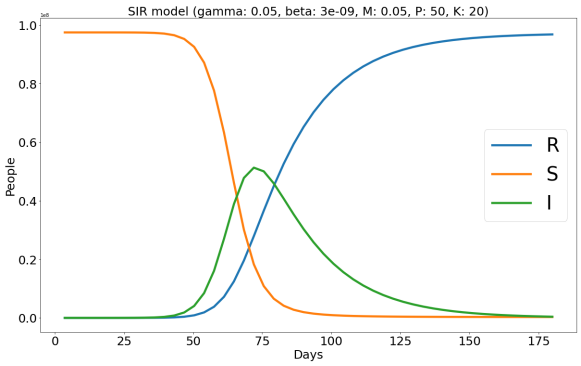
Simulated data will be made available on request. However, we have open-sourced our implementation of all the numerical schemes addressed in this paper in the following link: https://github.com/mquan1409/SIR_Relaxation.

ACKNOWLEDGMENTS

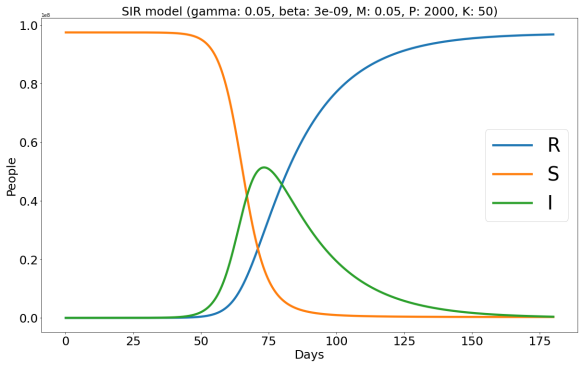
This research has received funding from the National Science Foundation (NSF) - DMS2451193. Specifically, V. A. K., P. M. Q., and J. A. extend their gratitude for the invaluable support provided by NSF. V. A. K. and J. A. also hold deep appreciation for the Florida A&M University Rattler Research program, its esteemed committee, and Dr. Tiffany W. Ardley. Their unwavering dedication has facilitated an exceptional academic journey for the mentee (J. A.) and the mentor (V. A. K.).

Furthermore, V. A. K. wishes to express many thanks to Dr. Charles Weatherford (Florida A&M University) and Dr. Ziad Musslimani (Florida State University). Their support has been instrumental in shaping V. A. K.'s early research career. Lastly, this work was complete on a momentous personal milestone – the wedding day of V. A. K. and the bride, Huynh Thi Kim Ngan.

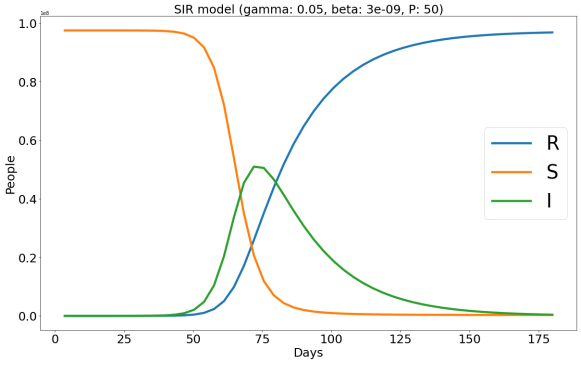
The authors would like to thank the anonymous reviewers for their valuable comments and suggestions, which significantly improved the quality of this manuscript.



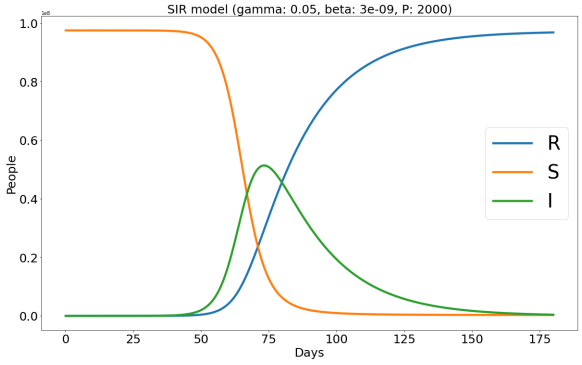
(a) Method #5 ($P = 50, K = 20$)



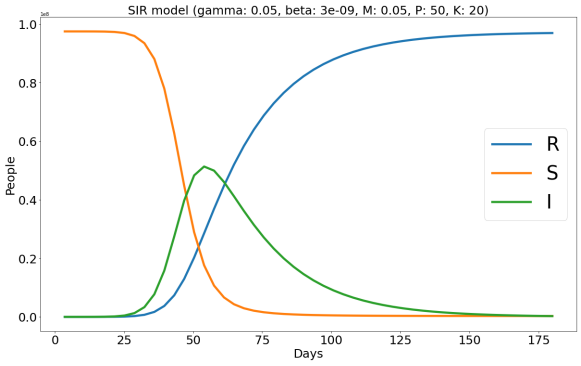
(b) Method #5 ($P = 2000, K = 50$)



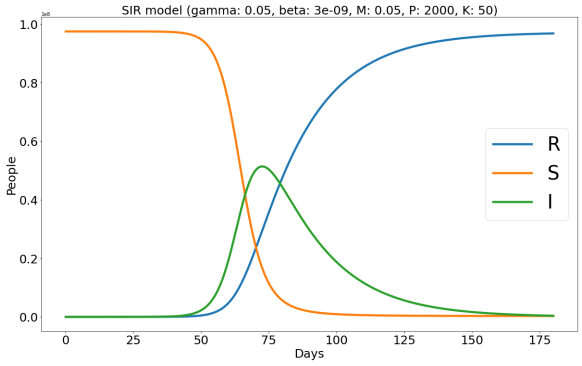
(c) Method #6 ($P = 50$)



(d) Method #6 ($P = 2000$)



(e) Method #1 ($P = 50, K = 20$)



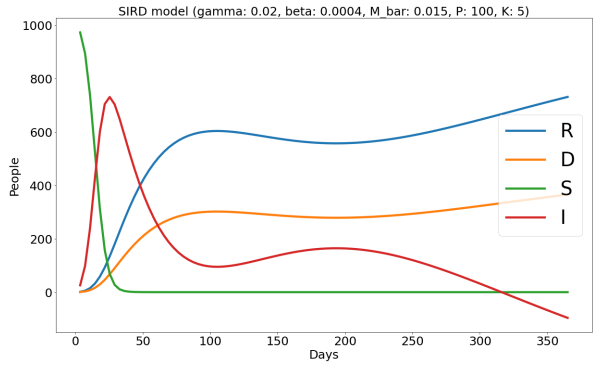
(f) Method #1 ($P = 2000, K = 50$)

Figure 2: Graphical illustrations of Test 2. Row 1: RK4-relaxation method. Row 2: direct RK4 method. Row 3: Euler-relaxation method.

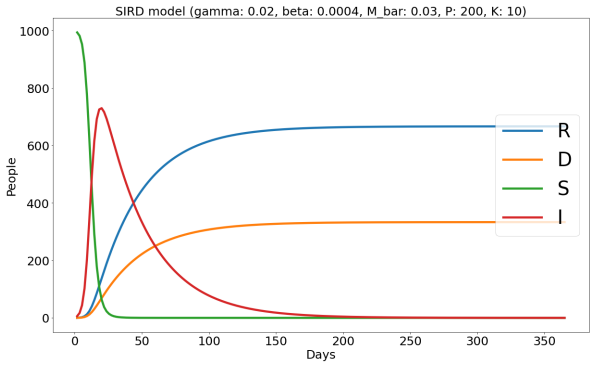
APPENDIX

Proof of Theorem 3

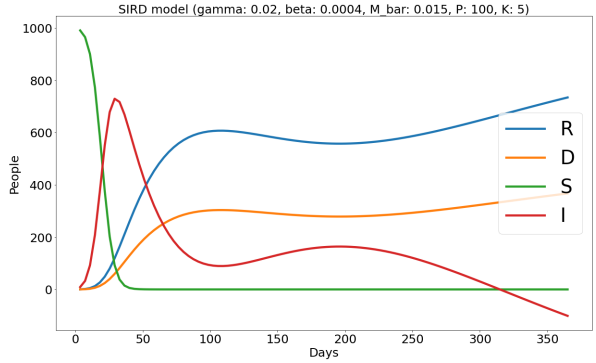
Step 1: Define $\mathcal{E}_k(t) = R_k(t) - R(t)$ for $k = 1, 2, 3, \ldots$. It follows from (6) and (4) that $\mathcal{E}_k$ satisfies the following differential equation:



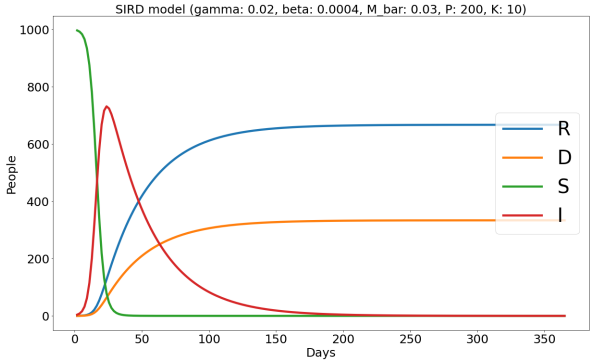
(a) Violating Euler-relaxation ($P = 100, K = 5, \overline{M} = 0.015$)



(b) Euler-relaxation ($P = 200, K = 10, \overline{M} = 0.03$)



(c) Violating RK4-relaxation ($P = 100, K = 5, \overline{M} = 0.015$)



(d) RK4-relaxation ($P = 200, K = 10, \overline{M} = 0.03$)

Figure 3: Graphical illustrations of Test 3 – SIRD model. Row 1: Euler-relaxation method with violating and non-violating cases. Row 2: RK4-relaxation method with violating and non-violating cases. These illustrations serve to highlight a crucial observation: when the relaxation parameter is not suitably selected, the numerical solution loses its adherence to non-negativity preservation.

$$\begin{aligned}\mathcal{E}_k'(t) + M\mathcal{E}_k(t) &= -g(R_{k-1}(t)) + g(R(t)) + M\mathcal{E}_{k-1}(t)\\ &= p(R(t)) - p(R_{k-1}(t)),\end{aligned} \tag{38}$$

where we have denoted $p(r) = g(r) - Mr$ for $r \geq 0$. Herewith, by Theorems 1 and 2, we are allowed to consider $r \geq 0$. We can compute that $p'(r) = -\gamma n\mu e^{-\mu r} + \gamma - M$. Then, for $M \geq \gamma$ and since $r \geq 0$, we estimate that

$$\gamma - \gamma n\mu - M \leq p'(r) = -\gamma n\mu e^{-\mu r} + \gamma - M < 0,$$

which shows

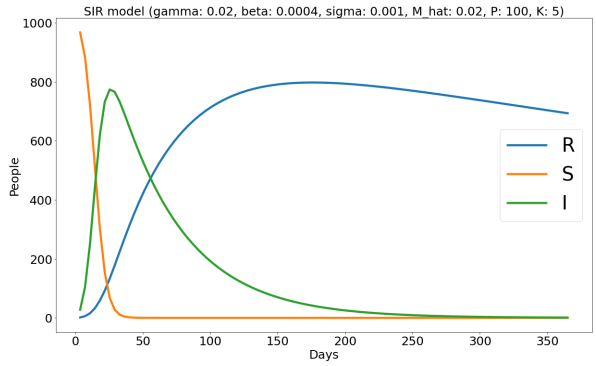
$$|p'(r)| \leq M + \gamma n\mu - \gamma. \tag{39}$$

Therefore, the left-hand side of (38) can be bounded from above by

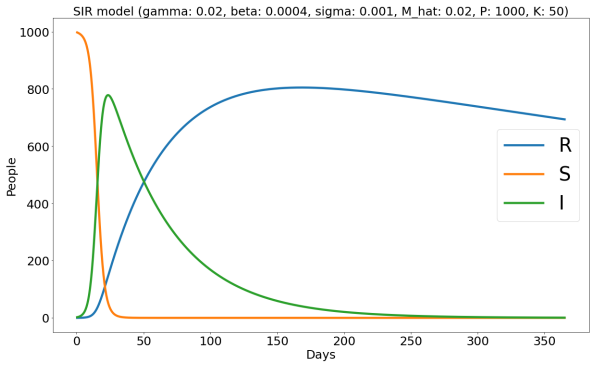
$$\mathcal{E}_k'(t) + M\mathcal{E}_k(t) \leq (M + \gamma n\mu - \gamma)\,|\mathcal{E}_{k-1}(t)|.$$

Using the Hölder inequality, we find that

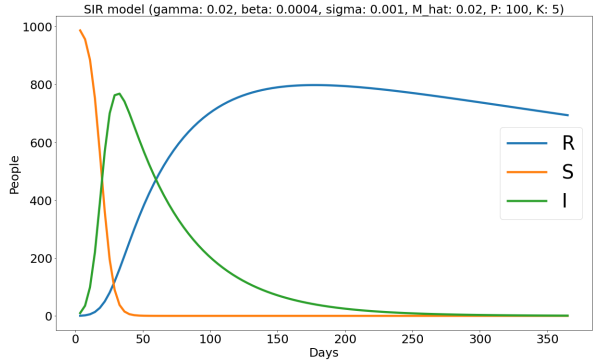
$$\begin{aligned}e^{2Mt}|\mathcal{E}_k(t)|^2 &\leq (M + \gamma n\mu - \gamma)^2 \left(\int_0^t e^{Ms}|\mathcal{E}_{k-1}(s)|\,ds\right)^2\\ &\leq (M + \gamma n\mu - \gamma)^2 \int_0^t e^{2Ms}ds \int_0^t |\mathcal{E}_{k-1}(s)|^2\,ds.\end{aligned}$$



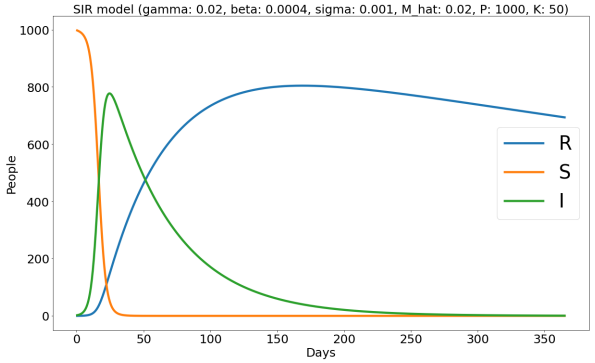
(a) Method #1 ($P = 100, K = 5$)



(b) Method #1 ($P = 1000, K = 50$)



(c) Method #5 ($P = 100, K = 5$)



(d) Method #5 ($P = 1000, K = 50$)

Figure 4: Graphical illustrations of Test 3 – SIR model with background death. Row 1: Euler-relaxation method with different values of P and K. Row 2: RK4-relaxation method with diverse P and K values. In these illustrations, numerical stability and non-negativity preservation are observed.

Thus, we deduce that

$$|\mathcal{E}_k(t)|^2 \leq \frac{1}{2M}(M + \gamma n\mu - \gamma)^2 \left(1 - e^{-2Mt}\right) \int_0^t |\mathcal{E}_{k-1}(s)|^2 \, ds.$$

By the elementary inequality $e^{-x} + x \geq 1$, we obtain the following estimate

$$|\mathcal{E}_k(t)|^2 \leq (M + \gamma n\mu - \gamma)^2 t \int_0^t |\mathcal{E}_{k-1}(s)|^2 \, ds. \tag{40}$$

Step 2: By induction, we can show that for any $2 \leq k \in \mathbb{N}$

$$|\mathcal{E}_k(t)|^2 \leq (M + \gamma n\mu - \gamma)^{2k} t \int_0^t s_1 \int_0^{s_1} \dots s_{k-1} \int_0^{s_{k-1}} |\mathcal{E}_0(s_k)|^2 \, ds_k ds_{k-1} \dots ds_1. \tag{41}$$

It follows from (40) that (41) holds true for $k = 2$. Indeed,

$$|\mathcal{E}_2(t)|^2 \leq (M + \gamma n\mu - \gamma)^2 t \int_0^t |\mathcal{E}_1(s_1)|^2 \, ds_1 \leq (M + \gamma n\mu - \gamma)^4 t \int_0^t s_1 \int_0^{s_1} |\mathcal{E}_0(s_2)|^2 \, ds_2 ds_1.$$

Assume that (41) holds true for $k = k_0$. We show that it also holds true for $k = k_0 + 1$. By (40), we have

$$\begin{aligned}
&\left|\mathcal{E}_{k_0+1}(t)\right|^2 \\
&\leq (M+\gamma n\mu-\gamma)^2 t \int_0^t \left|\mathcal{E}_{k_0}(s)\right|^2 ds \\
&\leq (M+\gamma n\mu-\gamma)^2 t \int_0^t (M+\gamma n\mu-\gamma)^{2k_0} s \int_0^s s_1 \int_0^{s_1} \dots s_{k_0-1} \int_0^{s_{k_0-1}} \left|\mathcal{E}_0(s_{k_0})\right|^2 ds_{k_0} ds_{k_0-1} \dots ds_1 ds \\
&\leq (M+\gamma n\mu-\gamma)^{2(k_0+1)} \int_0^t s_1 \int_0^{s_1} \dots s_{k_0} \int_0^{s_{k_0}} \left|\mathcal{E}_0(s_{k_0+1})\right|^2 ds_{k_0+1} ds_{k_0} \dots ds_1.
\end{aligned}$$

Hence, we complete Step 2.

Step 3: By (41), observe that $0 \leq s_k \leq s_{k-1} \leq \dots \leq s_1 \leq t$. Combining this, (41) and the fact that $R \in C^1$ gives

$$\begin{aligned}
\left|\mathcal{E}_k(t)\right|^2 &\leq (M+\gamma n\mu-\gamma)^{2k} t^{k+1} \max_{0\leq t\leq T} \left|\mathcal{E}_0(t)\right|^2 \int_0^t s_1 \int_0^{s_1} \dots s_{k-1} \int_0^{s_{k-1}} ds_k ds_{k-1} \dots ds_1 \\
&\leq (M+\gamma n\mu-\gamma)^{2k} \frac{t^{k+1}}{k!} \max_{0\leq t\leq T} \left|\mathcal{E}_0(t)\right|^2.
\end{aligned}$$

Note that we have the k and time independence of $M+\gamma n\mu-\gamma$ and $t \leq T$. Moreover, we know that $\mathcal{E}_0(t) = R_0(t) - R(t) = -R(t)$ by the choice $R_0(t) = 0$. Therefore, in view of the fact that $\lim_{k\to\infty} \frac{Q^k}{k!} = 0$ for any constant $Q > 0$, we can always find $\overline{k} > 0$ such that for any $k \geq \overline{k}$,

$$(M+\gamma n\mu-\gamma)^{2k} \frac{T^{k+1}}{k!} < 1. \tag{42}$$

Hence, we obtain the strong convergence of the sequence $\{R_k\}_{k=0}^{\infty}$ toward the true solution R.

Proof of Corollary 4

We define $\mathcal{E}_k(t) = R_k(t) - R(t)$ and $p(r) = g(r) - Mr$ as above. Multiplying (38) by $\mathcal{E}_k(t)$ and using (39) yield

$$\frac{1}{2}\frac{d}{dt}\mathcal{E}_k^2(t) + M\mathcal{E}_k^2(t) = \left[p(R(t)) - p(R_{k-1}(t))\right]\mathcal{E}_k(t) \leq \frac{M+\gamma n\mu-\gamma}{2}\mathcal{E}_{k-1}^2(t) + \frac{M+\gamma n\mu-\gamma}{2}\mathcal{E}_k^2(t).$$

Equivalently, we obtain

$$\frac{d}{dt}\mathcal{E}_k^2(t) + (M-\gamma n\mu+\gamma)\mathcal{E}_k^2(t) \leq (M+\gamma n\mu-\gamma)\mathcal{E}_{k-1}^2(t).$$

Notice that by the choice $M \geq \gamma$, it holds true that $M > \gamma n\mu - \gamma$ when $n\mu < 1$. Using the integrating factor $e^{(M-\gamma n\mu+\gamma)t}$ and taking integration with respect to t, we get

$$\begin{aligned}
\mathcal{E}_k^2(t) &\leq e^{-(M-\gamma n\mu+\gamma)t}(M+\gamma n\mu-\gamma) \int_0^t e^{(M-\gamma n\mu+\gamma)s}\mathcal{E}_{k-1}^2(s)\, ds \\
&\leq e^{-(M-\gamma n\mu+\gamma)t}\left[e^{(M-\gamma n\mu+\gamma)t} - 1\right]\frac{M+\gamma n\mu-\gamma}{M-\gamma n\mu+\gamma} \max_{0\leq t\leq T} \mathcal{E}_{k-1}^2(t).
\end{aligned}$$

Henceforth, we obtain

$$\max_{0\leq t\leq T} \left|\mathcal{E}_k(t)\right| \leq \left(\frac{M+\gamma n\mu-\gamma}{M-\gamma n\mu+\gamma}\right)^{1/2} \max_{0\leq t\leq T} \left|\mathcal{E}_{k-1}(t)\right|. \tag{43}$$

By induction and the fact that $R_0(t) = 0$, we deduce

$$\max_{0\le t\le T} |\mathcal{E}_k(t)| \le \left(\frac{M+\gamma n\mu-\gamma}{M-\gamma n\mu+\gamma}\right)^{k/2} \max_{0\le t\le T} |\mathcal{E}_0(t)| = \left(\frac{M+\gamma n\mu-\gamma}{M-\gamma n\mu+\gamma}\right)^{k/2} \max_{0\le t\le T} |R(t)|.$$

Since $M+\gamma n\mu-\gamma < M-\gamma n\mu+\gamma$ when $n\mu < 1$, we obtain the target estimate (7).

[1] N. T. J. Bailey. *The mathematical theory of infectious diseases and its applications*. Griffin, 1975.

[2] N. S. Barlow and S. J. Weinstein. Accurate closed-form solution of the SIR epidemic model. *Physica D: Nonlinear Phenomena*, 408:132540, 2020.

[3] G. Bärwolff. Modeling of COVID-19 propagation with compartment models. *Mathematische Semesterberichte*, 68(2):181–219, 2021.

[4] J. Caldwell and D. K. S. Ng. *Mathematical Modelling: Case Studies and Projects*. Springer, 2004.

[5] Y. Chakir. Global approximate solution of SIR epidemic model with constant vaccination strategy. *Chaos, Solitons & Fractals*, 169:113323, 2023.

[6] D. Conte, N. Guarinoa, G. Paganoa, and B. Paternoster. Positivity-preserving and elementary stable nonstandard method for a COVID-19 SIR model. *Dolomites Research Notes on Approximation*, 15:65–77, 2022.

[7] D. J. D. Earn, J. Dushoff, and S. A. Levin. Ecology and evolution of the flu. *Trends in Ecology & Evolution*, 17(7):334–340, 2002.

[8] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Springer Berlin Heidelberg, 2008.

[9] T. Harko, F. S. N. Lobo, and M. K. Mak. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation*, 236:184–194, 2014.

[10] M. T. Hoang. Positivity and boundedness preserving nonstandard finite difference schemes for solving Volterra's population growth model. *Mathematics and Computers in Simulation*, 199:359–373, 2022.

[11] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 1927.

[12] A. Khaleque and P. Sen. An empirical analysis of the Ebola outbreak in West Africa. *Scientific Reports*, 7(1), 2017.

[13] M. M. Khalsaraei, A. Shokri, H. Ramos, and S. Heydari. A positive and elementary stable nonstandard explicit scheme for a mathematical model of the influenza disease. *Mathematics and Computers in Simulation*, 182:397–410, 2021.

[14] V. A. Khoa, E. R. Ijioma, and N. N. Ngoc. Strong convergence of a linearization method for semi-linear elliptic equations with variable scaled production. *Computational and Applied Mathematics*, 39(4), 2020.

[15] V. A. Khoa, E. R. Ijioma, and N. N. Ngoc. Correction to: Strong convergence of a linearization method for semi-linear elliptic equations with variable scaled production. *Computational and Applied Mathematics*, 40(1), 2021.

[16] N. A. Kudryashov, M. A. Chmykhov, and M. Vigdorowitsch. Analytical features of the SIR model and their applications to COVID-19. *Applied Mathematical Modelling*, 90:466–473, 2021.

[17] H. Ma, Q. Zhang, and X. Xu. Positivity-preserving numerical method for a stochastic multi-group SIR epidemic model. *Computational Methods in Applied Mathematics*, 23(3):671–694, 2022.

[18] B. F. Maier and D. Brockmann. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*, 368(6492):742–746, 2020.

[19] O. D. Makinde. Adomian decomposition approach to a SIR epidemic model with constant vaccination strategy. *Applied Mathematics and Computation*, 184(2):842–848, 2007.

[20] R. E. Mickens. Numerical integration of population models satisfying conservation laws: NSFD methods. *Journal of Biological Dynamics*, 1(4):427–436, 2007.

[21] K. Mitra and I. S. Pop. A modified L-scheme to solve nonlinear diffusion problems. *Computers & Mathematics with Applications*, 77(6):1722–1738, 2019.

[22] N. Z. Monteiro, R. R. Pereira, B. M. Rocha, R. W. dos Santos, S. R. Mazorche, and A. F. D. Loula. A novel second-order ADI scheme for solving epidemic models with cross-diffusion. *Journal of Computational Science*, 81:102341, 2024.

[23] W. Piyawong, E. H. Twizell, and A. B. Gumel. An unconditionally convergent finite-difference scheme for the SIR model. *Applied Mathematics and Computation*, 146(2-3):611–625, 2003.

[24] D. Prodanov. Comments on some analytical and numerical aspects of the SIR model. *Applied Mathematical Modelling*, 95:236–243, 2021.

[25] D. Prodanov. Asymptotic analysis of the SIR model and the Gompertz distribution. *Journal of Computational and Applied Mathematics*, 422:114901, 2023.

[26] S. Sharma, V. Dolean, P. Jolivet, B. Robinson, J. D. Edwards, T. Kendzerska, and A. Sarkar. Scalable computational algorithms for geospatial COVID-19

spread using high performance computing. *Mathematical Biosciences and Engineering*, 20(8):14634–14674, 2023.

[27] D. Shi and C. Li. Superconvergence analysis of two-grid methods for bacteria equations. *Numerical Algorithms*, 86(1):123–152, 2020.

[28] Y. Shi, X. Yang, and Z. Zhang. Construction of a new time-space two-grid method and its solution for the generalized Burgers' equation. *Applied Mathematics Letters*, 158:109244, 2024.

[29] S. Side, A. M. Utami, Sukarna, and M. I. Pratama. Numerical solution of SIR model for transmission of tuberculosis by Runge-Kutta method. *Journal of Physics: Conference Series*, 1040:012021, 2018.

[30] T. C. Sideris. *Ordinary Differential Equations and Dynamical Systems*. Atlantis Press, 2013.

[31] M. Slodicka. Error estimates of an efficient linearization scheme for a nonlinear elliptic problem with a nonlocal boundary condition. *ESAIM: Mathematical Modelling and Numerical Analysis*, 35(4):691–711, 2001.

[32] Z. Wang and Q. Zhang. Positivity-preserving numerical method and relaxed control for stochastic Susceptible-Infected-Vaccinated epidemic model with Markov switching. *Journal of Computational Biology*, 30(6):695–725, May 2023.

[33] X. Yang, W. Qiu, H. Zhang, and L. Tang. An efficient alternating direction implicit finite difference scheme for the three-dimensional time-fractional telegraph equation. *Computers & Mathematics with Applications*, 102:233–247, 2021.

[34] X. Yang and Z. Zhang. Analysis of a new NFV scheme preserving DMP for two-dimensional subdiffusion equation on distorted meshes. *Journal of Scientific Computing*, 99(3), 2024.

[35] X. Yang and Z. Zhang. Superconvergence analysis of a robust orthogonal Gauss collocation method for 2D fourth-order subdiffusion equations. *Journal of Scientific Computing*, 100(3), 2024.