



OPEN

Optimal control for a SIR epidemic model with limited quarantine

Rocío Balderrama¹, Javier Peressutti², Juan Pablo Pinasco^{1,3}, Federico Vazquez⁴ & Constanza Sánchez de la Vega^{1,4}✉

Social distance, quarantines and total lock-downs are non-pharmaceutical interventions that policymakers have used to mitigate the spread of the COVID-19 virus. However, these measures could be harmful to societies in terms of social and economic costs, and they can be maintained only for a short period of time. Here we investigate the optimal strategies that minimize the impact of an epidemic, by studying the conditions for an optimal control of a Susceptible-Infected-Recovered model with a limitation on the total duration of the quarantine. The control is done by means of the reproduction number $\sigma(t)$, i.e., the number of secondary infections produced by a primary infection, which can be arbitrarily varied in time over a quarantine period T to account for external interventions. We also assume that the most strict quarantine (lower bound of σ) cannot last for a period longer than a value τ . The aim is to minimize the cumulative number of ever-infected individuals (recovered) and the socioeconomic cost of interventions in the long term, by finding the optimal way to vary $\sigma(t)$. We show that the optimal solution is a single *bang-bang*, i.e., the strict quarantine is turned *on* only once, and is turned *off* after the maximum allowed time τ . Besides, we calculate the optimal time to begin and end the strict quarantine, which depends on T , τ and the initial conditions. We provide rigorous proofs of these results and check that are in perfect agreement with numerical computations.

The Covid-19 pandemic outbreak raises an unprecedented series of decisions in different countries around the world. Since vaccines and effective pharmaceutical treatments were not initially available, governments had decided to impose non-pharmaceutical interventions like social distance, quarantines and total lock-downs as the most effective tools to mitigate the spread of the disease. Although these kinds of measures are helpful in reducing the virus transmission and giving time to health systems to adapt, they could be extremely stressful in terms of economic and social costs, and, in longer periods, tend to have less compliance with the population.

In this article we consider the classical SIR model introduced by Kermack and McKendrick¹ and widely used in epidemiology^{2,3}, where the population is divided in compartments of Susceptible, Infected and Recovered (or Removed) individuals. As it is usual in SIR models, we assume that people who have recovered develop immunity and, therefore, would not be able to get infected nor infect others. We consider that infection and recovery rates are allowed to change over time, and that are homogeneous among the population, instead of heterogeneous rates, not depending on age, individual protection measures, or awareness^{4,5}. Let us observe that these heterogeneous rates implies the existence of effective rates, obtained as weighted means of the individual rates⁶. However, for age-structured models, better results are obtained by considering integro-differential equations, and different tools from mathematical control theory are needed⁷. We also assume a mean-field hypothesis that implies random interactions between any pair of agents, unlike other works^{8–10} where interactions are mediated by an underlying network of contacts. In this case, by weighted means of agents degrees, it is possible to derive a system equivalent to a SIR model¹¹. Also, a full proof of a derivation using probabilistic tools was studied recently¹² together with the optimal control problem for vaccination.

Optimal control problems for a system governed by a SIR or a SEIR model (with the addition of the Exposed compartment) with pharmaceutical interventions as vaccination or treatment were widely studied^{13–16}. On its part, in the field on optimal control problems, non-pharmaceutical interventions were studied mostly for a control consisting of isolation acting only on the infected^{17–20}. Taking into account that there is a window of time when the infected are not detected, in this article we will consider that the quarantine is applied to the whole population. Non-pharmaceutical interventions can range from a mild mitigation policy to a strong suppression

¹Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón I, C1428EGA Buenos Aires, Argentina. ²Departamento de Física, Instituto de Física de Mar del Plata (IFIMAR) CONICET, UNMDP, Universidad Nacional de Mar del Plata, Funes 3350, 7600 Mar del Plata, Argentina. ³IMAS-CONICET, Ciudad Universitaria, Pabellón I, C1428EGA Buenos Aires, Argentina. ⁴Instituto de Cálculo, FCEN, Universidad de Buenos Aires and CONICET, C1428EGA Buenos Aires, Argentina. ✉email: csfvega@gmail.com

policy. As discussed by Ferguson et al.²¹, a suppression policy “*aims to reverse epidemic growth, reducing case numbers to low levels and maintaining that situation indefinitely*”. Suppression can be achieved by restricting travels, closing schools and nonessential businesses, banning social gatherings, and asking citizens to shelter in place. These measures, often referred to as a lockdown, are highly restrictive on social rights and damaging to the economy. In contrast, a mitigation policy “*focuses on slowing but not necessarily stopping epidemic spread*”. Mitigation measures may involve discouraging air travel while encouraging remote working, requiring companies to provide physical separation between workers, banning large gatherings, isolating the vulnerable, and identifying and quarantining contagious individuals and their recent contacts.

A critical parameter in the SIR model is the basic reproduction number R_0 , defined as the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection. At the beginning of the epidemic, when no one in the population is immune, infected individuals will infect R_0 other people on average. Let us observe that, for $R_0 < 1$, the number of new cases decline, and when $R_0 > 1$, the number of new cases grows. However, at any time $t > 0$, the effective reproduction number R_t replaces R_0 , since the number of contacts between infected and susceptible agents is reduced due to the interactions with recovered individuals that are immune. Hence, the epidemic grows until a sufficient fraction of the population becomes infected, and after reaching a peak starts to gradually decline. Following Ferguson et al.²¹, for the Covid-19 the suppression phase can achieve $R_0 < 1$, while the mitigation measures are unlikely to bring R_0 below 1. Therefore, the number of new cases are expected to decline during the suppression phase and to start rising again during the mitigation phase, although at a slower rate than in a non-intervention scenario.

In this work we assume that the intervention will occur in a preset period of time T , as proposed by Greenhalgh²² and recently by Ketcheson²³, since it is unrealistic that interventions can be sustained indefinitely. Also, the lockdown or (strict quarantine) can last at most for a period $\tau < T$, the maximum time that the population will adhere. Now, there are several interesting questions related to the implementation of the measures:

1. When should the suppression policy begin in $[0, T]$?
2. Is it convenient to split the maximum time τ into different intervals?
3. Is it better to apply a strong lockdown followed by mild mitigation measures or not?

In this article we study the previous questions using optimal control tools and numerical computations^{24–27}. The answers clearly depend on the goal, which in our case is to minimize the overall impact of the epidemics in terms of the final number of infected individuals and the social and economic cost of the interventions, which we assume to increase as the quarantine becomes more strict. To account for quarantine measures, we consider a time-dependent reproduction number. Using an optimal control approach we show that the optimal strategy is of a single *bang-bang* type, that is, the lockdown or strict quarantine is applied in a single interval of time. Moreover, we characterize the time to start and finish the lockdown during the intervention phase. Let us remark that these questions make sense also in SIHR models, which include hospitalized individuals, since it must be necessary to keep the maximum of the hospitalized group below some threshold²⁸.

Recently, many works have appeared dealing with these and related issues. The optimal time to start the suppression measures that maximizes this type of objective function was studied by Ketcheson²³, where it was proved that a bang-bang control is optimal. However, in that work it is assumed that the lockdown corresponds to a zero reproduction number, something that is impossible to achieve in the real world. Moreover, it is assumed that the strict lockdown can last during the whole intervention, which seems to be impracticable. This problem was also analyzed for a different objective function²⁹, i.e., minimizing the peak of infected individuals, for which they proved that the optimal policy is not bang-bang. Besides, Kruse and Strack³⁰ minimize a functional that depends on the number of infectives during the intervention plus a term that measures the social and economic cost of interventions, and prove that the optimal control is bang-bang, but they do not investigate the optimal time to start the suppression policy.

The second question is suggested by the strategy proposed by Ferguson et al.²¹: the lockdown must be turned *on* and *off* several times based on the incidence of the virus in the population. A control-theoretic approach was considered in several works^{28,31–33}, although no time limits for the interventions were imposed. We shall see that the optimal policy is of bang-bang type, which consists on turning the lockdown *on* only once and turning it *off* after the maximum allowed time τ , in agreement with other authors^{23,30}.

Finally, the third question involves both suppression and mitigation phases, and one of the policies was colorfully characterized as *the hammer and the dance* in³⁴: a strict lockdown, followed by mitigation measures in order to keep under control the propagation of the disease. However, our main results indicate that the best strategy actually depends on the initial condition, determined by the relation between R_0 and the initial fraction of susceptible individuals x_0 . On the one hand, when x_0 is smaller than $1/R_0$ the optimal strategy consists on applying a strong lockdown right at the beginning of the intervention period $[0, T]$ for the maximum time period τ , followed by a mild mitigation measure until the end of the intervention (strong-mild strategy). On the other hand, when x_0 is larger than $1/R_0$ the optimal strategy is to apply mild mitigation measures at the beginning of the intervention, followed by a strong lockdown, and then a mild mitigation measure again in some cases (mild-strong-mild strategy). In this case, the optimal time to start the strong lockdown depends non-trivially on the initial condition.

The paper is organized as follows. We start by describing the basic SIR model in “[The SIR model](#)” section, and by introducing the SIR model with control in “[The SIR control model](#)” section. We then present the main results of the article supported by numerical simulations in “[Results](#)” section, followed by a discussion and conclusions including future work in “[Discussion and conclusions](#)” section. Finally, in “[Methods](#)” section we provide rigorous proofs of the results by applying Pontryagin’s maximum principle to the control problem, and we prove that

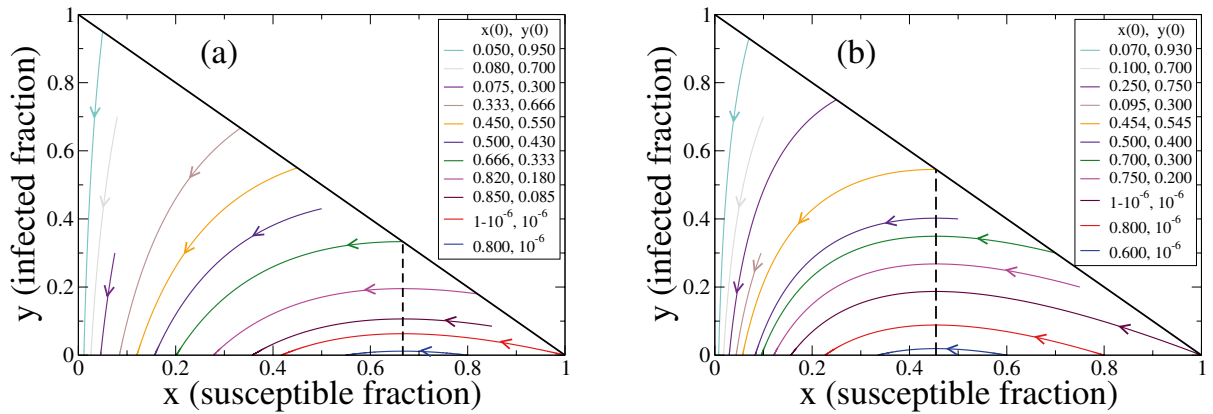


Figure 1. Trajectories of the system in the $x - y$ phase space for the SIR model Eq. (1) with basic reproduction number $\sigma = 1.5$ (a) and $\sigma = 2.2$ (b), where x and y are the fractions of susceptible and infected individuals, respectively. Each curve corresponds to a trajectory starting from a given initial condition $(x(0), y(0))$, as indicated in the legends. Arrows denote the direction of the time evolution of x and y . The vertical dashed line corresponds to the critical value $x = 1/\sigma$ where $y' = 0$.

the optimal control is bang-bang in Lemma 2. The main results that characterize the optimal control are given in Theorem 2, Corollary 1 and Theorem 3, for different case scenarios.

The SIR model

The basic SIR compartmental model of infectious diseases introduced by Kermack and McKendrik¹ considers a population of individuals that is divided in three compartments with homogeneous characteristics: Susceptible (S), Infected (I), and Recovered or Removed (R). The fraction of susceptible, infected and recovered individuals at time t is denoted by $x(t)$, $y(t)$ and $z(t) = 1 - x(t) - y(t)$, respectively. It is assumed that each infected individual is in contact with an average number of β random individuals per unit time, and that infects only those who are susceptible ($S \rightarrow I$ transition), generating new infections at an average rate $\beta x(t)$. Besides, each infected individual recovers at a rate γ ($I \rightarrow R$ transition). Births and deaths are neglected, and the recovered population is assumed to no longer infect others and cannot be reinfected. The infection and recovery rates β and γ , respectively, are related to the basic reproduction number σ by $\sigma = \beta/\gamma$, that is, the mean number of infections produced by a single infected individual in a totally susceptible population ($x = 1$) during its mean infectious period $1/\gamma$. Then, the evolution of the system is governed by the following set of coupled nonlinear ordinary differential equations (see Hethcote²⁶, section 2.1):

$$x'(t) = -\gamma \sigma x(t) y(t), \tag{1a}$$

$$y'(t) = \gamma \sigma x(t) y(t) - \gamma y(t), \tag{1b}$$

with $(x(0), y(0)) \in \mathcal{D} = \{(x_0, y_0) : x_0 > 0, y_0 > 0, x_0 + y_0 \leq 1\}$. Here x' and y' are short notations for the time derivatives dx/dt and dy/dt , respectively. The region \mathcal{D} is forward-invariant and there exists a unique solution for all time²⁶. Then, since $y_0 > 0$, the proportion of infectious individuals is positive at any time. Even though the temporal dynamics of Eq. (1) depends on both σ and γ , the set of system's trajectories on the $x - y$ space depends only on the basic reproduction number σ because γ only affects the overall time scale of the system. In Fig. 1 we depict typical trajectories starting from different initial conditions $x(0), y(0)$, for $\sigma = 1.5$ [panel (a)] and $\sigma = 2.2$ [panel (b)].

The system of Eq. (1) is at equilibrium if $y(t) = 0$. This equilibrium is stable only if $x(t) \leq 1/\sigma$, a condition referred to as herd immunity. If this condition is not satisfied at the initial time ($x(0) > 1/\sigma$), then $y(t)$ first increases until it reaches its maximum value at a time t for which $x(t) = 1/\sigma$ (dashed vertical lines in Fig. 1), and then decreases and approaches zero asymptotically, i.e., $y_\infty \equiv \lim_{t \rightarrow \infty} y(t) = 0$. That is, for $y(0) > 0, \gamma \geq 0$ and $\sigma \geq 0$ is $y(t)$ larger than zero for any finite time $t \geq 0$. The fraction of susceptible individuals $x(t)$ is strictly decreasing, and its value in the long time limit $x_\infty \equiv \lim_{t \rightarrow \infty} x(t)$ is always positive. Therefore, the state of the system in the long time limit consists only of susceptible and recovered individuals, $x_\infty + z_\infty = 1$, where $z_\infty \equiv \lim_{t \rightarrow \infty} z(t)$. Also, it is known that $x_\infty \in (0, 1/\sigma)$ (see Theorem 2.1 of the work by Hethcote²⁶).

The SIR control model

We now extend the classical SIR model to address the problem of controlling the spread of an epidemics with no access to vaccination, where the only possible control is isolation. We model this non-pharmaceutical intervention via a time dependent reproduction number $\sigma(t)$ that can be varied in the interval $[\sigma_s, \sigma_m]$, where σ_s corresponds to a more strict isolation ("strict" quarantine) than σ_m ("mild" quarantine), with $0 \leq \sigma_s < \sigma_m$, and assume that this intervention can only be applied over a finite time interval $[0, T]$. Here T is the length of the intervention period. After the intervention, the restrictions are removed, thus the disease spreads freely and $\sigma(t) = \sigma_f \geq \sigma_m$ for all $t > T$. We think of the control parameter $\sigma(t)$ as capturing political measures such as

social distancing, and the lockdown of businesses, schools, universities and other institutions. Then, the system evolves according to the following set of coupled nonlinear ordinary differential equations:

$$x'(t) = -\gamma \sigma(t) x(t) y(t), \quad (2a)$$

$$y'(t) = \gamma \sigma(t) x(t) y(t) - \gamma y(t), \quad (2b)$$

with $(x(0), y(0)) \in \mathcal{D} = \{(x_0, y_0) : x_0 > 0, y_0 > 0, x_0 + y_0 \leq 1\}$, $\sigma(t) \in [\sigma_s, \sigma_m]$ for $t \in [0, T]$ and $\sigma(t) = \sigma_f$ for $t > T$, where $0 \leq \sigma_s < \sigma_m \leq \sigma_f$. We also assume that during the intervention period $[0, T]$ it is not possible to impose an extremely restrictive isolation for a long time. Thus, we consider that the strict quarantine—corresponding to σ_s —can last at most for a fixed time period τ , with $\tau \in (0, T)$. Once the period of intervention is finished at time T we compute $x_\infty(x(T), y(T), \sigma_f) = \lim_{t \rightarrow \infty} x(t)$, where $(x(t), y(t))$ is the solution of the system of Eq. (2) with initial condition $(x(T), y(T))$ and constant reproduction number $\sigma(t) \equiv \sigma_f$ for $t > T$. Note that in this case, from Hethcote²⁶ we deduce that $x_\infty(x(T), y(T), \sigma_f) \in (0, 1/\sigma_f)$.

While political measures reduce the spread of the disease, they often come at an important economic and social cost. A long and strict quarantine can be very effective at reducing contagions, but at the expense of having a negative impact on the economy. Our goal is to find the optimal control on the SIR model described above that minimizes the total damage of a pandemic in terms of both, the total number of infections and also the socioeconomic costs. We model this trade-off by considering a global cost capturing the total number of individuals that were infected during the epidemics, i.e., those who are recovered in the long-time limit z_∞ , and the socioeconomic cost of shutting down society during the intervention on $[0, T]$, which we assume to increase as σ decreases (more restrictions). In order to find the optimal $\sigma(t)$ it proves convenient to work with the fraction of susceptible individuals in the long term x_∞ instead. Then, given that minimizing z_∞ is equivalent to maximizing x_∞ , since $x_\infty + z_\infty = 1$, we define the functional

$$J(x, y, \sigma) := x_\infty(x(T), y(T), \sigma_f) + C(\sigma), \quad (3)$$

where

$$C(\sigma) := \int_0^T L(\sigma(t)) dt. \quad (4)$$

Our goal is to maximize the functional J , which has the following interpretation. The first term of J is the fraction of individuals that remain susceptible in the long term x_∞ , and that we want to keep as large as possible subject to the condition of maximizing the second term of J as well, the functional $C(\sigma)$. The functional $C(\sigma)$ is taken to be inversely proportional to the socioeconomic cost of the intervention ($C(\sigma)$ increases as the cost decreases), as the function L is assumed to be a monotonously increasing function of σ . Then, an increase of the socioeconomic cost is achieved by decreasing σ (more restrictions or stricter quarantine), which leads to decreasing L and consequently $C(\sigma)$. Therefore, we see that there is a non-trivial competition between the two terms of Eq. (3), given that by decreasing σ the value of x_∞ increases, while $C(\sigma)$ decreases.

In the next section we describe the main results about the optimal control and we test them via numerical simulations.

Results

As mentioned in the last section, the optimal control is given by the shape of $\sigma(t)$ that minimizes both, the final number of infected individuals and the socioeconomic costs, which corresponds to maximizing the functional J from Eqs. (3) and (4). From now on we restrict ourselves to the case where the socioeconomic cost of imposing a quarantine is linear in the control $\sigma(t)$. This is a simplified first approach that narrows the analysis of the general problem formulated in "The SIR control model" section but, as we shall see, has the advantage of providing further insight into the structure of the optimal policy. The assumption that the cost of socioeconomic measures that reduce the transmission rate is linear in σ can be interpreted in the context of social distancing as given by Kruse and Strack³⁰: "Shutting down half of the economy for two days is equally costly as shutting down the whole economy for a single day". Then, we consider that the function L in Eq. (4) is a linear and increasing function that depends only on the control σ , that is $L(\sigma(t)) = \kappa \sigma(t)$, which satisfies the condition of being a monotonically increasing function of σ expressed in the last section. In this case, the parameter $\kappa \geq 0$ could be interpreted as the assessment that a policy maker gives to the socioeconomic impact of the quarantine compared to the final number of infected individuals. In this regard, κ is a fixed real number that can be chosen small enough by a government that intends to reduce the final number of infected individuals regardless the socioeconomic impact, or can be chosen large enough by a government that can face a large number of final infected and intends to control the socioeconomic impact. Here we mainly focus on the case where κ is small (see condition Eq. (33) in "Methods" section). Also, when κ is large enough we prove that the optimal strategy consists in calling off the lockdown and take mild mitigation measures for all the intervention period (see Lemma 6 in "Methods" section).

Under these conditions, we prove in "Methods" section that the optimal control is of the form of a single *bang-bang*. This consists on turning the strict quarantine *on* only once and switching it *off* after the maximum allowed time τ or, eventually, when the intervention ends at time T , depending on the initial condition and the values of the strict and mild reproduction numbers σ_s and σ_m , respectively. Then, the problem is reduced to find the optimal time to start the strict quarantine, which we call t^* , and its length called η^* . We also show in "Methods" section that the length of the strict quarantine η^* for the optimal control could be less than the maximum time

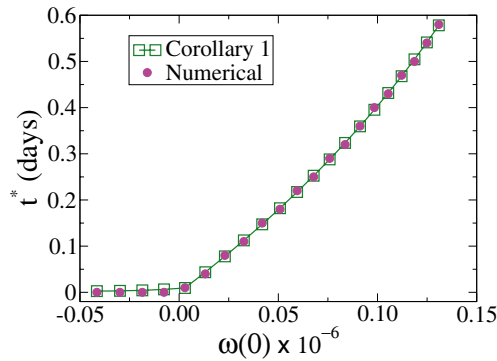


Figure 2. Optimal initial time of the strict quarantine t^* vs $w(0)$ for $\tau = 10$ and $\kappa = 0$. The other parameters are $\gamma = 0.1, \sigma_m = \sigma_f = 1.5, \sigma_s = 0$ and $T = 260$.

τ in some cases, as we shall see below. This means that, surprisingly, sometimes it is more convenient to make a shorter use of the strict quarantine to obtain better results in terms of pandemic costs.

We analyze the optimal control in three different case scenarios: i) $\kappa = 0, \sigma_m = \sigma_f$ and $\sigma_s = 0$, ii) $\kappa = 0, \sigma_m = \sigma_f$ and $\sigma_s > 0$ and iii) $\kappa > 0$ and $0 < \sigma_s < \sigma_m < \sigma_f$. The optimal times t^* and η^* for each case are given in Corollary 1, Theorems 2 and 3, respectively, of “Methods” section, where the interested reader can find rigorous proofs. The optimal control for the different cases, given by t^* and η^* , is summarized and numerically tested below for specific parameter values. For that, we integrate the system of Eq. (2) using the Adams’ method^{35,36}, for various time periods τ of the strict quarantine ($\sigma = \sigma_s$) in the bang-bang control, starting at time t but ending before T , which is the control period ($\sigma(t) = \sigma_f$ for $t > T$). That is, the value of $\sigma(t)$ adopts the following form, depending on t, τ and T :

$$\sigma(r) = \begin{cases} \sigma_m & \text{for } 0 \leq r < t, \\ \sigma_s & \text{for } t \leq r < t + \eta, \\ \sigma_m & \text{for } t + \eta \leq r < T, \\ \sigma_f & \text{for } r \geq T, \end{cases} \tag{5}$$

where

$$\eta = \begin{cases} \tau & \text{for } t + \tau \leq T \text{ and} \\ T - t & \text{for } t + \tau > T. \end{cases} \tag{6}$$

We start by testing the simplest case $\kappa = 0, \sigma_m = \sigma_f$ and $\sigma_s \geq 0$, and we then test the most general case $\kappa > 0$ and $0 < \sigma_s < \sigma_m < \sigma_f$. We take the value $\gamma = 0.1/\text{day}$ for the recovery rate, which corresponds to a mean recovery time of 10 days that falls within the range of Covid-19 estimates²³. This value of γ sets the time scales of the system. For now on all time scales are given in units of “days”, even though we omit units for the sake of simplicity.

Case $\kappa = 0, \sigma_m = \sigma_f$ and $\sigma_s = 0$ (Corollary 1). We first analyze the case $\kappa = 0$ with a bang-bang control in the interval $[0, T]$ that consists of a mild quarantine ($\sigma_m = 1.5$) and an extremely strict and unrealistic quarantine ($\sigma_s = 0$) during which there are no infections. The other parameter in the simulations is $T = 260$, together with the initial condition $y_0 = y(t = 0) = 10^{-6}$ and $x_0 = x(t = 0) = 1 - 10^{-6}$. We can see from Corollary 1 that the optimum initial time of the strict quarantine is $t^* = 0$ for $x_0 \leq 1/\sigma_f$ ($w(0) \leq 0$), while for $x_0 > 1/\sigma_f$ ($w(0) > 0$) is given by

$$t^* = \begin{cases} \bar{t} & \text{for } 0 \leq \tau \leq \bar{\tau}, \text{ where } x_{T-\tau, \tau}(T - \tau) \leq \frac{1}{\sigma_f} \text{ and } \eta = \tau, \\ T - \tau & \text{for } \bar{\tau} < \tau < \tilde{\tau}, \text{ where } \frac{1}{\sigma_f} < x_{T-\tau, \tau}(T - \tau) \leq \frac{1}{\sigma_f(1-e^{-\gamma\tau})} \text{ and } \eta = T - \tau, \\ \tilde{t} & \text{for } \tau \geq \tilde{\tau}, \text{ where } x_{T-\tau, \tau}(T - \tau) > \frac{1}{\sigma_f(1-e^{-\gamma\tau})}, \text{ and } \eta = T - \tilde{t}. \end{cases} \tag{7}$$

where $\bar{t} = T - \bar{\tau} \in [0, T - \tau]$ is the unique value, independent from $\tau \in [0, \bar{\tau}]$, such that $x_{\bar{t}, \tau}(\bar{t}) = \frac{1}{\sigma_f}$. Likewise, $\tilde{t} = T - \tilde{\tau} \in [T - \tau, T]$ is the unique value, independent from $\tau \in [\tilde{\tau}, T]$, such that $x_{\tilde{t}, T-\tilde{t}}(\tilde{t}) = \frac{1}{\sigma_f(1-e^{-\gamma(T-\tilde{t})})}$. As $\kappa = 0, x_\infty$ reaches a maximum value when the strict quarantine starts at the optimal time t^* [see Eq. (12)]. In Fig. 2 we plot t^* vs $w(0)$ for $\tau = 10$, calculated from Eq. (54) (squares) and by estimating the maximum of x_∞ (circles). We can see that t^* takes values close to zero for $w(0) \leq 0$. In the rest of this section we consider the case $w(0) > 0$.

The behaviour of t^* from Eq. (7) for $x_0 > 1/\sigma_f$ ($w(0) > 0$) is tested in Fig. 3a, where we compare numerical results (circles) with that obtained from Eq. (7) (squares, Corollary 1). We observe that the agreement between numerical computations and Corollary 1 is very good. Figure 3b is an auxiliary plot that shows how to obtain

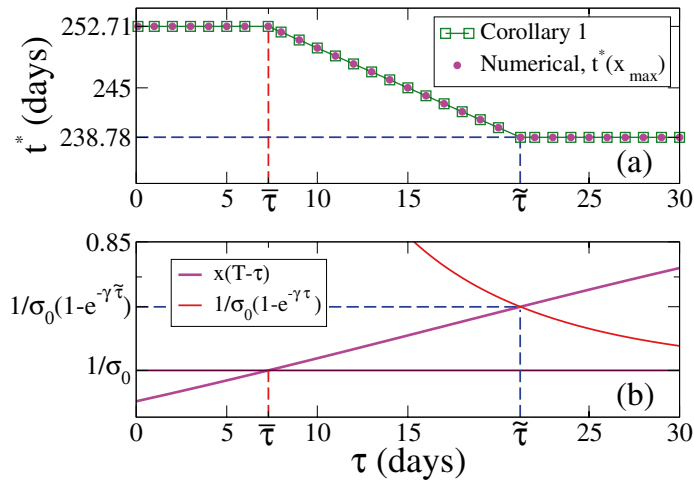


Figure 3. (a) Optimal initial time t^* vs strict quarantine length τ for $\kappa = 0$ and $x_0 > 1/\sigma_f$. (b) Graphical determination of the times $\bar{\tau}$ and $\tilde{\tau}$ that define the three regions for the different behaviours of t^* . The parameters are $\gamma = 0.1, \sigma_m = \sigma_f = 1.5, \sigma_s = 0$ and $T = 260$. The initial condition corresponds to $x(0) = 1 - 10^{-6}, y(0) = 10^{-6}$ ($R_0 = \sigma_f x(0) > 1$). The optimum times of the first and last regions are $\bar{t} \simeq 252.71$ and $\tilde{t} \simeq 238.78$, respectively, determined by $\bar{\tau} \simeq 7.29$ and $\tilde{\tau} \simeq 21.22$.

graphically the optimum times $\bar{t} \simeq 252.71$ and $\tilde{t} \simeq 238.78$ that define the three different regimes of t^* from Eq. (7). These times are obtained by estimating the values of τ for which the curve $x_{t,T-t}(T - \tau)$ crosses the lines $1/\sigma_f$ and $1/[\sigma_f(1 - e^{-\gamma\tau})]$, which happens at $\bar{\tau} \simeq 7.29$ and $\tilde{\tau} \simeq 21.22$, respectively.

Remark 1 The effective reproductive number $R_t^\sigma \equiv \sigma x_\sigma(t)$ represents the mean number of individuals that an agent infects during its infectious period, at time t . It is interesting to note that the optimal time from Eq. (7) can be rewritten in terms of R_t^σ as

$$t^* = \begin{cases} \bar{t} & \text{for } 0 \leq \tau \leq \bar{\tau}, \text{ where } R_{T-\tau}^{\sigma_f} \leq 1 \text{ and } \eta = \tau, \\ T - \tau & \text{for } \bar{\tau} \leq \tau \leq \tilde{\tau}, \text{ where } 1 < R_{T-\tau}^{\sigma_f} \leq \frac{1}{1-e^{-\gamma\tau}} \text{ and } \eta = T - \tau, \\ \tilde{t} & \text{for } \tau > \tilde{\tau}, \text{ where } R_{T-\tau}^{\sigma_f} > \frac{1}{1-e^{-\gamma\tau}}, \text{ and } \eta = T - \tilde{\tau}. \end{cases} \tag{8}$$

Here $R_{T-\tau}^{\sigma_f} = \sigma_f x_{\sigma_f}(T - \tau)$, $\bar{\tau} = T - \bar{t}$ and $\tilde{\tau} = T - \tilde{t}$, where \bar{t} and \tilde{t} are determined from the relations

$$R_{\bar{t}}^{\sigma_f} = 1 \quad \text{and} \quad R_{\tilde{t}}^{\sigma_f} = \frac{1}{1 - e^{-\gamma(T-\tilde{t})}}. \tag{9}$$

In Fig. 4 we show the evolution of the system in the $x - y$ phase space for a given τ and various t (right panels), together with the evolution of $\sigma(t)$ (left panels), which describe the three different behaviours of t^* . All curves start at $(x_0, y_0) = (1 - 10^{-6}, 10^{-6})$ and follow the top curve with mild quarantine ($\sigma = \sigma_f$) until the strict quarantine starts at t ($\sigma = \sigma_s = 0$), vertically falling down up to a lower level curve when the mild quarantine starts, and finally following this curve until the fixed point $(x_\infty, 0)$ is asymptotically reached. The vertical trajectory describes the evolution within the strict quarantine where $x(t)$ remains constant, given that $\sigma(t) = \sigma_s = 0$ in that period. The optimum time t^* that leads to the maximum of x_∞ corresponds to the time for which $y(t)$ drops to the lowest level curve in the interval $[t, t + \eta]$ (pink curve). For $\tau = 6 < 7.29 = \bar{\tau}$ (Fig. 4 top panels) we see that the maximum of x_∞ is reached starting the strict quarantine at $t^* = \bar{t} = 252.71$, where the effective reproduction number is $R_{\bar{t}}^{\sigma_f} = R_{\bar{t}+\eta}^{\sigma_f} = 1$, and thus there is no new outbreak when the strict quarantine is released ($\frac{dy}{dt}|_{t+\eta} = 0$). In this case the entire quarantine period $\eta = \tau$ is used. For $\bar{\tau} < \tau = 12 < \tilde{\tau} = 21.22$ (Fig. 4 middle panels) the optimum initial time is $t^* = T - \tau = 248 < \bar{t}$, obtained by still using the entire strict quarantine period but starting earlier than \bar{t} . Finally, for $\tau = 26 > \tilde{\tau}$ (Fig. 4 bottom panels) the optimum is $t^* = \tilde{t} = 238.78 > T - \tau = 234$, where it turns more effective to use the strict quarantine for a shorter time $T - t^* < \tau$. Notice that implementing a shorter but later strict quarantine is more efficient than using a longer and earlier strict quarantine, as we can see by comparing $\sigma(t)$ for $\eta = 26$ and $\eta = 21.22$ in the bottom panels for the $\tau = 26$ case.

Case $\kappa = 0, \sigma_m = \sigma_f$ and $\sigma_s > 0$ (Theorem 2). We now analyze the case $\kappa = 0, \sigma_m = \sigma_f = 1.5$ and $\sigma_s = 0.3 > 0$, with $T = 260$. This corresponds to a strict quarantine that is softer than in the previous case $\sigma_s = 0$, and during which there are infections. Initially is $y_0 = 10^{-6}$ and $x_0 = 1 - 10^{-6}$. We can see from Theorem 2 that the optimum initial time of the strict quarantine t^* for $w(0) > 0$ is

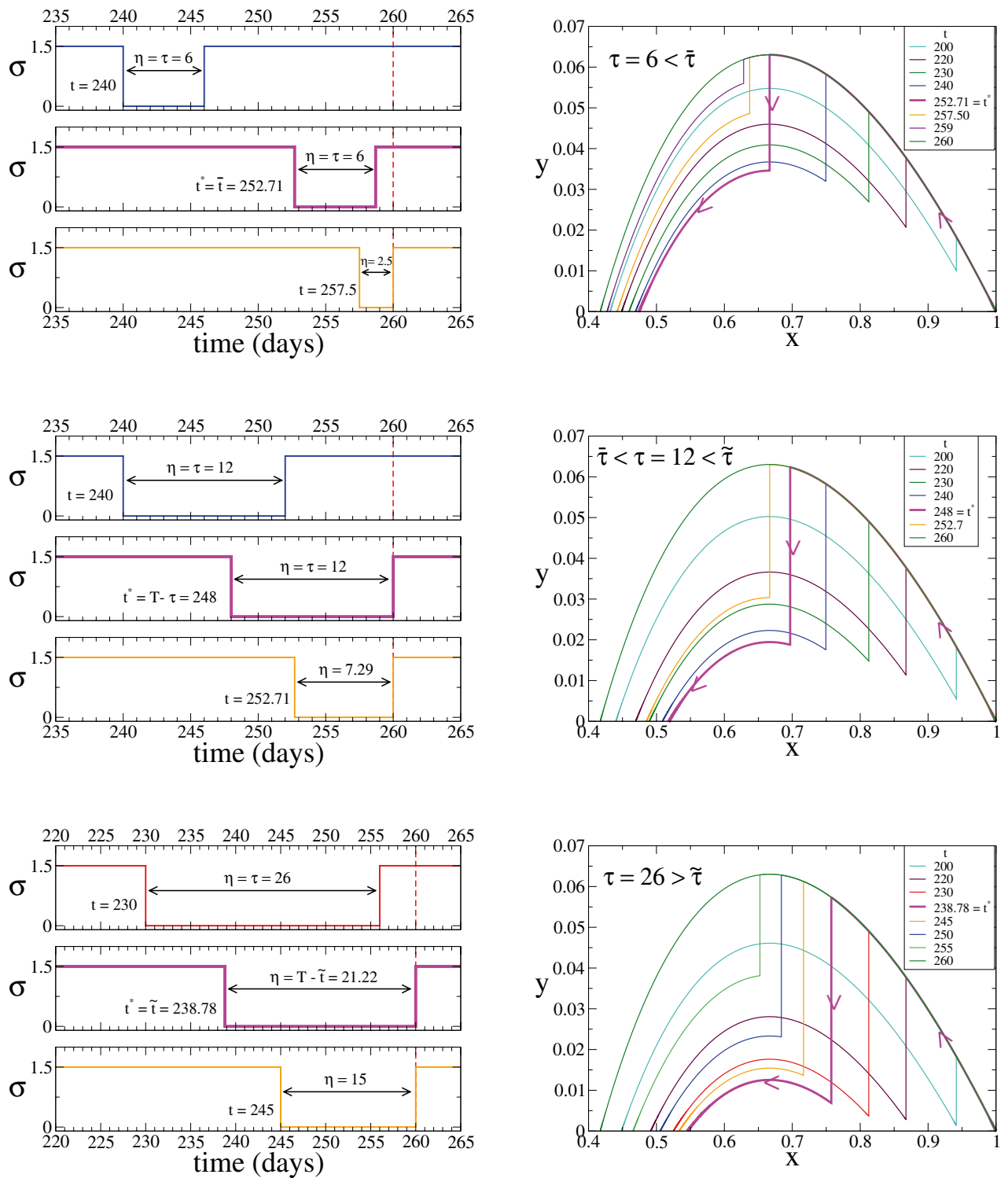


Figure 4. Case $\kappa = 0, \sigma_m = \sigma_f$ and $\sigma_s = 0$ (Corollary 1). System's trajectory in the $x - y$ phase space (right panels), for $\gamma = 0.1, \kappa = 0, \sigma_m = \sigma_f = 1.5, \sigma_s = 0$ and $T = 260$, and the three values of τ indicated in the legends corresponding to the different regimes of the optimum time t^* (pink lines). Left panels show the time evolution of σ for three different initial times t of the strict quarantine in each case. The optimum times are $t^* = \bar{t} \simeq 252.71$ for $\tau = 6$ (top panels), $t^* = T - \tau = 248$ for $\tau = 12$ (middle panels) and $t^* = \tilde{t} \simeq 238.78$ for $\tau = 26$ (bottom panels).

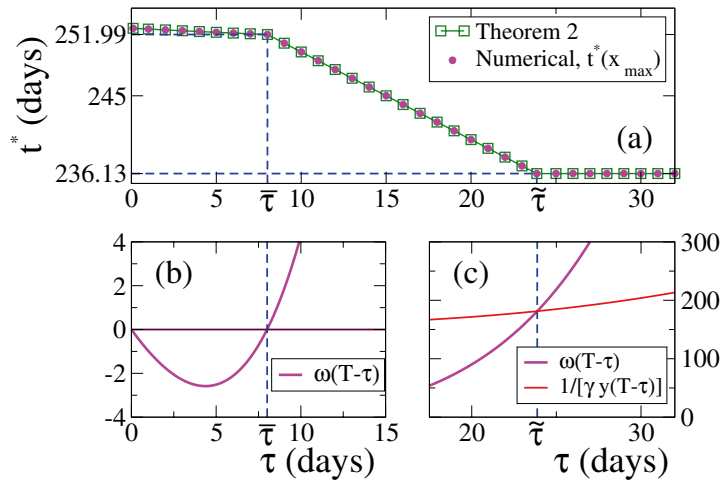


Figure 5. (a) Optimal initial time t^* vs strict quarantine length τ for $\kappa = 0$ and $w(0) > 0$. (b) and (c) Graphical determination of the times $\bar{\tau}$ and $\tilde{\tau}$, respectively, which define the three regions for the different behaviours of t^* . The parameters are $\gamma = 0.1, \sigma_f = 1.5, \sigma_s = 0.3, T = 260$. The initial condition corresponds to $x(0) = 1 - 10^{-6}, y(0) = 10^{-6}$. The optimum time for the region $\tau > \tilde{\tau} \approx 23.87$ is $\tilde{t} \approx 236.13$, while t^* has a slight dependence on τ for $\tau \leq \bar{\tau} \approx 8.01$.

$$t^* = \begin{cases} \bar{t} & \text{for } 0 \leq \tau \leq \bar{\tau}, \text{ where } w(0) \geq 0, w(T - \tau) \leq 0 \text{ and } \eta = \tau, \\ T - \tau & \text{for } \bar{\tau} \leq \tau \leq \tilde{\tau}, \text{ where } 0 < w(T - \tau) \leq \frac{1}{\gamma y_{T-\tau, \tau}(T-\tau)} \text{ and } \eta = T - \tau, \\ \tilde{t} & \text{for } \tau > \tilde{\tau}, \text{ where } w(T - \tau) > \frac{1}{\gamma y_{T-\tau, \tau}(T-\tau)}, \text{ and } \eta = T - \tilde{t}, \end{cases} \quad (10)$$

where $\bar{t} \in [0, T - \tau]$ is the unique value, depending on $\tau \in [0, \bar{\tau}]$, such that $w(\bar{t}) = 0$. On the other hand, $\tilde{t} \in [T - \tau, T]$ is the unique value, independent from $\tau \in [\tilde{\tau}, T]$, such that $w(\tilde{t}) = \frac{1}{\gamma y_{T-\tau, \tau}(\tilde{t})}$. The dependence and independence of $w(t)$ on τ for $t \in [0, T - \tau]$ and $t \in [T - \tau, T]$, respectively, can be seen from the definition of $w(t)$ in Eq. (42).

In Fig. 5a we compare numerical results (circles) with results from Eq. (10) (squares, Theorem 2), where we see a very good agreement. At t^* , x_∞ reaches a maximum. Unlike the $\sigma_s = 0$ case, for $\sigma_s = 0.3 > 0$ the optimal time t^* in the $0 \leq \tau \leq \bar{\tau} \approx 8.01$ interval depends on τ , that is, $t^* = \bar{t}(\tau)$, while for $\tau > \tilde{\tau} \approx 23.87$ is $t^* = \tilde{t} \approx 236.13$ independent of τ . Figure 5b,c show that the optimal times \bar{t} and \tilde{t} are estimated, respectively, as the values of $t = T - \tau$ for which the curve $w(T - \tau)$ crosses the horizontal line 0 and the curve $1/[\gamma y_{t, T-t}(T - \tau)]$.

Figure 6 is analogous to Fig. 4 for the $\sigma_s = 0$ case, and depicts the three different behaviours of t^* . Curves are similar to those of $\sigma_s = 0$, where the main difference is that for $\sigma_s = 0.3 > 0$ the trajectory of the system within the strict quarantine in the $x - y$ space is described by a diagonal line (see inset of top-right panel), given that σ_s is larger than zero and thus $x(t)$ decreases in this period. At the optimum time t^* , $y(t)$ drops to the lowest level curve in the interval $[t, t + \eta]$ (pink curves).

General case $\kappa > 0$ and $0 < \sigma_s < \sigma_m < \sigma_f$ (Theorem 3). In this section we analyze the most general case $\kappa = 10^{-5} > 0$, with a mild quarantine ($\sigma_m = 1.5$) together with a strict quarantine ($\sigma_s = 0.3 < \sigma_m$) during the control interval $t \in [0, T]$, and with $\sigma(t) = \sigma_f = 2.2 > \sigma_m$ for the case of no restrictions after the control period $t > T$. We take $T = 320$, and the rest of the parameters are the same as those in the previous studied cases. Then, from Theorem 3 the optimum initial time t^* is given by

$$t^* = \begin{cases} \bar{t} & \text{for } 0 \leq \tau \leq \bar{\tau}, \text{ where } w(0) \geq 0, w(T - \tau) \leq 0 \text{ and } \eta = \tau, \\ T - \tau & \text{for } \bar{\tau} \leq \tau \leq \tilde{\tau}, \text{ where } 0 < w(T - \tau) \leq \alpha(T - \tau) \text{ and } \eta = T - \tau, \\ \tilde{t} & \text{for } \tau > \tilde{\tau}, \text{ where } w(T - \tau) > \alpha(T - \tau), \text{ and } \eta = T - \tilde{t}, \end{cases} \quad (11)$$

where $\bar{t} \in [0, T - \tau]$ is a unique value that depends on $\tau \in [0, \bar{\tau}]$ and satisfies $w(\bar{t}) = 0$, while $\tilde{t} \in [T - \tau, T]$ is a unique value independent of $\tau \in [\tilde{\tau}, T]$ that satisfies $w(\tilde{t}) = \alpha(\tilde{t})$. Here $\alpha(t)$ is given by Eq. (36), whereas the dependence and independence of $w(t)$ on τ for $t \in [0, T - \tau]$ and $t \in [T - \tau, T]$, respectively, is seen in the definition of $w(t)$ in Eq. (35).

Given that we consider here $\kappa > 0$, J reaches a maximum at the optimum time t^* (see Eq. (12)). Figure 7a shows the behaviour of t^* as a function of τ , where we observe a very good agreement between numerical results (circles) and Theorem 3 (squares). We also see that t^* depends slightly on τ in the $0 \leq \tau \leq \bar{\tau} \approx 9.65$ interval, while $t^* = \tilde{t} \approx 291.46$ for $\tau > \tilde{\tau} \approx 28.54$. The optimal times \bar{t} and \tilde{t} are estimated as the values of $t = T - \tau$ for which the curve $w(T - \tau)$ crosses the horizontal line 0 and the curve $\alpha(T - \tau)$, respectively (Fig. 7b,c).

In the right panels of Fig. 8 we show the system's evolution in the $x - y$ space for three different values of τ corresponding to the different behaviour of t^* . Unlike the previously studied cases where $\sigma_m = \sigma_f$ (Figs. 4 and 6), here we observe that the curves $(x(t), y(t))$ may exhibit up to three different regimes within the control period

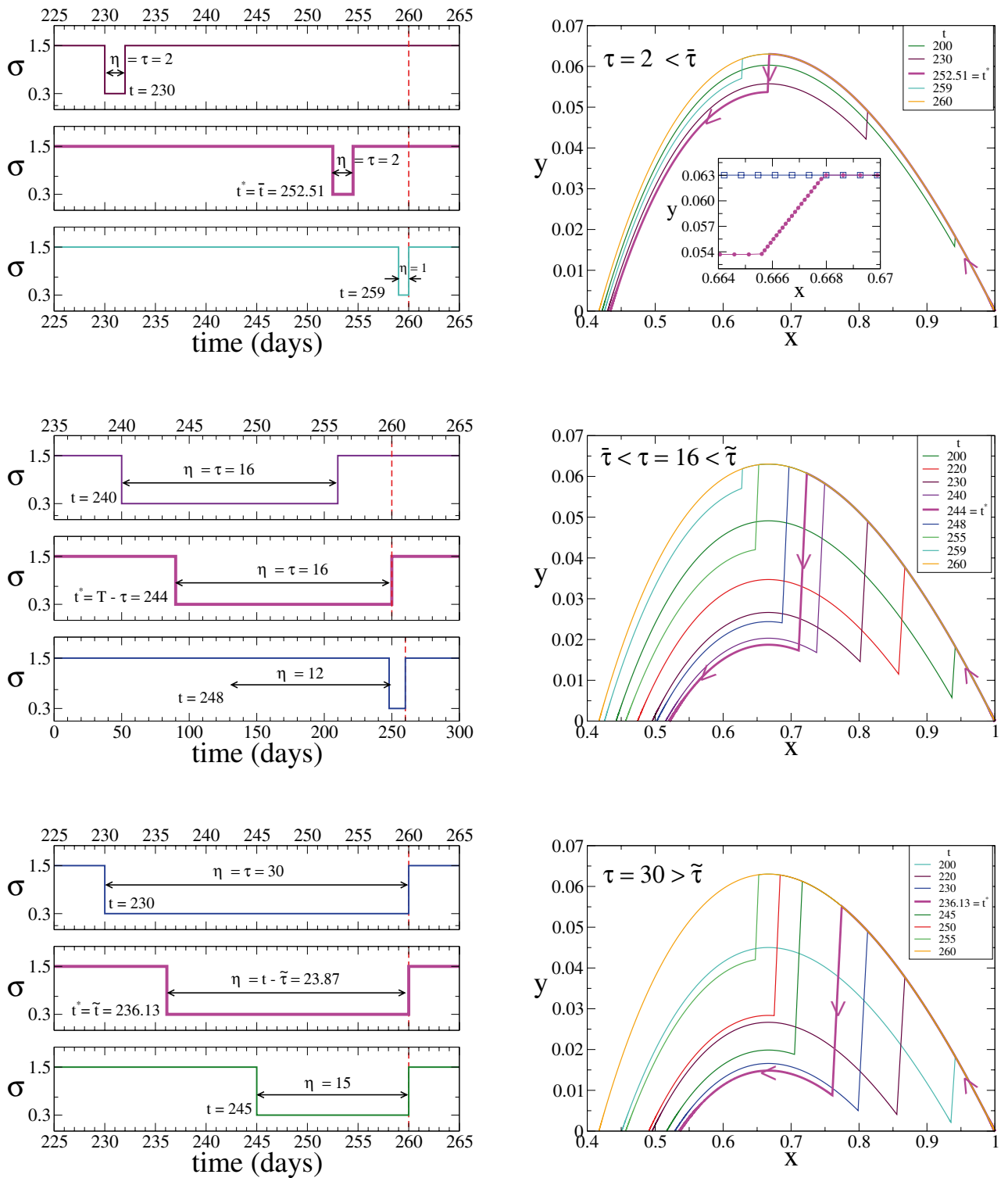


Figure 6. Case $\kappa = 0, \sigma_m = \sigma_f$ and $\sigma_s > 0$ (Theorem 2). System's trajectory in the $x - y$ phase space (right panels), for $\gamma = 0.1, \kappa = 0, \sigma_m = \sigma_f = 1.5, \sigma_s = 0.3$ and $T = 260$, and the three values of τ indicated in the legends corresponding to the different regimes of the optimum time t^* (pink lines). Left panels show the time evolution of σ for three different initial times t of the strict quarantine in each case. The optimum times are $t^* = \bar{t} \simeq 252.51$ for $\tau = 2$ (top panels), $t^* = T - \tau = 244$ for $\tau = 16$ (middle panels) and $t^* = \bar{t} \simeq 236.13$ for $\tau = 30$ (bottom panels).

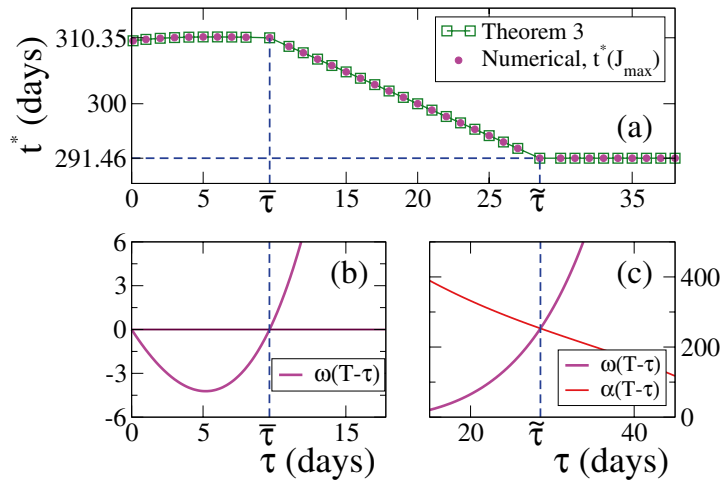


Figure 7. (a) Optimal initial time t^* vs strict quarantine length τ for $\kappa = 10^{-5}$ and $w(0) > 0$. (b) and (c) Graphical determination of the times $\bar{\tau}$ and $\tilde{\tau}$, respectively, which define the three regions for the different behaviours of t^* . The parameters are $\gamma = 0.1, \sigma_f = 2.2, \sigma_s = 0.3, \sigma_m = 1.5$ and $T = 320$. The initial condition corresponds to $x(0) = 1 - 10^{-6}, y(0) = 10^{-6}$. The optimum time for the region $\tau > \tilde{\tau} \simeq 28.54$ is $t^* \simeq 291.46$, while t^* has a slight dependence on τ for $\tau \leq \bar{\tau} \simeq 9.65$.

T , which is due to the fact that σ jumps three times in that interval: from σ_m to σ_s at time t , from σ_s to σ_m at $t + \tau$ and from σ_m to σ_f at T . This can be clearly seen in the $t^* = 310.35$ curve for $\tau = 5 < \bar{\tau}$ (inset of top-right panel of Fig. 8). For τ in the other two regions ($\tau = 18$ and 34), the strict quarantine ends at T for t^* , and thus σ jumps twice and $(x(t), y(t))$ exhibits two regimes in $[0, T]$ (insets of middle-right and bottom-right panels). As in the previously studied cases, $y(t)$ drops to the lowest level curve in the interval $[t, t + \eta]$ for the optimum time t^* (pink curves).

Discussion and conclusions

In this paper, we have studied an optimal control problem on a SIR dynamics, with a control on the reproduction number $\sigma(t)$ and a limitation in the duration of the intervention T and strict quarantine. Based on the Pontryagin’s maximum principle, we have given first order necessary conditions with an overall cost of the epidemic that takes into account both the maximization of the susceptible population in the long term (equivalently, a minimization of the ever infected population) and a penalization of the lockdown associated to a social and economic cost of the epidemic. We also point out that we have employed a novel proof to establish our analytical results. Moreover, some numerical examples have been provided to show the validity of our theoretical results.

Given a fixed time of intervention T where control strategies can be applied, and a strict quarantine period $\tau < T$ that represents the maximum time lapse for the stronger intervention, we proved that the optimal strategy is bang-bang when the term representing the socioeconomic cost of the objective functional is linear with respect to the control σ . More precisely, the optimal solution consists of switching at most twice between a mild ($\sigma = \sigma_m$) and a strict ($\sigma = \sigma_s < \sigma_m$) quarantine, where the latter lasts at most a time period τ .

Although some studies have supported the idea that a too soon or too late intervention may not minimize the total mortality, we found a broader scenario. This is because the optimal solution takes the value $\sigma = \sigma_s$ corresponding to the lockdown on an interval $[t^*, t^* + \eta] \subseteq [0, T]$, with t^* and $\eta \leq \tau$ depending on the initial fractions of susceptible and infected individuals x_0 and y_0 , respectively, and the parameters $\gamma, \tau, \sigma_f, \sigma_s, \sigma_m$ and T . In fact, we showed that, in some cases, the optimal strategy consists of taking $t^* = 0$ or $t^* + \eta = T$ (see Theorem 2 items 1 and 3-4 respectively). However, for an initial condition that corresponds to a real-life case scenario in which the percentage of the population that is infected is small when non-pharmaceutical interventions start, we obtained that the optimal strategy consists on delaying the beginning of the lockdown (items 2-4 from Theorem 2). For the case $\tau \ll T$ and $x_0 < 1/R_0$, this optimum consists in applying a mild mitigation policy ($\sigma = \sigma_m$) at the beginning of the intervention, followed by a strong suppression policy ($\sigma = \sigma_s$) and then a mild mitigation again (mild–strict–mild strategy). Here the optimum time t^* to start the strict quarantine corresponds to one that leaves the effective reproduction number $\sigma(t^* + \tau)x(t^* + \tau)$ at or just below the threshold value 1.0 when the strict quarantine is released, preventing a new outbreak. For the case $\tau \lesssim T$ the optimum corresponds to a mild–strict mitigation strategy, with a strict quarantine that starts late and lasts for a period shorter than τ . Surprisingly, it turns more effective to implement a short strict quarantine that starts late than a long strict quarantine that starts early.

We remark that these are optimal strategies within the basic SIR model defined in Eq. (2), which describe in an oversimplified manner the spread of an epidemic on an infinitely large population of individuals with homogeneous recovery, contagion and contact rates, where stochastic fluctuations due to finite-size effects are neglected. Then, stochastic fluctuations in the SIR model on finite populations may play a major role at the beginning of the epidemics if the fraction of infected individuals is relatively small, and thus starting with a strict quarantine

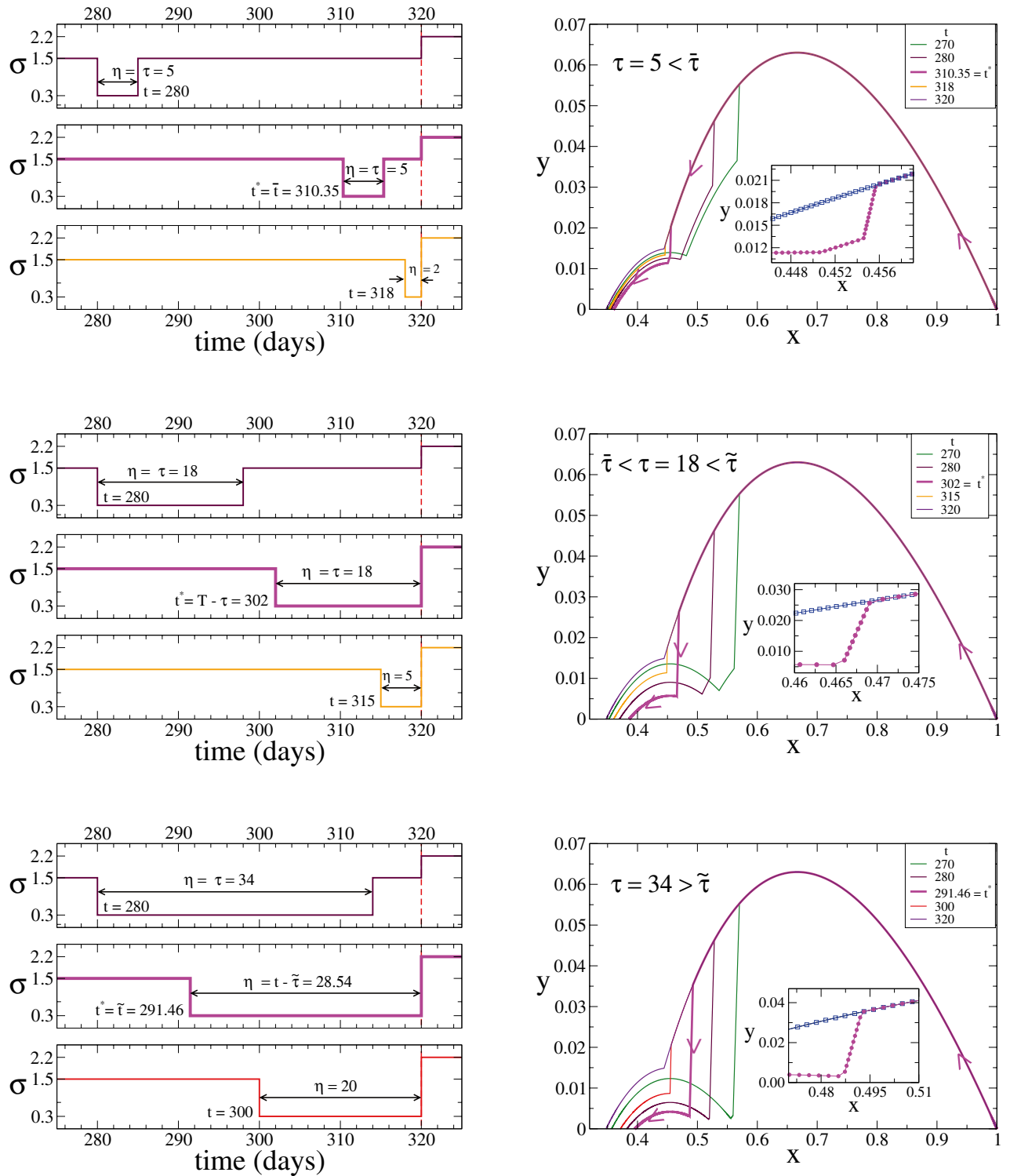


Figure 8. Case $\kappa > 0$ and $0 < \sigma_s < \sigma_m < \sigma_f$ (Theorem 3). System's trajectory in the $x - y$ phase space (right panels), for $\gamma = 0.1, \kappa = 10^{-5}, \sigma_s = 0.3, \sigma_m = 1.5, \sigma_f = 2.2$ and $T = 320$, and the three values of τ indicated in the legends corresponding to the different regimes of the optimum time t^* (pink lines). Left panels show the time evolution of σ for three different initial times t of the strict quarantine in each case. The optimum times are $t^* = \bar{t} \simeq 310.35$ for $\tau = 5$ (top panels), $t^* = T - \tau = 302$ for $\tau = 18$ (middle panels) and $t^* = \tilde{t} \simeq 291.46$ for $\tau = 34$ (bottom panels).

may prove more effective if we want to drive the epidemic to extinction. However, we expect that the results presented in this article hold in the limit of very large populations.

We have also studied the possibility of implementing intermittent quarantines, and the possibility of applying suppression measures first, followed by mitigation measures. In both cases, if the total duration of measures is limited, we have shown that they are not optimal in order to maximize the fraction of susceptible individuals at the end of the pandemic.

A major concern with respect to the current COVID-19 crisis is the possibility of an overload of available treatment resources. Since the hospitalized individuals are a fraction of the infected population, a natural objective is to keep the number of infected individuals below some threshold for all times. In a future work we intend to extend our analytic results including a running state constraint that takes this restriction into account. It might also be interesting to study the agent-based version of the SIR model, which naturally accounts for finite-size fluctuations, in order to investigate the role played by stochastic fluctuations in the different optimal strategies described above. It would be worthwhile to explore how the results are affected by the heterogeneity in recovery and infection parameters related to age and social stratum. Finally, we also aim to study the role of an underlying network of contacts, and changes in contact rates due to individual measures triggered by fear of contagion.

Methods

Formalization of the optimal control problem. As said in “Results” section, we assume that the function L [integrand of $C(\sigma)$] depends linearly on the control $\sigma(t)$. This is a simplified first approach which provides further insight into the structure of the optimal policy. In fact, under this linearity assumption we will prove that the optimal control must be bang–bang (Lemma 2), that is, the strict quarantine is turned *on* and *off*. Thus, we consider that $L(\sigma(t)) = \kappa\sigma(t)$ with $\kappa \geq 0$. In this case the functional J reads

$$J(x, y, \sigma) = x_\infty(x(T), y(T), \sigma_f) + \kappa \int_0^T \sigma(t) dt. \tag{12}$$

Moreover, we consider a restriction on the admissible controls σ that assumes that the control can take the value σ_s for at most τ time, and also takes into account a maximum economic cost that the policy maker can afford. In regard to the latter, we consider a maximum cost for imposing the strict lock down for the entire period τ . Smaller values of σ represent stricter measures and thus a larger socioeconomical cost. Therefore, we impose an inferior bound to the average of σ on $[0, T]$, meaning an upper bound for the socioeconomical cost. Thus, we consider the restriction

$$\int_0^T \sigma^*(t) dt \geq \sigma_s \tau + \sigma_m (T - \tau). \tag{13}$$

Note that any control $\sigma \in [\sigma_s, \sigma_m]$ satisfying Eq. (13) takes the value σ_s for a period no longer than τ . In fact, if σ is a control that takes the value σ_s for a longer period of time than τ , for instance $\tilde{\tau} > \tau$, then we would have that $\int_0^T \sigma(t) dt \leq \sigma_s \tilde{\tau} + \sigma_m (T - \tilde{\tau}) < \sigma_s \tau + \sigma_m (T - \tau)$ contradicting the inequality from Eq. (13).

We are now in conditions to formalize the problem of finding the optimal control that maximizes the functional J . Given $(x_0, y_0) \in \mathcal{D}$, T, τ fixed satisfying $0 < \tau < T, 0 \leq \sigma_s < \sigma_m \leq \sigma_f$, we then consider the following optimal control problem with an objective function J :

$$\max J(x, y, \sigma) := x_\infty(x(T), y(T), \sigma_f) + \int_0^T L(\sigma(t)) dt \tag{14a}$$

$$\text{s.t. } x' = -\gamma\sigma(t)xy, \quad x(0) = x_0, \quad t \in [0, T], \tag{14b}$$

$$y' = \gamma\sigma(t)xy - \gamma y, \quad y(0) = y_0, \quad t \in [0, T], \tag{14c}$$

$$\sigma_s \leq \sigma(t) \leq \sigma_m \tag{14d}$$

$$\int_0^T \sigma(t) dt \geq \sigma_s \tau + \sigma_m (T - \tau) \tag{14e}$$

In what follows we compute the partial derivatives of $x_\infty(x(t), y(t), \sigma)$ with respect to $x(t)$ and $y(t)$ in the same way that it is done in²³. We begin reviewing the solution of the SIR model without control Eq. (1) as done by²³. It can be shown¹ that $x(t)$ satisfies $x(t)e^{\sigma z(t)} = x_0 e^{\sigma z_0}$ which combined with the identity $z(t) = 1 - x(t) - y(t)$ implies that

$$\mu(x(t), y(t), \sigma) := x(t)e^{-\sigma(x(t)+y(t))}$$

is constant in time for any solution of Eq. (1). The trajectories in Fig. 1 are also contour lines of μ . Since $y_\infty = 0$, we then have that $x_\infty = x_0 e^{\sigma(x_\infty - x_0 - y_0)} = \mu(x_0, y_0, \sigma) e^{\sigma x_\infty}$. Then $w = -\sigma x_\infty$ satisfies the equation $w e^w = -\sigma \mu(x_0, y_0, \sigma)$ and therefore $w = W_0(-\sigma \mu(x_0, y_0, \sigma))$ where W_0 is the principal branch of Lambert’s W -function³⁷, and thus for any $(x, y) \in \mathcal{D}$

$$x_\infty(x, y, \sigma) = -\frac{1}{\sigma} W_0(\sigma \mu(x, y, \sigma)).$$

From this expression we can compute the partial derivatives of $x_\infty(x(t), y(t), \sigma)$ with respect to $x(t)$ and $y(t)$ in the same way that it is done in²³.

$$\frac{\partial x_\infty(x(t), y(t), \sigma)}{\partial x(t)} = \frac{1 - \sigma x(t)}{x(t)} \frac{x_\infty(x(t), y(t), \sigma)}{1 - \sigma x_\infty(x(t), y(t), \sigma)}, \tag{15a}$$

$$\frac{\partial x_\infty(x(t), y(t), \sigma)}{\partial y(t)} = -\frac{\sigma x_\infty(x(t), y(t), \sigma)}{1 - \sigma x_\infty(x(t), y(t), \sigma)}. \tag{15b}$$

In order to solve problem Eq. (14) by means of the Pontryagin’s Maximum Principle, we add a new state variable given by $v(t) = \int_0^t \sigma(s) ds$ and consider $\sigma : [0, T] \rightarrow [\sigma_s, \sigma_m]$ in the class of Lebesgue-measurable functions (so that we have an existence result for optimal solution). Thus, we can study the equivalent optimal control problem:

$$\max J(x, y, v, \sigma) := x_\infty(x(T), y(T), \sigma_f) + \int_0^T L(\sigma(t)) dt \tag{16a}$$

$$s.t. \quad x'(t) = -\gamma \sigma(t) x(t) y(t), \quad x(0) = x_0, \quad t \in [0, T], \tag{16b}$$

$$y'(t) = \gamma \sigma(t) x(t) y(t) - \gamma y(t), \quad y(0) = y_0, \quad t \in [0, T], \tag{16c}$$

$$v'(t) = \sigma(t), \quad v(0) = 0, \quad t \in [0, T], \tag{16d}$$

$$\sigma(t) \in [\sigma_s, \sigma_m], \quad \text{a.e. } t \in [0, T] \tag{16e}$$

$$v(T) \geq \sigma_m T + (\sigma_s - \sigma_m) \tau. \tag{16f}$$

We will refer to a 4-tuple (x, y, v, σ) as an admissible process of the underlying control system if the control σ is a measurable function and the state (x, y, v) is an absolutely continuous vector function satisfying Eqs. (16b)–(16f). The optimal control problem consists in finding an optimal admissible process $(x^*, y^*, v^*, \sigma^*)$ that maximizes the cost J . In this case, we refer to the control σ^* as optimal control.

Next, we give a result on existence of solution for the optimal control problem Eq. (16).

Proposition 1 *The optimal control problem Eq. (16) admits a solution.*

Proof Since $L(\sigma) = \kappa \sigma$ is continuous, convex and satisfies that there exists a constant α_0 such that for all $\sigma \in [\sigma_s, \sigma_m]$ it holds $L(\sigma) \geq \alpha_0$, the proof follows directly from Theorem 23.11²⁴. Using that the control space is closed, solutions of the system of Eqs. (16b), (16c) satisfy that $0 \leq x + y \leq 1$ and the application $x_\infty(x, y)$ is continuous, it is straightforward to prove conditions (a) to (f) from Theorem 23.11. Moreover, taking $\sigma(t) \equiv \sigma_m$, we see that the unique solution of the system of Eqs. (16b)–(16d) together with σ gives an admissible process for which J is finite completing the hypothesis of Theorem 23.11. \square

The optimal control is bang-bang. In what follows, we derive the necessary conditions for problem Eq. (16) where J is given by Eq. (12). We consider the Hamiltonian H

$$H(x, y, v, \sigma, \lambda) = \lambda_0 L(\sigma) - \lambda_1(\gamma \sigma x y) + \lambda_2(\gamma \sigma x y - \gamma y) + \lambda_3 \sigma \tag{17}$$

where $\lambda_0 \geq 0, \lambda \in \mathbb{R}^3$. The necessary conditions for a maximum process $(x^*, y^*, v^*, \sigma^*)$ on $[0, T]$ are the following^{24,25}: There exists a real number $\lambda_0 \geq 0$, the adjoint variable $\lambda : [0, T] \rightarrow \mathbb{R}^3$ which is absolutely continuous, and $\beta \in \mathbb{R}$ such that $(\lambda_0, \lambda(t), \beta) \neq 0$ for every t and the following conditions hold:

1. The adjoint variables $\lambda_1(t), \lambda_2(t)$ satisfy a.e. $t \in [0, T]$

$$\lambda_1'(t) = (\lambda_1(t) - \lambda_2(t)) \gamma \sigma(t) y(t), \tag{18a}$$

$$\lambda_2'(t) = (\lambda_1(t) - \lambda_2(t)) \gamma \sigma(t) x(t) + \gamma \lambda_2(t), \tag{18b}$$

with final time conditions (using the abbreviation x_∞ for $x_\infty(x(T), y(T), \sigma_f)$)

$$\lambda_1(T) = \lambda_0 \frac{\partial x_\infty}{\partial x(T)} = \lambda_0 \frac{1 - \sigma_f x(T)}{x(T)} \frac{x_\infty}{1 - \sigma_f x_\infty}, \tag{19a}$$

$$\lambda_2(T) = \lambda_0 \frac{\partial x_\infty}{\partial y(T)} = -\lambda_0 \frac{\sigma_f x_\infty}{1 - \sigma_f x_\infty}, \tag{19b}$$

and the adjoint variable λ_3 satisfies

$$\lambda_3'(t) = 0, \tag{20}$$

$$\lambda_3(T) = \beta \geq 0 \text{ and } \lambda_3(T)(v(T) - \sigma_m(T - \tau) - \sigma_s \tau) = 0 \tag{21}$$

obtaining $\lambda_3(t) = \beta$ for all $t \in [0, T]$.

2. For a.e. $t \in [0, T]$

$$\begin{aligned} & \lambda_0 L(\sigma^*(t)) + \sigma^*(t) (\gamma x^*(t) y^*(t) (\lambda_2(t) - \lambda_1(t)) + \beta) \\ &= \max_{\sigma: \sigma_s \leq \sigma \leq \sigma_m} \lambda_0 L(\sigma) + \sigma (\gamma x^*(t) y^*(t) (\lambda_2(t) - \lambda_1(t)) + \beta). \end{aligned}$$

Using that $L(\sigma) = \kappa \sigma$ and defining

$$\phi(t) = \lambda_0 \kappa + \beta + \gamma x^*(t) y^*(t) (\lambda_2(t) - \lambda_1(t)), \tag{22}$$

we obtain for a.e. $t \in [0, T]$,

$$\sigma^*(t) \phi(t) = \max_{\sigma: \sigma_s \leq \sigma \leq \sigma_m} \sigma \phi(t). \tag{23}$$

3. There exists a constant C such that for a.e. $t \in [0, T]$

$$\lambda_0 L(\sigma^*(t)) + \sigma^*(t) (\gamma x^*(t) y^*(t) (\lambda_2(t) - \lambda_1(t)) + \beta) - \gamma \lambda_2(t) y^*(t) = C,$$

thus, for all $t \in [0, T]$

$$\sigma^*(t) \phi(t) - \gamma \lambda_2(t) y^*(t) = C. \tag{24}$$

We have the following result:

Lemma 1 *The optimal control problem is normal (that is, the multiplier $\lambda_0 \neq 0$).*

Proof Assume $\lambda_0 = 0$. From Eqs. (18a), (18b) with final time conditions $\lambda_1(T) = \lambda_2(T) = 0, \lambda_1(t) = \lambda_2(t) = 0$ for all $t \in [0, T]$. Since the multipliers $(\lambda_0, \lambda(t), \beta) \neq 0$, then $\lambda_3(t) \equiv \beta > 0$ yielding $\phi(t) \equiv \kappa + \beta > 0$. Therefore, from the optimality condition given in Eq. (23), $\sigma^*(t) = \sigma_m$ a.e. $t \in [0, T]$ contradicting the complementarity condition $v(T) = \sigma_m(T - \tau) + \sigma_s \tau$ given in Eq. (21). Thus, we can assume $\lambda_0 = 1$, and the proof is finished. \square

Lemma 2 *Let $L(\sigma(t)) = \kappa \sigma(t)$ with $\kappa \geq 0$ and let σ^* be an optimal control of problem Eq. (16). Then $\sigma^*(t)$ is a bang-bang control.*

Proof Assume $\phi(t) = 0$ on an interval $[a, b] \subset [0, T]$, then computing its derivative we obtain

$$0 = -\gamma x^*(t) y^*(t) \lambda_1(t) \text{ for all } t \in (a, b).$$

Thus, from Eq. (18a) $\lambda_1(t) = \lambda_2(t) = 0$ for all $t \in (a, b)$ and therefore for all $t \in [0, T]$, contradicting the end point conditions. Then, there cannot be singular arcs and the control σ^* is given by:

$$\sigma^*(t) = \begin{cases} \sigma_m & \text{if } \phi(t) > 0 \\ \sigma_s & \text{if } \phi(t) < 0. \end{cases} \tag{25}$$

\square

Lemma 3 *Let (x_0, y_0) be given and (x, y, v, σ) an admissible process. Then for $t \geq 0$*

$$x_\infty(x(t), y(t), \sigma_f) \geq x_\infty(x_0, y_0, \sigma_f),$$

and therefore

$$x_\infty(x_0, y_0, \sigma_f) \leq x_\infty(x(T), y(T), \sigma_f) < 1/\sigma_f. \tag{26}$$

Proof See²³.

In the next lemma we will see that the switching function changes sign at most two times, concluding that an optimal control σ^* jumps at most twice. \square

Lemma 4 *The switching function ϕ given in Eq. (22) changes sign at most twice.*

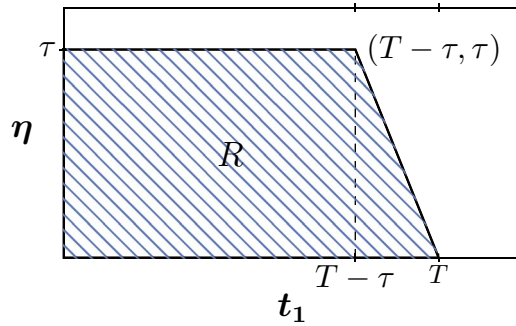


Figure 9. Graphic of set R.

Proof The proof follows by analysing the phase diagram of λ_1, λ_2 . We begin by noting that a solution (λ_1, λ_2) of the system of Eqs. (18a), (18b) cannot cross both semilines $\lambda_1 = \lambda_2 > 0$ and $\lambda_1 = \lambda_2 < 0$. This is a consequence of condition Eq. (24). In fact, assume there exist $s_1, s_2 \in [0, T]$ such that $\lambda_1(s_1) = \lambda_2(s_1) > 0$ and $\lambda_1(s_2) = \lambda_2(s_2) < 0$. Evaluating Eq. (22) on $t = s_i$ for $i = 1, 2$ we have that $\phi(s_i) = \kappa + \lambda_3(T) \geq 0$ and from Eq. (24), $\lambda_2(s_i) = -\frac{C}{\gamma y^*(s_i)}$ if $\kappa + \lambda_3(T) = 0$ or, using Eq. (25), $\lambda_2(s_i) = \frac{\sigma_m(\kappa + \lambda_3(T)) - C}{\gamma y^*(s_i)}$ if $\kappa + \lambda_3(T) > 0$, both cases contradicting that $\lambda_2(s_1)$ and $\lambda_2(s_2)$ had opposite signs.

Since we have end time conditions on T we go backwards from $(\lambda_1(T), \lambda_2(T))$ with $\lambda_1(T) > 0$ and $\lambda_2(T) < 0$. From Eq. (18a), for $\lambda_1 < \lambda_2, \lambda_1' < 0$, thus λ_1 is decreasing and for the semiplane $\lambda_1 > \lambda_2$ we have that λ_1 is increasing. Also, from Eq. (18b), for $\lambda_1 = \lambda_2 > 0$ we have that λ_2 is increasing and for $\lambda_1 = \lambda_2 < 0, \lambda_2$ is decreasing. Finally, from Eqs. (18a) and (18b), for $\lambda_2 = 0$ and $\lambda_1 > 0$, both λ_1 and λ_2 are increasing.

Thus, since the end time conditions are on the region of the phase diagram with $\lambda_1 > \lambda_2$ and $\lambda_2 < 0$ we have that the solution backwards in time moves to the left where λ_1 decreases and λ_2 keeps being negative. At some point in time it could cross the semiline $\lambda_1 = \lambda_2 < 0$ (note that there cannot be touch points). If the solution crosses this line it cannot cross the semiline $\lambda_1 = \lambda_2 > 0$ for a previous time and thus it stays on the region $\lambda_1 < \lambda_2$ for all previous times.

From the definition of ϕ Eq. (22), for $\lambda_1 < \lambda_2$, we have $\phi > 0$. For $\lambda_1 = \lambda_2, \phi = \kappa + \lambda_3(T) \geq 0$ and for $\lambda_1 > \lambda_2, \phi$ could become negative. Since $\phi'(t) = \gamma x^*(t) y^*(t) \lambda_1(t)$, we see that for $t \in [0, T]$ such that $(\lambda_1(t), \lambda_2(t))$ is on the region $\lambda_1 > \lambda_2$ the function ϕ decreases for such t 's with $\lambda_1(t) < 0$ and increases for $\lambda_1(t) > 0$. Also, let $s_0, s_1 \in [0, T]$ such that $s_0 < s_1, \lambda_1(s_0) = \lambda_2(s_0) < 0$ and $\lambda_2(s_1) < \lambda_1(s_1) = 0$, then $\phi(s_0) = \kappa + \lambda_3(T), \phi$ reaches the minimum value on $[t_0, T]$ at s_1 and $\phi(T) = \kappa + \lambda_3(T) - \frac{\gamma y^*(T) x_\infty}{1 - \sigma_f x_\infty} < \phi(s_0)$.

Thus, we conclude that ϕ has at most two zeros on $[0, T]$ and the proof is finished. □

Theorem 1 Let $(x^*, y^*, v^*, \sigma^*)$ be an optimal process, then:

$$\sigma^*(t) = \begin{cases} \sigma_m & \text{for } 0 \leq t < t_1 \\ \sigma_s & \text{for } t_1 \leq t < t_1 + \eta \\ \sigma_m & \text{for } t_1 + \eta \leq t \leq T \end{cases} \tag{27}$$

with $0 \leq \eta \leq \tau$.

Proof Since the optimal control must be bang-bang satisfying Eq. (25), from Lemma 4 it has at most two jumps and from Eq. (13), it takes the value σ_s at most for τ time. Thus the proof is completed.

As a consequence of Theorem 1 we have that if $(x^*, y^*, v^*, \sigma^*)$ is an optimal process, then the optimal control σ^* is a piecewise constant function having at most two jumps and therefore its unique associated state (x^*, y^*, v^*) is a piecewise continuously differentiable function.

Characterization of the optimal control. In this section we will give the main Theorem of the article, that characterizes the switching times t_1 and $t_1 + \eta$ (where t_1 is the beginning of the lockdown and η is its duration) for an optimal control.

Let us consider the compact set (see Fig. 9)

$$R = \{(t_1, \eta) \in \mathbb{R}^2 : 0 \leq \eta \leq \tau, 0 \leq t_1 \leq T - \eta\}.$$

Given $(t_1, \eta) \in R$, for simplicity of notation, we will denote $t_0 = 0, t_2 = t_1 + \eta, t_3 = T$.

Moreover, given $(t_1, \eta) \in R$, we will denote the solution of equation Eqs. (16b), (16c) for $s \geq t$ by

$$\Psi(s, t, x, y, \sigma), \text{ with } \sigma \in \{\sigma_s, \sigma_m\} \text{ and initial data } (x, y) \in \mathcal{D} \text{ at time } t, \tag{28}$$

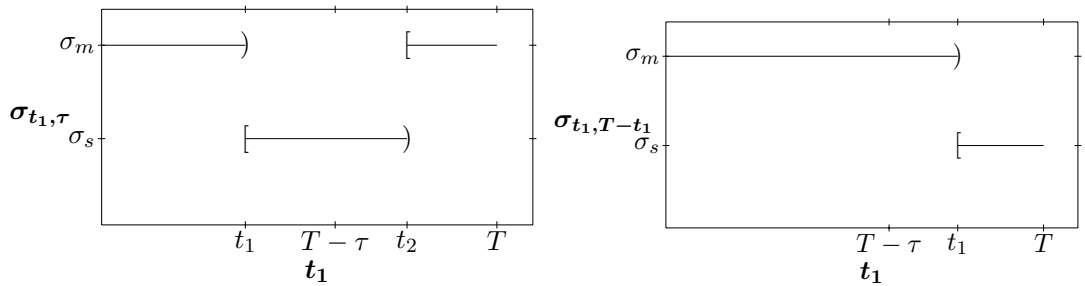


Figure 10. Control for $(t_1, \eta) \in P$. In (a) we have a mild-strict-mild quarantine in the intervention interval $[0, T]$. In (b) we have a mild-strict quarantine in the intervention interval $[0, T]$.

and $(x_1, y_1) = \Psi(t_1, t_0, x_0, y_0, \sigma_m), (x_2, y_2) = \Psi(t_2, t_1, x_1, y_1, \sigma_s)$. Then, if we call $(x_{t_1, \eta}, y_{t_1, \eta})$, the solution of equation Eqs. (16b), (16c) associated to the control σ given by equation Eq. (27) with initial data $(x(0), y(0)) = (x_0, y_0)$, we have that

$$(x_{t_1, \eta}(s), y_{t_1, \eta}(s)) = \begin{cases} \Psi(s, t_0, x_0, y_0, \sigma_m) & \text{for } 0 \leq s \leq t_1 \\ \Psi(s, t_1, x_1, y_1, \sigma_s) & \text{for } t_1 < s \leq t_2 \\ \Psi(s, t_2, x_2, y_2, \sigma_m) & \text{for } t_2 < s \leq T. \end{cases} \tag{29}$$

From Theorem 1 we need to determine the maximum of the function

$$J(t_1, \eta) = x_\infty(x_{t_1, \eta}(T), y_{t_1, \eta}(T), \sigma_f) + \kappa(\sigma_s \eta + \sigma_m(T - \eta)) \tag{30}$$

on the compact set R .

In order to do that, we need to compute the derivatives of J .

After some computations (see Eqs. (73) and (76) from Appendix) we obtain that

$$\frac{\partial J}{\partial t_1}(t_1, \eta) = \left[\frac{\gamma^2(\sigma_m - \sigma_s)y_{t_1, \eta}(T)y_{t_1, \eta}(t_1)x_{\infty, t_1, \eta}}{(1 - \sigma_f x_{\infty, t_1, \eta})} \right] \cdot \left[\int_{t_1}^{t_2} \frac{\sigma_f x_{t_1, \eta}(r) - 1}{y_{t_1, \eta}(r)} dr - \gamma y_2 \int_{t_2}^T \frac{\sigma_f x_{t_1, \eta}(r) - 1}{y_{t_1, \eta}(r)} dr \int_{t_1}^{t_2} \frac{\sigma_m x_{t_1, \eta}(r) - 1}{y_{t_1, \eta}(r)} dr \right] \tag{31}$$

and

$$\begin{aligned} \frac{\partial J}{\partial \eta}(t_1, \eta) &= \frac{\gamma x_{\infty, t_1, \eta}}{1 - \sigma_f x_{\infty, t_1, \eta}} y_{t_1, \eta}(t_2)(\sigma_m - \sigma_s) \left(1 - (\sigma_f - \sigma_m)y_{t_1, \eta}(T)\gamma \int_{t_2}^T \frac{x_{t_1, \eta}(r)}{y_{t_1, \eta}(r)} dr \right) \\ &\quad - \kappa(\sigma_m - \sigma_s) \\ &= \frac{\gamma x_{\infty, t_1, \eta}}{1 - \sigma_f x_{\infty, t_1, \eta}} (\sigma_m - \sigma_s)y_{t_1, \eta}(T) \left(1 - \gamma y_{t_1, \eta}(t_2) \int_{t_2}^T \frac{\sigma_f x_{t_1, \eta}(r) - 1}{y_{t_1, \eta}(r)} dr \right) \\ &\quad - \kappa(\sigma_m - \sigma_s) \end{aligned} \tag{32}$$

where $x_{\infty, t_1, \eta} = x_\infty(x_{t_1, \eta}(T), y_{t_1, \eta}(T), \sigma_f)$. Note that for $(t_1, \eta) \in R, \frac{\partial J}{\partial \eta}(t_1, \eta) > 0$ if and only if

$$\frac{x_{\infty, t_1, \eta}}{1 - \sigma_f x_{\infty, t_1, \eta}} \gamma y_{t_1, \eta}(T) \left(1 - \gamma y_{t_1, \eta}(t_2) \int_{t_2}^T \frac{\sigma_f x_{t_1, \eta}(r) - 1}{y_{t_1, \eta}(r)} dr \right) > \kappa. \tag{33}$$

In the results given in the next sections we will assume that $\frac{\partial J}{\partial \eta}(t_1, \eta) > 0$ for all $(t_1, \eta) \in R$ and therefore in the following remark we analyse the derivatives of J restricted to the superior border of R (see Eq. (37) and (38)) which will be used later.

Remark 2 Assume that $\frac{\partial J}{\partial \eta}(t_1, \eta) > 0$ for all $(t_1, \eta) \in R$, then the maximum value of J on R must be attained at the superior border

$$P = \{(t_1, \tau), t_1 \in [0, T - \tau]\} \cup \{(t_1, T - t_1), t_1 \in [T - \tau, T]\}. \tag{34}$$

Thus, in this case, for (t_1, τ) with $t_1 \in [0, T - \tau)$ we have $t_2 = t_1 + \tau$ and for $(t_1, T - t_1)$ with $t_1 \in [T - \tau, T]$ we have $t_2 = T$, and the control is as in Fig. 10

Following, we define a continuous function w for $t \in [0, T]$ which is the second factor in equation Eq. (31) and thus gives information on the monotonicity of J (see also Eqs. (40) and (41) below).

$$w(t) = \begin{cases} \int_t^{t+\tau} \frac{\sigma_f x_{t, \tau}(r) - 1}{y_{t, \tau}(r)} dr - \gamma y_{t, \tau}(t + \tau) \int_{t+\tau}^T \frac{\sigma_f x_{t, \tau}(r) - 1}{y_{t, \tau}(r)} dr \int_t^{t+\tau} \frac{\sigma_m x_{t, \tau}(r) - 1}{y_{t, \tau}(r)} dr & \text{for } 0 \leq t \leq T - \tau \\ \int_t^T \frac{\sigma_f x_{t, T-t}(r) - 1}{y_{t, T-t}(r)} dr & \text{for } T - \tau < t \leq T, \end{cases} \tag{35}$$

and for $t \in [T - \tau, T]$, we define

$$\alpha(t) = \frac{1}{\gamma y_{t,T-t}(t)} \left(1 - \kappa \frac{1 - \sigma_f x_{\infty,t,T-t}}{\gamma y_{t,T-t}(T) x_{\infty,t,T-t}} \right). \tag{36}$$

Then we have that for $(t_1, \eta) \in P$

$$\frac{\partial J}{\partial t_1}(t_1, \eta) = \frac{x_{\infty,t_1,\eta}}{(1 - \sigma_f x_{\infty,t_1,\eta})} \gamma^2 (\sigma_m - \sigma_s) y_{t_1,\eta}(T) y_{t_1,\eta}(t_1) w(t_1), \tag{37}$$

and for $(t_1, T - t_1) \in P$ with $t_1 \in [T - \tau, T]$

$$\begin{aligned} \frac{\partial J}{\partial \eta}(t_1, T - t_1) &= \frac{x_{\infty,t_1,\eta}}{1 - \sigma_f x_{\infty,t_1,\eta}} \gamma (\sigma_m - \sigma_s) y_{t_1,\eta}(T) - \kappa (\sigma_m - \sigma_s) \\ &= \frac{x_{\infty,t_1,\eta}}{1 - \sigma_f x_{\infty,t_1,\eta}} \gamma^2 (\sigma_m - \sigma_s) y_{t_1,\eta}(T) y_{t_1,\eta}(t_1) \alpha(t_1). \end{aligned} \tag{38}$$

We consider \tilde{J} the continuous function defined as the restriction of $J(t_1, \eta)$ to P , that is

$$\tilde{J}(t_1) = \begin{cases} J(t_1, \tau) & \text{for } t_1 \in [0, T - \tau], \\ J(t_1, T - t_1) & \text{for } t_1 \in [T - \tau, T]. \end{cases} \tag{39}$$

From Eq. (37) we have that

$$\tilde{J}'(t_1) = \gamma^2 (\sigma_m - \sigma_s) \frac{x_{\infty,t_1,\tau}}{(1 - \sigma_f x_{\infty,t_1,\tau})} y_{t_1,\tau}(t_1) y_{t_1,\tau}(T) w(t_1) \quad \text{for } t_1 \in [0, T - \tau), \tag{40}$$

and from Eqs. (37) and (38), we get

$$\begin{aligned} \tilde{J}'(t_1) &= \frac{dJ}{dt_1}(t_1, T - t_1) - \frac{dJ}{d\eta}(t_1, T - t_1) \\ &= \gamma^2 (\sigma_m - \sigma_s) \frac{x_{\infty,t_1,T-t_1}}{1 - \sigma_f x_{\infty,t_1,T-t_1}} y_{t_1,T-t_1}(t_1) y_{t_1,T-t_1}(T) (w(t_1) - \alpha(t_1)) \end{aligned} \tag{41}$$

for $t_1 \in (T - \tau, T]$.

In the next subsection, we prove the main result of this article (Theorem 2) for the case $\sigma_m = \sigma_f, \sigma_s \geq 0$ and $\kappa = 0$. Then, we derive the result for $\sigma_s = 0$ (Corollary 1) in order to compare our result with the one obtained in²³.

Case $\sigma_m = \sigma_f$ and $\kappa = 0$. For $\sigma_m = \sigma_f$, from Eq. (74) with $i = 2$, we obtain

$$w(t) = \begin{cases} \frac{y_{t,\tau}(t + \tau)}{y_{t,\tau}(T)} \int_t^{t+\tau} \frac{\sigma_f x_{t,\tau}(r) - 1}{y_{t,\tau}(r)} dr & \text{for } 0 \leq t \leq T - \tau \\ \int_t^T \frac{\sigma_f x_{t,T-t}(r) - 1}{y_{t,T-t}(r)} dr & \text{for } T - \tau < t \leq T. \end{cases} \tag{42}$$

In addition, from Eqs. (32) and (74) and using $\kappa = 0$ we have that

$$\frac{\partial J}{\partial \eta}(t_1, \eta) = \frac{x_{\infty,t_1,\eta}}{1 - \sigma_f x_{\infty,t_1,\eta}} \gamma (\sigma_f - \sigma_s) y_{t_1,\eta}(t_1 + \eta) > 0 \tag{43}$$

for all $(t_1, \eta) \in R$.

In this case, the sign and zeros of $w(t)$ for $t \in [0, T - \tau]$ are the same as those for the function

$$z(t) = \int_t^{t+\tau} \frac{\sigma_f x_{t,\tau}(r) - 1}{y_{t,\tau}(r)} dr, \tag{44}$$

where $z(t)$ can be interpreted as the average on the time window $[t, t + \tau]$ of the difference between the effective reproduction number $\sigma_f x_{t,\tau}(r)$ and the threshold 1.0, weighted by the inverse of $y_{t,\tau}$. Looking at the phase diagram of Fig. 1 we see that the trajectories travel through the contour line $\mu(x_0, y_0, \sigma_m)$ until the strict lockdown is activated, and then descends for the time the lock down lasts (always less than τ) to another contour line of μ with σ_m . For $\kappa = 0$, the zero of $w(t)$ (t^*) captures the moment when this travel to a lower contour line is faster, in the sense that leaves the trajectory on the lowest possible contour line at the end of the strict lockdown and, therefore, at the maximum of x_{∞} . The function $z(t)$ can also be seen as an external parameter that becomes zero at the optimal time t^* for which J reaches its maximum, i.e., $z(t^*) = 0$. In the next remark we discuss the sign of $z(t)$ for $t \in [0, T - \tau]$ when $\sigma_m = \sigma_f$.

Remark 3 Given $(x_0, y_0) \in \mathcal{D}$ with $x_0 > 1/\sigma_f$, assume there exists $s_1 \in [0, T - \tau]$ such that the solution $\Psi_1(s_1, 0, x_0, y_0, \sigma_m) = 1/\sigma_f$, that is $x_{s_1,\tau}(s_1) = 1/\sigma_f$ (red line in Fig. 11). Then, for $t \in [s_1, T - \tau]$, $x_{t,\tau}(t) \leq 1/\sigma_f$

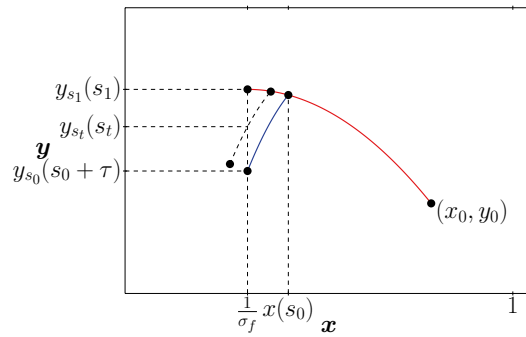


Figure 11. Trajectories $(x_{t,\tau}, y_{t,\tau})$ for $t \in [s_0, s_1]$ when $\sigma_m = \sigma_f$.

and therefore $x_{t,\tau}(r) < 1/\sigma_f$ for $r \in (t, t + \tau]$ implying $z(t) < 0$. Additionally, assume there exists $s_0 \in [0, T - \tau]$ such that $x_{s_0,\tau}(s_0 + \tau) = 1/\sigma_f$ (blue line). In this case, it is clear that $s_0 < s_1$ and also that for all $t \in (s_0, s_1)$ there exists a unique $s_t \in [t, t + \tau]$ such that $x_{t,\tau}(s_t) = 1/\sigma_f$. Moreover we can conclude that for $t \in [0, s_0]$, $x_{t,\tau}(r) \geq 1/\sigma_f$ for all $r \in (t, t + \tau)$ and therefore $z(t) > 0$. If s_1 defined before does not exist, that is $x_{s,\tau}(s) > 1/\sigma_f$ for all $s \in [0, T - \tau]$, then we take $s_1 = T - \tau$. Likewise, if s_0 does not exist, that is for all $s \in [0, T - \tau]$, $x_s(s + \tau) < 1/\sigma_f$, then we take $s_0 = 0$ and conclude that in either case, the sign of $z(t)$ for $t \geq 0$, is determined in the complement of $[s_0, s_1]$.

In the next lemma we prove that z is a decreasing function on the interval (s_0, s_1) introduced in Remark 3.

Lemma 5 Let $s_0, s_1 \in [0, T - \tau]$ be given by Remark 3. Then the function z defined in Eq. (44) is decreasing on (s_0, s_1) . Moreover, $z(t) > 0$ for $t < s_0$, $z(t) < 0$ for $t > s_1$ and consequently w changes sign at most once on $[0, T - \tau]$.

Proof Since the duration of the strict quarantine is τ fixed, for simplicity of notation in this proof we neglect the subindex τ from solutions x and y . Given $t \in (s_0, s_1)$, we define the auxiliary functions

$$g_t(s) = \sigma_f \sigma_s x_t(s) y_t(s) + (\sigma_f x_t(s) - 1)(\sigma_s x_t(s) - 1), \tag{45a}$$

$$f_t(s) = \frac{\sigma_f x_t(s) - 1}{y_t(s)}, \tag{45b}$$

$$i_t(s) = \frac{x_t(s)}{y_t(s)} (\sigma_f x_t(s) + \sigma_f y_t(s) - 1) + \gamma g_t(s) \int_t^s \frac{x_t(r)}{y_t(r)} dr. \tag{45c}$$

By computing the derivative for $s \in (t, t + \tau)$ we obtain

$$g'_t(s) = -\gamma \sigma_s^2 x_t(s) y_t(s) (\sigma_f x_t(s) + \sigma_f y_t(s) - 1), \tag{46a}$$

$$f'_t(s) = -\gamma \frac{g_t(s)}{y_t(s)}, \tag{46b}$$

$$i'_t(s) = -\gamma \sigma_s x_t(s) \left(\gamma \sigma_s y_t(s) \int_t^s \frac{x_t(r)}{y_t(r)} dr + 1 \right) (\sigma_f x_t(s) + \sigma_f y_t(s) - 1), \tag{46c}$$

and we have that

$$\begin{aligned} z'(t) &= f_t(t + \tau) - f_t(t) \\ &+ \gamma (\sigma_s - \sigma_f) y_1 \int_t^{t+\tau} \left[\frac{x_t(s)}{y_t^2(s)} (\sigma_f x_t(s) + \sigma_f y_t(s) - 1) + \gamma \left(\int_t^s \frac{x_t(r)}{y_t(r)} dr \right) \frac{g_t(s)}{y_t(s)} \right] ds \\ &= f_t(t + \tau) - f_t(t) + \gamma (\sigma_s - \sigma_f) y_1 \int_t^{t+\tau} \frac{i_t(s)}{y_t(s)} ds \end{aligned} \tag{47}$$

Note that both g'_t and i'_t have the opposite sign of $(\sigma_f x_t(s) + \sigma_f y_t(s) - 1)$. First, assume $\sigma_f x_t(s) + \sigma_f y_t(s) - 1 > 0$ for all $s \in (t, t + \tau)$, then from Eq. (46a), g_t is a decreasing function. Moreover, since $\sigma_s x_t(t + \tau) - 1 < 0$ and $\sigma_f x_t(t + \tau) - 1 < 0$ for $t \in (s_0, s_1)$, we deduce that $g_t(s) > g_t(t + \tau) > 0$. In addition, from Eq. (45c) we obtain i_t is positive.

On the other side, from the fact that $x + y$ is a decreasing function, if we assume that there exists $s_3 \in [t, t + \tau]$ such that $\sigma_f x_t(s_3) + \sigma_f y_t(s_3) = 1$ we have that for $s < s_3$, $\sigma_f x_t(s) + \sigma_f y_t(s) - 1 > 0$ and for

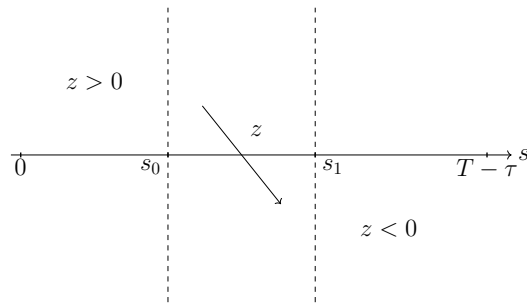


Figure 12. Behaviour of $z(t)$ when $\sigma_f = \sigma_m$.

$s > s_3$, $\sigma_f x_t(s) + \sigma_f y_t(s) - 1 < 0$. Therefore, g_t and i_t attains a global minimum on $[t, t + \tau]$ at s_3 and thus $g_t(s) \geq g_t(s_3) = \sigma_f y_t(s_3) > 0$ and $i_t(s) \geq i_t(s_3) = \gamma g_t(s_3) \int_t^{s_3} \frac{x_t(r)}{y_t(r)} dr > 0$ for all $s \in [t, t + \tau]$. Furthermore, from Eq. (46b) we have that $f_t(s)$ is also a decreasing function.

Thus, we have proved that for $t \in (s_0, s_1)$, $f_t(s)$ is decreasing on $(t, t + \tau)$ and $i_t(s) > 0$ for all $s \in [t, t + \tau]$, yielding from Eq. (47) that $z'(t) < 0$ for all $t \in (s_0, s_1)$. Finally, from Remark 3 we deduce that z changes sign at most once on $[0, T - \tau]$ (see Fig. 12) concluding that w changes sign at most once on $[0, T - \tau]$.

From Eq. (40), we see that the monotonicity of J on $[0, T - \tau]$ depends on the sign of w . For all the foregoing, if J attains its maximum at $t_1 = t^* \in [0, T - \tau]$, we can interpret $w(t)$ as an external parameter that shifts the extremal point of J from t^* to zero.

In the next theorem we assume that $x_0 < 1/\sigma_s$. This condition is always satisfied for $\sigma_s < 1$. □

Theorem 2 Let $0 \leq \sigma_s < \sigma_m = \sigma_f$ with $\sigma_s < 1$, $\kappa = 0$ and w be given by Eq. (42). Then the optimal control is unique and is given by

$$\sigma^*(s) = \begin{cases} \sigma_f & \text{for } 0 \leq s < t^*, \\ \sigma_s & \text{for } t^* \leq s < t^* + \eta, \\ \sigma_f & \text{for } t^* + \eta \leq s < T, \end{cases} \tag{48}$$

where

1. For $w(0) \leq 0$: $t^* = 0$ and $\eta = \tau$.
2. For $w(0) > 0$ and $w(T - \tau) \leq 0$: $t^* = \bar{t}$ and $\eta = \tau$ where \bar{t} is the unique value on $[0, T - \tau]$ such that $w(\bar{t}) = 0$.
3. For $0 < w(T - \tau) \leq \frac{1}{\gamma y_{T-\tau, \tau}(T - \tau)}$: $t^* = T - \tau$ and $\eta = \tau$.
4. For $w(T - \tau) > \frac{1}{\gamma y_{T-\tau, \tau}(T - \tau)}$: $t^* = \tilde{t}$ where \tilde{t} is the unique value on $[T - \tau, T]$ such that $w(\tilde{t}) = \frac{1}{\gamma y_{\tilde{t}, T-\tilde{t}}(\tilde{t})}$ and $\eta = T - \tilde{t}$.

Proof From Eq. (43) and Remark 2 the maximum value of J on R must be attained at the superior border P defined in Eq. (34). Therefore, from Eqs. (40) and (41) we have

$$\tilde{J}'(t_1) = \gamma^2(\sigma_f - \sigma_s) \frac{x_{\infty, t_1, \tau}}{(1 - \sigma_f x_{\infty, t_1, \tau})} y_{t_1, \tau}(t_1) y_{t_1, \tau}(T) w(t_1) \quad \text{for } t_1 \in [0, T - \tau], \tag{49}$$

and using that for $\kappa = 0$, $\alpha(t_1) = \frac{1}{\gamma y_{t_1, T-t_1}(t_1)}$, then

$$\tilde{J}'(t_1) = \gamma^2(\sigma_f - \sigma_s) \frac{x_{\infty, t_1, T-t_1}}{1 - \sigma_f x_{\infty, t_1, T-t_1}} y_{t_1, T-t_1}(t_1) y_{t_1, T-t_1}(T) \left(w(t_1) - \frac{1}{\gamma y_{t_1, T-t_1}(t_1)} \right) \tag{50}$$

for $t_1 \in (T - \tau, T]$. Note that from Eq. (75) with $i = 1$ and $j = 0$, for $t \in [T - \tau, T]$ it holds the identity

$$\gamma y_{t_1, T-t_1}(T) \left(w(t_1) - \frac{1}{\gamma y_{t_1, T-t_1}(t_1)} \right) = \gamma(\sigma_f - \sigma_s) h(t_1) - 1 \tag{51}$$

where

$$h(t) = y_{t, T-t}(T) \int_t^T \frac{x_{t, T-t}(s)}{y_{t, T-t}(s)} ds \tag{52}$$

is a decreasing function in $[T - \tau, T]$. In fact, using that $u_1(s)$ is a positive function [see Eqs. (68)) and (62a)] it is easy to see that

$$\frac{d}{dt_1} \left(\frac{y_{t_1}(T)}{y_{t_1}(s)} \right) < 0$$

and therefore, $h'(t_1) < 0$ for $t_1 \in (T - \tau, T)$. From Eqs. (50) and (51) we also have that for $t_1 \in [T - \tau, T]$

$$\tilde{J}'(t_1) = \gamma(\sigma_f - \sigma_s) \frac{x_{\infty, t_1, T-t_1}}{1 - \sigma_f x_{\infty, t_1, T-t_1}} y_{t_1, T-t_1}(t_1) (\gamma(\sigma_f - \sigma_s) h(t_1) - 1) \tag{53}$$

□

We consider the following cases:

1. If $w(0) \leq 0$, then from Lemma 5, $w(t_1) < 0$ for all $t_1 \in (0, T - \tau]$. Thus, from Eq. (49) we have

$$\tilde{J}'(t_1) \leq 0 \text{ for all } t_1 \in [0, T - \tau].$$

Moreover, using that $w(T - \tau) < 0$, the positivity of $\gamma y_{T-\tau, \tau}(T - \tau)$ and Eq. (51), we obtain that $h(T - \tau) < \frac{1}{\gamma(\sigma_f - \sigma_s)}$ and being h a decreasing function, from Eq. (53) we deduce that

$$\tilde{J}'(t_1) < 0 \text{ for all } t_1 \in (T - \tau, T].$$

Therefore $t^* = 0$.

2. Since $w(0) > 0$ and $w(T - \tau) \leq 0$, from Lemma 5 there exists a unique $\bar{t} \in (0, T - \tau]$ such that $w(\bar{t}) = 0$, $w(t_1) > 0$ for all $t_1 \in [0, \bar{t})$ and $w(t_1) < 0$ for all $t_1 \in (\bar{t}, T]$.

Moreover, since $w(T - \tau) \leq 0$, in the same way as for the previous item, we have

$$\tilde{J}'(t_1) < 0 \text{ for all } t_1 \in (T - \tau, T]$$

and from Eq. (49), we obtain

$$\tilde{J}'(t_1) \geq 0 \text{ for } t_1 < \bar{t} \text{ and } \tilde{J}'(t_1) < 0 \text{ for } t_1 > \bar{t},$$

concluding that $t^* = \bar{t}$.

3. Since

$$0 < w(T - \tau) < \frac{1}{\gamma y_{T-\tau, \tau}(T - \tau)},$$

from Lemma 5, $w(t_1) > 0$ for all $t_1 \in [0, T - \tau]$. On the other side, since $w(T - \tau) < \frac{1}{\gamma y_{T-\tau, \tau}(T - \tau)}$, from Eq. (52) we have that $h(T - \tau) \leq \frac{1}{\gamma(\sigma_f - \sigma_s)}$ and using that h is a continuous and decreasing function, we obtain that $h(t_1) < \frac{1}{\gamma(\sigma_f - \sigma_s)}$ for all $t_1 \in (T - \tau, T]$. Thus,

$$\tilde{J}'(t_1) > 0 \text{ for } t_1 \in [0, T - \tau] \text{ and } \tilde{J}'(t_1) < 0 \text{ for } t_1 \in (T - \tau, T].$$

Therefore, $t^* = T - \tau$ and $\eta = \tau$.

4. Since

$$w(T - \tau) > \frac{1}{\gamma y_{T-\tau, \tau}(T - \tau)} > 0,$$

from Lemma 5, we have $w(t_1) > 0$ for all $t_1 \in [0, T - \tau]$. On the other side, since $w(T - \tau) > \frac{1}{\gamma y_{T-\tau, \tau}(T - \tau)}$, then $h(T - \tau) > \frac{1}{\gamma(\sigma_f - \sigma_s)}$ and using that $h(T) = 0$ and h is a continuous and decreasing function, there exists a unique \tilde{t} such that $h(\tilde{t}) = \frac{1}{\gamma(\sigma_f - \sigma_s)}$, $h(t) > \frac{1}{\gamma(\sigma_f - \sigma_s)}$ for $t \in [T - \tau, \tilde{t})$ and $h(t) < \frac{1}{\gamma(\sigma_f - \sigma_s)}$ for $t \in (\tilde{t}, T]$. Therefore,

$$\tilde{J}'(t_1) > 0 \text{ for } t_1 \in [0, \tilde{t}) \text{ and } \tilde{J}'(t_1) < 0 \text{ for } t_1 \in (\tilde{t}, T].$$

Consequently, $t^* = \tilde{t}$ and $\eta = T - \tilde{t}$.

Case $\sigma_s = 0$, $\sigma_m = \sigma_f$ and $\kappa = 0$. Let $w(t)$ defined as in Eq. (35). Note that for $\sigma_s = 0$ and $\sigma_m = \sigma_f$, from Eq. (75) with $i = j = 2$, we obtain

$$w(t) = \begin{cases} \frac{y_{t, \tau}(t + \tau)(\sigma_f x_{t, \tau}(t) - 1)}{\gamma y_{t, \tau}(T) y_{t, \tau}(t)} (e^{\gamma \tau} - 1) & \text{for } 0 \leq t \leq T - \tau \\ \frac{\sigma_f x_{t, T-t}(t) - 1}{\gamma y_{t, T-t}(t)} (e^{\gamma(T-t)} - 1) & \text{for } T - \tau < t \leq T. \end{cases} \tag{54}$$

It is easy to observe that in this case the sign of $w(t)$ on $[0, T - \tau]$ is given by $(\sigma_f x_{t,\tau}(t) - 1)$. Moreover, w changes sign at most once on $[0, T - \tau]$, going from positive to negative values.

Corollary 1 *Let $\sigma_s = 0, \sigma_m = \sigma_f > 0$ and $k = 0$. Then, the optimal control is unique and is given by*

$$\sigma^*(s) = \begin{cases} \sigma_f & \text{for } 0 \leq s < t^*, \\ 0 & \text{for } t^* \leq s < t^* + \eta, \\ \sigma_f & \text{for } t^* + \eta \leq s \leq T, \end{cases}$$

where

1. For $x_0 \leq \frac{1}{\sigma_f}$: $t^* = 0$ and $\eta = \tau$.
2. For $x_0 > \frac{1}{\sigma_f}$ and $x_{T-\tau,\tau}(T - \tau) \leq \frac{1}{\sigma_f}$: $t^* = \bar{t}$ and $\eta = \tau$, where \bar{t} is the unique value on $[0, T - \tau]$ such that $x_{\bar{t},\tau}(\bar{t}) = \frac{1}{\sigma_f}$.
3. For $\frac{1}{\sigma_f} < x_{T-\tau,\tau}(T - \tau) \leq \frac{1}{\sigma_f(1 - e^{-\gamma\tau})}$: $t^* = T - \tau$ and $\eta = \tau$.
4. For $x_{T-\tau,\tau}(T - \tau) > \frac{1}{\sigma_f(1 - e^{-\gamma\tau})}$: $t^* = \tilde{t}$ and $\eta = T - \tilde{t}$, where \tilde{t} is the unique value on $[T - \tau, T]$ such that $x_{\tilde{t},T-\tilde{t}}(\tilde{t}) = \frac{1}{\sigma_f(1 - e^{-\gamma(T-\tilde{t})})}$.

Proof The proof follows from Theorem 2 using the fact that

$$\text{sign}(w(t)) = \text{sign}(\sigma_f x_{t,\tau}(t) - 1), \quad \text{for } t \in [0, T - \tau],$$

and

$$w(T - \tau) = \frac{1}{\gamma y_{T-\tau,\tau}(T - \tau)} (\sigma_f x_{T-\tau,\tau}(T - \tau) - 1)(e^{\gamma\tau} - 1).$$

Note that in this case if we take $\tau = T$, the corollary is reduced to only two possible cases: $x_0 \leq \frac{1}{\sigma_f(1 - e^{-\gamma T})}$ or $x_0 > \frac{1}{\sigma_f(1 - e^{-\gamma T})}$, obtaining the same result as Ketcheson²³ in Theorem 3. □

General case. In this section we study the behaviour of optimal solutions for the general case when $0 \leq \sigma_s < \sigma_m \leq \sigma_f$ and $\kappa > 0$, that is, objective function J includes the term that accounts the running cost of the control and allows us to account for factors like the economic cost of intervention or heightened risks caused by hospital overflow.

Lemma 6 *Assume $\kappa > \frac{x_{\infty,t_1,\eta} \gamma y_{t_1,\eta}(T)}{1 - \sigma_f x_{\infty,t_1,\eta}} \left(1 - \gamma y_2 \int_{t_2}^T \frac{\sigma_f x_{t,\eta}(r) - 1}{y_{t,\eta}(r)} dr\right)$ for all $(t_1, \eta) \in R$, then the optimal control is given by $\sigma^* \equiv \sigma_m$.*

Proof From Eq. (32), we have that $\frac{\partial J}{\partial \eta}(t_1, \eta) < 0$ and therefore the maximum value of J on R is attained at the inferior border of R where $\eta = 0$ and $J(t_1, 0)$ is constant.

In the next theorem we give a general result including both the economic cost of intervention ($\kappa > 0$) and a mitigation phase different from the no intervention one, that is $\sigma_m < \sigma_f$. In the next section we give numerical simulations supporting this result. When $\sigma = \sigma_f$ and $\kappa = 0$ we recover Theorem 2. □

Theorem 3 *Let $0 \leq \sigma_s < \sigma_m \leq \sigma_f$ with $\sigma_s < 1, \kappa$ satisfying Eq. (33) for all $(t_1, \eta) \in R$ and let w and α defined as in Eqs. (35) and (36) respectively, then the optimal control is unique and is given by*

$$\sigma^*(s) = \begin{cases} \sigma_m & \text{for } 0 \leq s < t^*, \\ \sigma_s & \text{for } t^* \leq s < t^* + \eta, \\ \sigma_m & \text{for } t^* + \eta \leq s < T, \end{cases} \tag{55}$$

where

1. For $w(0) \leq 0$: $t^* = 0$ and $\eta = \tau$.
2. For $w(0) > 0$ and $w(T - \tau) \leq 0$: $t^* = \bar{t}$ and $\eta = \tau$ where \bar{t} is the unique value on $[0, T - \tau]$ such that $w(\bar{t}) = 0$.
3. For $0 < w(T - \tau) \leq \alpha(T - \tau)$: $t^* = T - \tau$ and $\eta = \tau$.

4. For $w(T - \tau) > \alpha(T - \tau)$: $t^* = \tilde{t}$ where \tilde{t} is the unique value on $[T - \tau, T]$ such that $w(\tilde{t}) = \alpha(\tilde{t})$ and $\eta = T - \tilde{t}$.

Data availability

All relevant data are within the paper.

Appendix

We begin by computing the derivatives of $x_{t_1, \eta}(T) = \Psi_1(T, t_1 + \eta, x_2, y_2, \sigma_m)$ and $y_{t_1, \eta}(T) = \Psi_2(T, t_1 + \eta, x_2, y_2, \sigma_m)$ with respect to t_1 .

We recall two properties for the solutions of ordinary differential equations. First, the relation between the derivative with respect to initial time and the derivatives with respect to initial data give us the equation

$$\frac{\partial \Psi_j(s, t, x, y, \sigma)}{\partial t} = \frac{\partial \Psi_j(s, t, x, y, \sigma)}{\partial x} \gamma \sigma x y - \frac{\partial \Psi_j(s, t, x, y, \sigma)}{\partial y} \gamma y (\sigma x - 1) \tag{56}$$

for Ψ defined in Eq. (28), with $s \geq t, \sigma \in \{\sigma_s, \sigma_m\}$, initial data $(x, y) \in \mathcal{D}$ at time t and $j = 1, 2$.

Second, the dependence of the solution $\Psi(s, t, x, y, \sigma)$ with respect to initial data x, y is given by the following known equations. For simplicity of notation, when there is no risk of confusion, we will denote $\Psi(s)$ for $\Psi(s, t, x, y, \sigma)$,

$$\begin{pmatrix} \frac{\partial \Psi_1}{\partial x} & \frac{\partial \Psi_1}{\partial y} \\ \frac{\partial \Psi_2}{\partial x} & \frac{\partial \Psi_2}{\partial y} \end{pmatrix}'(s) = \begin{pmatrix} -\gamma \sigma \Psi_2(s) & -\gamma \sigma \Psi_1(s) \\ \gamma \sigma \Psi_2(s) & \gamma (\sigma \Psi_1(s) - 1) \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial \Psi_1}{\partial x} & \frac{\partial \Psi_1}{\partial y} \\ \frac{\partial \Psi_2}{\partial x} & \frac{\partial \Psi_2}{\partial y} \end{pmatrix}(s). \tag{57}$$

with initial data

$$\begin{pmatrix} \frac{\partial \Psi_1}{\partial x} & \frac{\partial \Psi_1}{\partial y} \\ \frac{\partial \Psi_2}{\partial x} & \frac{\partial \Psi_2}{\partial y} \end{pmatrix}(t) = Id. \tag{58}$$

Then we call $\sigma_1 = \sigma_s, \sigma_2 = \sigma_m$ and for $i = 1, 2$,

$$u_i(s) = u(s, t_i, x_i, y_i, \sigma_i) = \frac{\partial \Psi_1}{\partial x_i}(s, t_i, x_i, y_i, \sigma_i) - \frac{\partial \Psi_1}{\partial y_i}(s, t_i, x_i, y_i, \sigma_i), \tag{59a}$$

$$v_i(s) = v(s, t_i, x_i, y_i, \sigma_i) = \frac{\partial \Psi_2}{\partial x_i}(s, t_i, x_i, y_i, \sigma_i) - \frac{\partial \Psi_2}{\partial y_i}(s, t_i, x_i, y_i, \sigma_i), \tag{59b}$$

for $s \in [t_i, t_{i+1}]$, and we have the system of equations on u_i and v_i

$$\begin{pmatrix} u_i'(s) \\ v_i'(s) \end{pmatrix} = \begin{pmatrix} -\gamma \sigma_i \Psi_2(s) & -\gamma \sigma_i \Psi_1(s) \\ \gamma \sigma_i \Psi_2(s) & \gamma (\sigma_i \Psi_1(s) - 1) \end{pmatrix} \cdot \begin{pmatrix} u_i(s) \\ v_i(s) \end{pmatrix}. \tag{60}$$

with initial data

$$\begin{pmatrix} u_i(t_i) \\ v_i(t_i) \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{61}$$

Therefore, after some computations and using Eqs. (56) and (59) we obtain for $s \in (t_1, t_1 + \eta)$

$$\frac{dx_{t_1, \eta}}{dt_1}(s) = -\gamma (\sigma_m - \sigma_s) x_1 y_1 u_1(s), \tag{62a}$$

$$\frac{dy_{t_1, \eta}}{dt_1}(s) = -\gamma (\sigma_m - \sigma_s) x_1 y_1 v_1(s), \tag{62b}$$

and for $s \in (t_1 + \eta, T]$

$$\begin{aligned} \frac{dx_{t_1, \eta}}{dt_1}(s) &= \gamma (\sigma_m - \sigma_s) x_2 y_2 u_2(s) \\ &\quad - \gamma (\sigma_m - \sigma_s) x_1 y_1 \left(\frac{\partial \Psi_1(s, t_2, x_2, y_2, \sigma_m)}{\partial x_2} u_1(t_2) + \frac{\partial \Psi_1(s, t_2, x_2, y_2, \sigma_m)}{\partial y_2} v_1(t_2) \right), \end{aligned} \tag{63a}$$

$$\begin{aligned} \frac{dy_{t_1, \eta}}{dt_1}(s) &= \gamma (\sigma_m - \sigma_s) x_2 y_2 v_2(s) \\ &\quad - \gamma (\sigma_m - \sigma_s) x_1 y_1 \left(\frac{\partial \Psi_2(s, t_2, x_2, y_2, \sigma_m)}{\partial x_2} u_1(t_2) + \frac{\partial \Psi_2(s, t_2, x_2, y_2, \sigma_m)}{\partial y_2} v_1(t_2) \right). \end{aligned} \tag{63b}$$

Moreover, using that for any $(x_i, y_i) \in \mathcal{D}$, $\Psi(s, t_i, x_i, y_i, \sigma_i)$, satisfies for $s \in [t_i, t_{i+1}]$

$$\Psi_1(s, t_i, x_i, y_i, \sigma_i)e^{-\sigma_i(\Psi_1(s, t_i, x_i, y_i, \sigma_i) + \Psi_2(s, t_i, x_i, y_i, \sigma_i))} = x_i e^{-\sigma_i(x_i + y_i)}, \tag{64}$$

we compute the derivatives with respect to x_i and y_i and using $\Psi_1(s) = \Psi_1(s, t_i, x_i, y_i, \sigma_i)$, we obtain for $s \in [t_i, t_{i+1}]$

$$\frac{\partial \Psi_1}{\partial x_i}(s) - \sigma_i \Psi_1(s) \left(\frac{\partial \Psi_1}{\partial x_i}(s) + \frac{\partial \Psi_2}{\partial x_i}(s) \right) = (1 - \sigma_i x_i) \frac{\Psi_1(s)}{x_i}, \tag{65}$$

$$\frac{\partial \Psi_1}{\partial y_i}(s) - \sigma_i \Psi_1(s) \left(\frac{\partial \Psi_1}{\partial y_i}(s) + \frac{\partial \Psi_2}{\partial y_i}(s) \right) = -\sigma_i \Psi_1(s). \tag{66}$$

Then, subtracting the last two equations

$$u_i(s) - \sigma_i \Psi_1(s, t_i, x_i, y_i, \sigma_i)(u_i(s) + v_i(s)) = \frac{\Psi_1(s, t_i, x_i, y_i, \sigma_i)}{x_i} \tag{67}$$

for $s \in [t_i, t_{i+1}]$ and therefore using Eq. (67), from Eq. (60) we have that u_i satisfies the ordinary differential equation

$$u_i'(s) = \gamma (\sigma_i \Psi_1(s, t_i, x_i, y_i, \sigma_i) - \sigma_i \Psi_2(s, t_i, x_i, y_i, \sigma_i) - 1) u_i(s) + \gamma \frac{\Psi_1(s, t_i, x_i, y_i, \sigma_i)}{x_i},$$

$$u_i(t_i) = 1.$$

In the rest of this section, for simplicity of notation we will denote x and y for $x_{t_1, \eta}$ and $y_{t_1, \eta}$, defined in Eq. (29), respectively.

Thus, for $s \in [t_i, t_{i+1}]$ when $i = 1, 2$, we obtain

$$u_i(s) = \frac{x(s)y(s)}{x_i y_i} + \gamma \frac{x(s)y(s)}{x_i} \int_{t_i}^s \frac{1}{y(r)} dr$$

$$= \frac{x(s)}{x_i} + \gamma \sigma_i \frac{x(s)y(s)}{x_i} \int_{t_i}^s \frac{x(r)}{y(r)} dr, \tag{68}$$

$$v_i(s) = -\frac{x(s)}{x_i} + \frac{(1 - \sigma_i x(s))y(s)}{x_i} \gamma \int_{t_i}^s \frac{x(r)}{y(r)} dr, \tag{69}$$

$$u_i(s) + v_i(s) = \frac{y(s)}{x_i} \gamma \int_{t_i}^s \frac{x(r)}{y(r)} dr. \tag{70}$$

Also, from Eqs. (60) and (66) we can prove for $\Psi(T) = \Psi(T, t_2, x_2, y_2, \sigma_m)$ that

$$\frac{\partial \Psi_1}{\partial x_2}(T) + \frac{\partial \Psi_2}{\partial x_2}(T) = \frac{y(T)}{y_2} + (1 - \sigma_m x_2)(u_2(T) + v_2(T)), \tag{71a}$$

$$\frac{\partial \Psi_1}{\partial y_2}(T) + \frac{\partial \Psi_2}{\partial y_2}(T) = \frac{y(T)}{y_2} - \sigma_m x_2 x_2 (u_2(T) + v_2(T)). \tag{71b}$$

Analogously,

$$\frac{dx_{t_1, \eta}}{d\eta}(T) = \gamma(\sigma_m - \sigma_s)x_2 y_2 u_2(T), \tag{72a}$$

$$\frac{dy_{t_1, \eta}}{d\eta}(T) = \gamma(\sigma_m - \sigma_s)x_2 y_2 v_2(T). \tag{72b}$$

We can now compute the derivatives of $J(t_1, \eta)$ given by Eq. (30). From Eq. (15),

$$\frac{\partial J}{\partial t_1}(t_1, \eta) = \frac{dx_{\infty}(x_{t_1, \eta}(T), y_{t_1, \eta}(T), \sigma_f)}{dt_1}$$

$$= \frac{x_{\infty, t_1, \eta}}{1 - \sigma_f x_{\infty, t_1, \eta}} \left(\frac{1 - \sigma_f x_{t_1, \eta}(T)}{x_{t_1, \eta}(T)} \frac{dx_{t_1, \eta}(T)}{dt_1} - \sigma_f \frac{dy_{t_1, \eta}(T)}{dt_1} \right) \tag{73}$$

$$= \frac{x_{\infty, t_1, \eta}}{(1 - \sigma_f x_{\infty, t_1, \eta})x_{t_1, \eta}(T)} \left((1 - \sigma_f x_{t_1, \eta}(T)) \frac{dx_{t_1, \eta}(T)}{dt_1} - \sigma_f x_{t_1, \eta}(T) \frac{dy_{t_1, \eta}(T)}{dt_1} \right),$$

and, from Eq. (63)

$$\begin{aligned}
 & (1 - \sigma_f x_{t_1, \eta}(T)) \frac{dx_{t_1, \eta}(T)}{dt_1} - \sigma_f x_{t_1, \eta}(T) \frac{dy_{t_1, \eta}(T)}{dt_1} \\
 &= \gamma(\sigma_m - \sigma_s) x_2 y_2 \left((1 - \sigma_f x_{t_1, \eta}(T)) u_2(T) - \sigma_f x_{t_1, \eta}(T) v_2(T) \right) \\
 & - \gamma(\sigma_m - \sigma_s) x_1 y_1 \left(u_1(t_2) \left((1 - \sigma_f x_{t_1, \eta}(T)) \frac{\partial \Psi_1(T, t_2, x_2, y_2, \sigma_m)}{\partial x_2} + \sigma_f x_{t_1, \eta}(T) \frac{\partial \Psi_2(T, t_2, x_2, y_2, \sigma_m)}{\partial x_2} \right) \right. \\
 & \left. + v_1(t_2) \left((1 - \sigma_f x_{t_1, \eta}(T)) \frac{\partial \Psi_1(T, t_2, x_2, y_2, \sigma_m)}{\partial y_2} + \sigma_f x_{t_1, \eta}(T) \frac{\partial \Psi_2(T, t_2, x_2, y_2, \sigma_m)}{\partial y_2} \right) \right),
 \end{aligned}$$

using Eqs. (65), (66), (67) and (71) we obtain

$$\begin{aligned}
 & (1 - \sigma_f x_{t_1, \eta}(T)) \frac{dx_{t_1, \eta}(T)}{dt_1} - \sigma_f x_{t_1, \eta}(T) \frac{dy_{t_1, \eta}(T)}{dt_1} \\
 &= \gamma(\sigma_m - \sigma_s) x_{t_1, \eta}(T) \left[(1 + (\sigma_m - \sigma_f)(u_2(T) + v_2(T)))(y_2 - y_1) + (\sigma_m - \sigma_s) x_1 y_1 (u_1(t_2) + v_1(t_2)) \right. \\
 & \left. - (\sigma_m - \sigma_f) x_1 y_1 \frac{y_{t_1, \eta}(T)}{y_2} (u_1(t_2) + v_1(t_2)) \right].
 \end{aligned}$$

Also, using that for $i = 1, 2$

$$\frac{1}{y_i} - \frac{1}{y_{i+1}} = - \int_{t_i}^{t_{i+1}} \left(\frac{1}{y} \right)'(r) dr = \gamma \int_{t_i}^{t_{i+1}} \frac{\sigma_i x(r) - 1}{y(r)} dr, \tag{74}$$

we have that

$$y_{i+1} - y_i + \gamma y_i y_{i+1} (\sigma_j - \sigma_i) \int_{t_i}^{t_{i+1}} \frac{x(r)}{y(r)} dr = \gamma y_i y_{i+1} \int_{t_i}^{t_{i+1}} \frac{\sigma_j x(r) - 1}{y(r)} dr. \tag{75}$$

Therefore, from Eqs. (70) and (75)

$$\begin{aligned}
 & \frac{(1 - \sigma_f x_{t_1, \eta}(T)) \frac{dx_{t_1, \eta}(T)}{dt_1} - \sigma_f x_{t_1, \eta}(T) \frac{dy_{t_1, \eta}(T)}{dt_1}}{\gamma^2 (\sigma_m - \sigma_s) x_{t_1, \eta}(T) y_{t_1, \eta}(T) y_1} \\
 &= \left(1 - \gamma y_2 \int_{t_2}^T \frac{\sigma_f x(r) - 1}{y(r)} dr \right) \int_{t_1}^{t_2} \frac{\sigma_m x(r) - 1}{y(r)} dr - (\sigma_m - \sigma_f) \int_{t_1}^{t_2} \frac{x(r)}{y(r)} dr \\
 &= \int_{t_1}^{t_2} \frac{\sigma_f x(r) - 1}{y(r)} dr - \gamma y_2 \int_{t_2}^T \frac{\sigma_f x(r) - 1}{y(r)} dr \int_{t_1}^{t_2} \frac{\sigma_m x(r) - 1}{y(r)} dr.
 \end{aligned} \tag{76}$$

Then, replacing in Eq. (73) we obtain

$$\begin{aligned}
 \frac{\partial J}{\partial t_1}(t_1, \eta) &= \frac{\gamma^2 (\sigma_m - \sigma_s) y_{t_1, \eta}(T) y_1 x_{\infty, t_1, \eta}}{1 - \sigma_f x_{\infty, t_1, \eta}} \\
 & \left[\int_{t_1}^{t_2} \frac{\sigma_f x(r) - 1}{y(r)} dr - \gamma y_2 \int_{t_2}^T \frac{\sigma_f x(r) - 1}{y(r)} dr \int_{t_1}^{t_2} \frac{\sigma_m x(r) - 1}{y(r)} dr \right].
 \end{aligned} \tag{77}$$

Note that for $\sigma_m = \sigma_f$ we have from Eq. (74) for $i = 2$ that

$$(1 - \sigma_f x_{t_1, \eta}(T)) \frac{dx_{t_1, \eta}(T)}{dt_1} - \sigma_f x_{t_1, \eta}(T) \frac{dy_{t_1, \eta}(T)}{dt_1} = \gamma^2 (\sigma_m - \sigma_s) x_{t_1, \eta}(T) y_1 y_2 \int_{t_1}^{t_2} \frac{\sigma_m x(r) - 1}{y(r)} dr,$$

and then,

$$\frac{\partial J}{\partial t_1}(t_1, \eta) = \frac{x_{\infty, t_1, \eta} \gamma^2 (\sigma_m - \sigma_s) y_1 y_2}{1 - \sigma_f x_{\infty, t_1, \eta}} \int_{t_1}^{t_2} \frac{\sigma_m x(r) - 1}{y(r)} dr. \tag{78}$$

On the other hand,

$$\begin{aligned}
 \frac{\partial J}{\partial \eta}(t_1, \eta) &= \frac{dx_{\infty}(x_{t_1, \eta}(T), y_{t_1, \eta}(T), \sigma_f)}{d\eta} - \kappa(\sigma_m - \sigma_s) \\
 &= \frac{x_{\infty, t_1, \eta}}{(1 - \sigma_f x_{\infty, t_1, \eta}) x_{t_1, \eta}(T)} \left((1 - \sigma_f x_{t_1, \eta}(T)) \frac{dx_{t_1, \eta}(T)}{d\eta} - x_{t_1, \eta}(T) \sigma_f \frac{dy_{t_1, \eta}(T)}{d\eta} \right) - \kappa(\sigma_m - \sigma_s),
 \end{aligned} \tag{79}$$

where $x_{\infty, t, \eta} = x_{\infty}(x_{t, \eta}(T), y_{t, \eta}(T), \sigma_f)$.

From Eqs. (59a), (59b) and (72), we obtain

$$\begin{aligned}
& \frac{1}{x_{t_1, \eta}(T)} \left((1 - \sigma_f x_{t_1, \eta}(T)) \frac{dx_{t_1, \eta}(T)}{d\eta} - \sigma_f x_{t_1, \eta}(T) \frac{dy_{t_1, \eta}(T)}{d\eta} \right) \\
&= \gamma y_2 (\sigma_m - \sigma_s) \frac{x_2}{x_{t_1, \eta}(T)} \left((1 - \sigma_f x_{t_1, \eta}(T)) u_2(T) - \sigma_f x_{t_1, \eta}(T) v_2(T) \right) \\
&= \gamma y_2 (\sigma_m - \sigma_s) \frac{x_2}{x_{t_1, \eta}(T)} \left((1 - \sigma_m x_{t_1, \eta}(T)) u_2(T) - \sigma_m x_{t_1, \eta}(T) v_2(T) - (\sigma_f - \sigma_m) x_{t_1, \eta}(T) (u_2(T) + v_2(T)) \right),
\end{aligned} \tag{80}$$

and using Eqs. (67), (70) and (75)

$$\begin{aligned}
&= \gamma y_2 (\sigma_m - \sigma_s) \frac{x_2}{x_{t_1, \eta}(T)} \left(\frac{x_{t_1, \eta}(T)}{x_2} - (\sigma_f - \sigma_m) x_{t_1, \eta}(T) (u_2(T) + v_2(T)) \right) \\
&= \gamma y_2 (\sigma_m - \sigma_s) \frac{x_2}{x_{t_1, \eta}(T)} \left(\frac{x_{t_1, \eta}(T)}{x_2} - (\sigma_f - \sigma_m) x_{t_1, \eta}(T) \frac{y_{t_1, \eta}(T)}{x_2} \gamma \int_{t_2}^T \frac{x(r)}{y(r)} dr \right) \\
&= \gamma y_2 (\sigma_m - \sigma_s) \left(1 - (\sigma_f - \sigma_m) y_{t_1, \eta}(T) \gamma \int_{t_2}^T \frac{x(r)}{y(r)} dr \right) \\
&= \gamma (\sigma_m - \sigma_s) \left(y_{t_1, \eta}(T) - \gamma y_2 y_{t_1, \eta}(T) \int_{t_2}^T \frac{\sigma_f x(r) - 1}{y(r)} dr \right).
\end{aligned} \tag{81}$$

Thus, we have

$$\frac{\partial J}{\partial \eta}(t_1, \eta) = \frac{x_{\infty, t_1, \eta}}{1 - \sigma_f x_{\infty, t_1, \eta}} \gamma (\sigma_m - \sigma_s) y_{t_1, \eta}(T) \left(1 - \gamma y_2 \int_{t_2}^T \frac{\sigma_f x(r) - 1}{y(r)} dr \right) - \kappa (\sigma_m - \sigma_s). \tag{82}$$

Received: 29 October 2021; Accepted: 12 July 2022

Published online: 22 July 2022

References

- Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Charact.* **115**(772), 700–721 (1927).
- Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, 1992).
- Brauer, F., Castillo-Chavez, C. & Castillo-Chavez, C. *Mathematical Models in Population Biology and Epidemiology* Vol. 2 (Springer, 2012).
- Funk, S., Gilad, E., Watkins, C. & Jansen, V. A. The spread of awareness and its impact on epidemic outbreaks. *Proc. Natl. Acad. Sci.* **106**(16), 6872–6877 (2009).
- Buono, C., Vazquez, F., Macri, P. A. & Braunstein, L. A. Slow epidemic extinction in populations with heterogeneous infection rates. *Phys. Rev. E* **88**(2), 022813 (2013).
- Giambiagi Ferrari, C., Pinasco, J. P. & Saintier, N. Coupling epidemiological models with social dynamics. *Bull. Math. Biol.* **83**(7), 1–23 (2021).
- Ferreira, E. J. Ecuaciones lmite para dinámicas markovianas: vacunación óptima en grafos aleatorios e intercambio de opiniones. Ph.D. Thesis, Universidad de Buenos Aires, (2021).
- Janson, S., Luczak, M. & Windridge, P. Law of large numbers for the SIR epidemic on a random graph with given degrees. *Random Struct. Algorithms* **45**(4), 726–763 (2014).
- Lagorio, C. *et al.* Quarantine-generated phase transition in epidemic spreading. *Phys. Rev. E* **83**(2), 026102 (2011).
- Velásquez-Rojas, F. & Vazquez, F. Interacting opinion and disease dynamics in multiplex networks: Discontinuous phase transition and nonmonotonic consensus times. *Phys. Rev. E* **95**(5), 052315 (2017).
- Volz, E. SIR dynamics in random networks with heterogeneous connectivity. *J. Math. Biol.* **56**(3), 293–310 (2008).
- Ferreira, E. J., Jonckheere, M. & Pinasco, J. P. SIR dynamics with vaccination in a large configuration model. *Appl. Math. Optim.* **84**(2), 1769–1818 (2021).
- Biswas, M. H. A., Paiva, L. T. & de Pinho, M. D. R. A SEIR model for control of infectious diseases with constraints. *Math. Biosci. Eng.* **11**(4), 761–784 (2014).
- Zhou, Y., Yang, K., Zhou, K. & Liang, Y. Optimal vaccination policies for an SIR model with limited resources. *Acta Biotheoretica* **62**(2), 171–181 (2014).
- Cori, A. *et al.* Key data for outbreak evaluation: Building on the Ebola experience. *Philos. Trans. R. Soc. B* **372**, 20160371 (2017).
- Ledzewicz, U. & Schättler, H. On optimal singular controls for a general SIR-model with vaccination and treatment. *Discrete Contin. Dyn. Syst.* **981–990**, 2011 (2011).
- Abouelkheir, I., El Kihal, F., Rachik, M. & Elmouki, I. Time needed to control an epidemic with restricted resources in SIR model with short-term controlled population: A fixed point method for a free isoperimetric optimal control problem. *Math. Comput. Appl.* **23**(4), 64 (2018).
- Hansen, E. & Day, T. Optimal control of epidemics with limited resources. *J. Math. Biol.* **62**(3), 423–451 (2011).
- El Kihal, F., Abouelkheir, I., Rachik, M. & Elmouki, I. Optimal control and computational method for the resolution of isoperimetric problem in a discrete-time SIRS system. *Math. Comput. Appl.* **23**(4), 52 (2018).
- Zhou, Y., Jianhong, W. & Min, W. Optimal isolation strategies of emerging infectious diseases with limited resources. *Math. Biosci. Eng.* **10**(5–6), 1691–701 (2013).
- Ferguson, N. *et al.* Report 9: Impact of non-pharmaceutical interventions (NPIS) to reduce covid19 mortality and healthcare demand. (Imperial College London, 2020).
- Greenhalgh, D. Some results on optimal control applied to epidemics. *Math. Biosci.* **88**(2), 125–158 (1988).
- Ketcheson, D. I. Optimal control of an sir epidemic through finite-time non-pharmaceutical intervention. *J. Math. Biol.* **83**(7), 1–21 (2021).
- Clarke, F. *Functional Analysis, Calculus of Variations and Optimal Control* Vol. 264 (Springer, 2013).
- Hartl, R. F., Sethi, S. P. & Vickson, R. G. A survey of the maximum principles for optimal control problems with state constraints. *SIAM Rev.* **37**(2), 181–218 (1995).

26. Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42**(4), 599–653 (2000).
27. Lenhart, S. & Workman, J. T. *Optimal Control Applied to Biological Models* (Chapman & Hall/CRC, 2007).
28. Köhler, J. *et al.* Robust and optimal predictive control of the COVID-19 outbreak. *Ann. Rev. Control* **51**, 525–539 (2020).
29. Morris, D. H., Rossine, F. W., Plotkin, J. B. & Levin, S. A. Optimal, near-optimal, and robust epidemic control. arXiv preprint [arXiv:2004.02209](https://arxiv.org/abs/2004.02209), (2020).
30. Kruse, T. & Strack, P. Optimal control of an epidemic through social distancing. Cowles Foundation Discussion Papers 2229, Cowles Foundation for Research in Economics, Yale University, (2020).
31. Palmer, A. Z., Zabinsky, Z. B. & Liu, S. Optimal control of COVID-19 infection rate with social costs. arXiv preprint [arXiv:2007.13811](https://arxiv.org/abs/2007.13811), (2020).
32. Tsay, C., Lejarza, F., Stadtherr, M. A. & Baldea, M. Modeling, state estimation, and optimal control for the US COVID-19 outbreak. *Sci. Rep.* **10**(1), 1–12 (2020).
33. Godara, P., Herminghaus, S. & Heidemann, K. M. A control theory approach to optimal pandemic mitigation. *PLoS ONE* **16**(2), e0247445 (2021).
34. Pueyo, T. Coronavirus: the hammer and the dance. Medium [Internet], (2020).
35. Zwillinger, D. *Handbook of Differential Equations* (Academic Press, 1989).
36. Abramowitz, M. & Stegun, I. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing. (Dover, 1972).
37. Pakes, A. G. Lambert's W meets Kermack-McKendrick epidemics. *IMA J. Appl. Math.* **80**(5), 1368–1386 (2015).

Author contributions

R.B., F.V. and C.S.d.l.V. wrote the main manuscript text and prepared the figures. All authors reviewed the main manuscript text.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.S.d.l.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022