Optimal intervention strategies for minimizing total incidence during an epidemic

Tom Britton* Lasse Leskelä†

February 16, 2022

Abstract

This article considers the minimization of the total number of infected individuals over the course of an epidemic in which the rate of infectious contacts can be reduced by time-dependent nonpharmaceutical interventions. The societal and economic costs of interventions are taken into account using a linear budget constraint which imposes a trade-off between short-term heavy interventions and long-term light interventions. We search for an optimal intervention strategy in an infinite-dimensional space of controls containing multiple consecutive lockdowns, gradually imposed and lifted restrictions, and various heuristic controls based for example on tracking the effective reproduction number. Mathematical analysis shows that among all such strategies, the global optimum is achieved by a single constant-level lockdown of maximum possible magnitude. Numerical simulations highlight the need of careful timing of such interventions, and illustrate their benefits and disadvantages compared to strategies designed for minimizing peak prevalence. Rather counterintuitively, adding restrictions prior to the start of a well-planned intervention strategy may even increase the total incidence.

1 Introduction

The recent pandemic has underlined the need for non-pharmaceutical interventions to help mitigating disease burden in the society, along with vaccines and medications. Despite a solid body of past literature and an enormous research effort during the past two years on epidemic modelling, certain fundamental questions related to the optimal control of epidemics still remain open. In this article we discuss the optimal employment of non-pharmaceutical interventions to mitigate disease burden under the assumption that interventions incur societal costs which are accumulated over

*Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden
†Department of Mathematics and Systems Analysis, Aalto University, FI-02015 Espoo, Finland

time. We focus on the minimization of long-term total incidence (the share of initially susceptible individuals who eventually become infected), and seek answers to questions of type:

> *Should interventions be imposed early, or later after prevalence has grown? Is it better to impose a one-month lockdown at 50% intervention level, or a milder two-month lockdown at 25% intervention level?*

Numerical simulations and control theory are routinely used to answer such questions on a case-by-case basis for models of unlimited complexity. An alternative approach, pursued here, is to search for universal mathematical principles characterizing the shape and size of optimal interventions in simple parsimonious models.

The simplest mathematical epidemic model, incorporating a time-dependent intervention strategy is arguably defined as follows. We assume that the transmission rate of infectious contacts at time t can be reduced by a factor

$$0 \leq u(t) \leq 1.$$

Under classical simplifying assumptions that recovery from disease gives full immunity, overall vaccination status in the population remains constant, there are no imported cases from other populations, population size remains constant, and the population is homogeneously mixing, the evolution of the epidemic can be modelled using differential equations

$$\begin{aligned} S' &= -(1-u)\beta SI, \\ I' &= (1-u)\beta SI - \gamma I, \\ R' &= \gamma I, \end{aligned} \tag{1}$$

where $\beta > 0$ is a baseline transmission rate in absence of interventions, $\gamma > 0$ is the recovery rate, and $S(t)$, $I(t)$, and $R(t)$ represent the shares of susceptible, infectious, and recovered individuals in the population, respectively. The special case with no interventions ($u(t) = 0$ for all t) reduces to the constant-rate version of the classical SIR model [1].

We write $(S, I, R) = (S_u, I_u, R_u)$ to emphasize that the epidemic trajectory depends on the chosen intervention strategy u. The societal and economic costs incurred by adopting an intervention strategy u can by measured by

- total cost $||u||_1 = \int_0^\infty u(t)\, dt$,
- total duration $||u||_0 = \int_0^\infty 1(u(t) > 0)\, dt$,
- maximum intervention level $||u||_\infty = \sup_{t \geq 0} u(t)$.

For this model, the basic reproduction number equals $\mathcal{R}_0 = \frac{\beta}{\gamma}$. In what follows, we will assume that $\mathcal{R}_0 > 1$ because otherwise epidemic outbreaks would not happen even in the absence of interventions [2].

1.1 Minimizing peak prevalence

Recent theoretical research [3–6] on minimizing disease burden has mostly focused on minimizing the peak prevalence

$$||I_u||_\infty \;=\; \sup_{t\geq 0} I_u(t).$$

Morris et al. [3] proved that the peak prevalence subject to interventions of bounded duration $||u||_0 \leq c_0$ is minimized by an intervention of form

$$u(t) \;=\; \begin{cases} 0, & t\in(0,t_1] \quad \text{(wait)} \\ 1-\frac{1}{(\beta/\gamma)S(t)}, & t\in(t_1,t_2] \quad \text{(maintain)} \\ 1, & t\in(t_2,t_3] \quad \text{(suppress)} \\ 0, & t\in(t_3,\infty) \quad \text{(relax)}, \end{cases} \tag{2}$$

and discovered that such interventions induce a second wave having a peak of same height as the first wave. Such a twin peaks phenomenon was also noted in [4]. Miclo, Spiro, and Weibull [5] studied a dual problem of minimizing the intervention cost $||u||_1$ subject to a bounded peak prevalence $||I_u||_\infty \leq c_\infty$, and proved that the optimum is of form

$$u(t) \;=\; \begin{cases} 0, & t\in(0,t_1] \quad \text{(wait)} \\ 1-\frac{1}{(\beta/\gamma)S(t)}, & t\in(t_1,t_2] \quad \text{(maintain)} \\ 0, & t\in(t_2,\infty) \quad \text{(relax)}. \end{cases} \tag{3}$$

The maintain phase in (2–3) is defined so that $I_u' = 0$, which keeps the infectious share at a constant level. A similar control problem restricted to a finite time horizon is analysed in [6].

1.2 Minimizing total incidence

Recent theoretical works on minimizing total incidence include [7–10]. Under different budget constraints on intervention costs, they all conclude that the optimal interventions are constant-level lockdowns with shape

$$u(t) \;=\; \begin{cases} 0, & t\in(0,t_1] \quad \text{(wait)} \\ c, & t\in(t_1,t_2] \quad \text{(suppress)} \\ 0, & t\in(t_2,\infty) \quad \text{(relax)}. \end{cases} \tag{4}$$

Feng, Iyer, and Li [7] considered on-off controls with finitely many switching times, with duration $||u||_0 = c_0$ and maximum level $||u||_\infty = c_\infty$. In parallel works, Bliman et al. [8] and Ketcheson [9] studied piecewise continuous interventions subject to $||u||_\infty \leq c_\infty$ and a bounded intervention time window $[0, T_0]$. Bliman and Duprez [10] considered piecewise continuous interventions with bounded duration $||u||_0 \leq c_0$ and level $||u||_\infty \leq c_\infty$, and

described a numerical method to determine the optimal start time t_1. The analysis in [10] represents the state-of-art, covering the full class of piecewise continuous intervention strategies, and taking properly into account the accumulation of intervention costs over time.

The results above provide elegant mathematical principles describing optimal intervention shapes, but they all are limited in one important aspect. Namely, they ignore the fact that interventions of higher magnitude usually inflict a higher societal cost, and as such cannot help in answering questions related to the trade-off between the magnitude and duration of an intervention.

2 Theoretical results

2.1 Optimal intervention strategy

Our goal is to minimize total incidence among intervention strategies with total time-aggregated cost bounded by $||u||_1 \leq c_1$ and maximum intervention level bounded by $||u||_\infty \leq c_\infty$ for some $0 < c_\infty \leq 1$. The latter constraint reflects the fact that a complete lockdown might be impossible to implement in practice. We search for a global optimum in an infinite-dimensional space of piecewise continuous intervention strategies containing multiple consecutive lockdowns, gradually imposed and lifted restrictions, and various heuristic controls based for example on tracking the effective reproduction number [11]. The following theorem shows that among all such intervention strategies, the global optimum is achieved by a simple constant-level lockdown.

Theorem 1. For any initial state with $S(0), I(0) > 0$, the total incidence among all piecewise continuous intervention strategies such that $||u||_1 \leq c_1$ and $||u||_\infty \leq c_\infty$ is minimized by an intervention of form (4) with level $c = c_\infty$, duration $t_2 - t_1 = c_1/c_\infty$, and a uniquely determined start time t_1.

Theorem 1 provides a simple answer to the simple but mathematically nontrivial question presented in the beginning, indicating that heavy lockdowns of short duration outperform light lockdowns of longer duration. The optimal start time t_1 may be numerically determined by solving a one-dimensional optimization problem, as described in [7, 10].

2.2 Upper and lower bounds

As a byproduct of the mathematical analysis needed for proving Theorem 1, we obtain universal upper and lower bounds for total incidence, valid for all intervention strategies with a finite cost. The upper bound corresponds to total incidence in the absence of interventions, and is expressed in terms of the limiting susceptible share $S_0(\infty)$ in a standard SIR epidemic with no

interventions, which is numerically obtained as the unique solution in the interval $(0, \frac{\gamma}{\beta})$ of equation

$$S_0(\infty) - \frac{\gamma}{\beta} \log S_0(\infty) \;=\; S(0) + I(0) - \frac{\gamma}{\beta} \log S(0).$$

Theorem 2. For any initial state with $S(0), I(0) > 0$ and for any intervention strategy with finite cost $||u||_1 < \infty$, the total incidence is at least $1 - \gamma/(\beta S(0))$ and at most $1 - S_0(\infty)/S(0)$

3 Numerical results

The performance of the optimal strategy in Theorem 1 is investigated by numerical simulations[1] using parameters with basic reproduction number $\mathcal{R}_0 = 3$ and an average infectious period of 5 days (Table 1), roughly in line with the first wave of the COVID-19 pandemic in spring 2020, e.g. [12]. The initial state corresponds to an importation of 1000 infectious individuals into a population of size 10 million.

Parameter	Value	Meaning
β	0.6	Transmission rate (per day)
γ	0.2	Recovery rate (per day)
$S(0)$	0.9999	Initial susceptible share
$I(0)$	0.0001	Initial infectious share

Table 1: Parameters used in numerical simulations.

3.1 Performance

When an epidemic with a basic reproduction number $\mathcal{R}_0 = 3$ hits an overwhelmingly susceptible population, eventually 94.0% of individuals will become infected if no interventions are imposed. Furthermore, at least $1 - 1/(S(0)\mathcal{R}_0) = 66.6\%$ will become infected under an arbitrary intervention with a finite total cost (Theorem 2). The top panel in Fig. 1 displays the total incidence that is achievable using optimal strategies with total cost $||u||_1$ bounded by $c_1 = 7.5, 15, 30$ and maximum intervention level bounded by $0 < c_\infty \leq 1$. It is seen that interventions with even a modest budget of $c_1 = 7.5$ may significantly reduce total incidence if they can be imposed at a sufficient magnitude. On the other hand, long but mild interventions have little effect on total incidence.

The bottom panel in Fig. 1 shows that the higher the maximal intervention level c_∞, the longer one should wait before imposing the intervention,

[1] https://github.com/lasseleskela/epidemic-models-with-control.

and that mild interventions should be started immediately. For a complete lockdown ($c_\infty = 1$), the optimal timing is to wait for 24.94 days, corresponding to the time instant at which the uncontrolled epidemic reaches herd immunity.

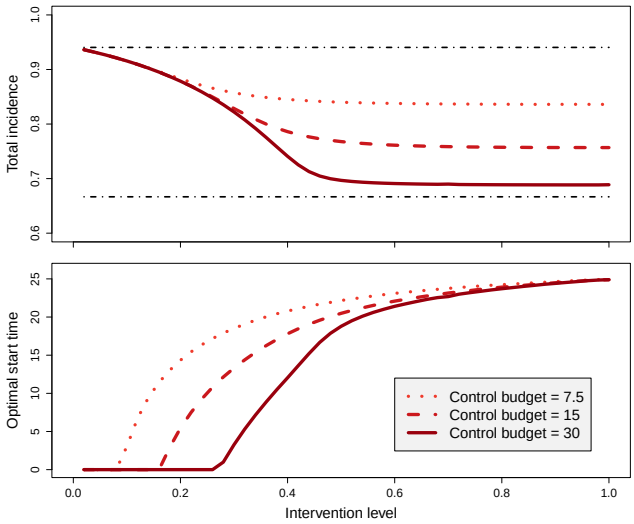

Figure 1: Minimum total incidence (top panel) achievable using interventions with total cost $||u||_1 \leq c_1$ and maximum level $||u||_\infty \leq c_\infty$ for $c_1 = 7.5, 15, 30$, and a full range of c_∞ (horizontal axis). The horizontal lines at levels 0.666 and 0.940 indicate the lower and upper bounds of Theorem 2. The bottom panel displays the start times of the optimal strategies corresponding to the (c_1, c_∞)-pairs.

3.2 Optimal start time

Fig. 2 displays the total incidence of a 20-day constant-level lockdown having level 0.75 and total cost 15, for different values of the start time, and underlines the importance of proper timing of such interventions. Mistiming the lockdown even by just one week may have a big effect on number of eventually infected individuals. Somewhat strikingly, starting too early is about equally as bad as starting too late. Similar findings have been reported in [10, 13]. Proper timing is crucial also when minimizing peak prevalence, as noted in [3].

3.3 Lack of monotonicity

Fig. 3 displays a striking phenomenon where adding restrictions before the start of an optimal intervention leads to more individuals eventually becoming infected. The original intervention (blue) with a duration 20 days and

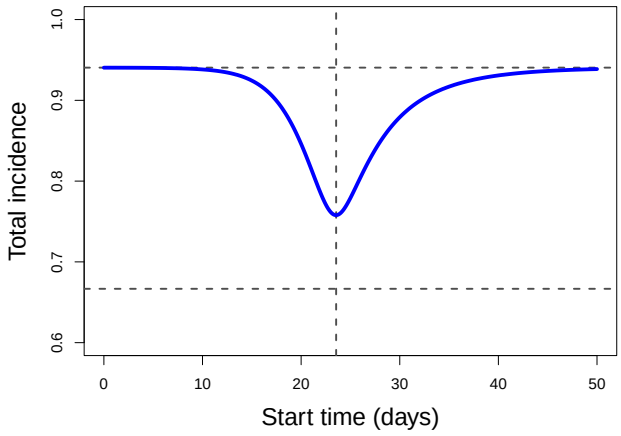

Figure 2: Total incidence of a constant-level lockdown of duration 20 days and level 0.75, for different values of the start time. The optimal start time $t_1 = 23.6$ yields a total incidence of 0.758. The horizontal lines at levels 0.666 and 0.940 indicate the lower and upper bounds of Theorem 2.

a constant level of 0.75, is started optimally at time 23.6. An alternative intervention (red) has the same level and end time but is started seven days earlier. Such a modification, aimed at mitigating disease burden by adding more restrictions, actually has an opposite effect and leads to 19.7% *more* infections in the long run. This negative outcome is caused by a second wave of infections which starts when the alternative intervention is lifted. This again underlines the important of proper timing: if the longer longer 27-day lockdown were started optimally at time 23.4, the total incidence would have been 0.723 which is not much higher than the best possible outcome of $1 - 1/(S(0)\mathcal{R}_0) = 0.666$ given in Theorem 2. Similar observations manifesting the counterintuitive lack of monotonicity has been reported in [14–16]. On a positive note, extending interventions from the end will never do harm in this way (see Lemma 8 in the appendix).

3.4 Comparison with peak-minimizing strategy

Fig. 4 displays time plots of a constant-level lockdown of type (4) designed to minimize total incidence and an intervention of type (3) designed to minimize peak prevalence. Both intervention strategies are constrained by a total budget $c_1 = 15$ and maximum intervention level $c_\infty = 0.75$. The latter strategy imposes restrictions earlier, and when these are lifted, the share of infectious individuals in the population is still relatively high, leading to more additional infections and a higher total incidence. As expected, the former strategy yields a lower total incidence than the latter, but at the cost of higher peak prevalence.

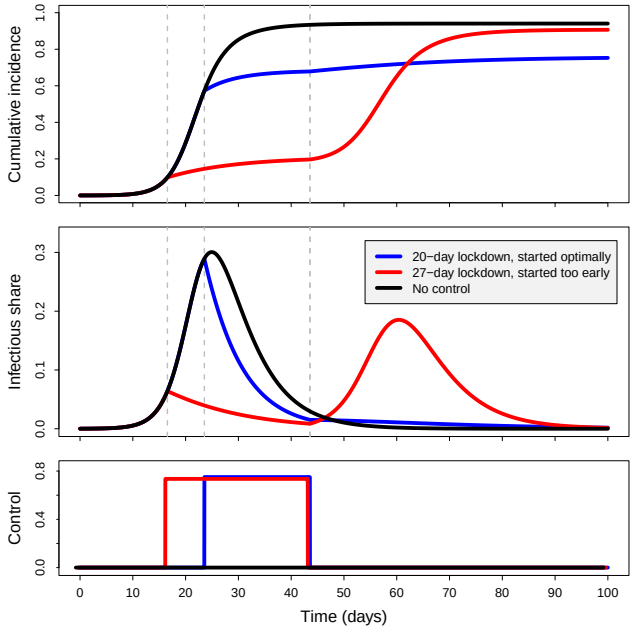

Figure 3: Intervening too early may do more harm than good. An optimally timed 20-day lockdown at 0.75 level is in effect during time interval (23.6,43.6) and leads to a total incidence of 0.758 (blue). A longer but badly timed lockdown at the same level is effective during time interval (16.6, 43.6) and leads to a total incidence of 0.907 (red) which is not far from the total incidence of 0.940 corresponding to an uncontrolled epidemic (black). For the longer 27-day lockdown, the optimal start time is 23.4 and yields total incidence 0.723.

4 Discussion

It was proven that, among all intervention strategies with cumulative cost not exceeding c_1 and magnitude never exceeding c_∞, a single lockdown at level c_∞, duration c_1/c_∞ started at an optimal time instant minimizes the total number of individuals eventually getting infected. As a consequence, assuming that a fixed proportion of infected individuals end up in hospitals and another fixed proportion will die, this strategy also minimizes the cumulative number of hospitalizations and case fatalities. We also saw that the total incidence is at least $1 - 1/(S(0)\mathcal{R}_0)$ for any intervention with a finite total cost. This reflects the fact that in the absence of future vaccinations, the prevalence would start rising again if the susceptible share would be above the herd immunity level $1 - 1/\mathcal{R}_0$ at a time when interventions are relaxed.

Numerically it was seen that milder restrictions for a longer time had very little effect on the total incidence, and also, somewhat surprisingly,

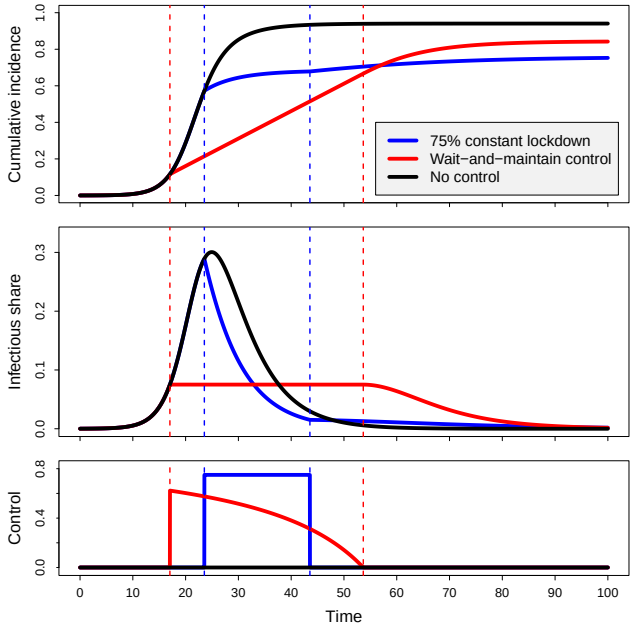

Figure 4: Time plots of cumulative incidence $1-S(t)/S(0)$ (top), prevalence $I(t)$ (middle), and intervention level $u(t)$ (bottom) for three strategies: (a) Total-incidence-minimizing single lockdown of 20 days at 75% level with $||u||_1 = 15$, and optimized start time = 23.6 (blue). (b) Peak-prevalence-minimizing strategy (3) with start time 17.0 and duration 36.6 corresponding to $||u||_1 = 15$ (red). (c) No control (black). The total incidence values for the three strategies are 0.758, 0.843, and 0.940. The corresponding peak prevalence values are 0.289, 0.075, and 0.300.

that restrictions imposed very early are not as effective as waiting until the prevalence has grown substantially before inserting maximal prevention. Rather counterintuitively, adding restrictions prior to the beginning of an existing intervention strategy may even result in a *larger* total incidence. On the contrary, as a by-product of our main proof, it was proven that adding restrictions *after* an intervention strategy has ended can only reduce the total incidence.

The optimization problem is formulated without having any fixed time horizon in mind. If for example a vaccine was known to become available not too far into the future, then this could lead to a rather different optimal prevention. Another assumption was that we were only willing to spend a finite cost c_1 for the cumulative preventions. For a severe disease this might not be the case — we might be willing to keep some restrictive level for a very long time. In such a case the optimal solution will also be different.

A common feature of epidemic models based on deterministic differential

equations is that the fraction of infected individuals never exactly reaches zero. In real epidemics, this might happen due to stochastic finite-population effects. In such cases, a feasible intervention strategy might be to aim for an early elimination of the disease by imposing massive restrictions early on, as has been advocated with SARS-CoV-2 in certain countries. The performance of such strategies cannot be analysed using the type of deterministic models considered here.

The class of interventions considered was assumed to have a bounded linear cost integrated over time. It is not obvious that the societal cost of preventive measures act linearly on the amount of prevention, and therefore nonlinear cost functionals might be relevant to consider. Similarly, changing an existing intervention level might incur additional societal costs, especially if such changes are carried out frequently. The analysis of optimal intervention strategies under such cost functions remains an open problem worthy of attention.

Other extensions worth considering would be to make the underlying epidemic model more realistic by incorporating seasonal effects, adding a latency period between which individuals have been infected but are not yet infectious, including a delay for the time it takes for an intervention decision to take effect, acknowledging different types of individuals and social structures such as households and workplaces, and relaxing the assumption that individuals recover at a constant rate. However, we do not expect major qualitative changes from these extensions, as opposed to considering nonlinear cost functions or unlimited intervention budgets.

5 Proofs

5.1 Proof of Theorem 1

Denote by $U(c_1, c_\infty)$ the set of piecewise continuous functions $u : [0, \infty) \to [0, 1]$ such that $||u||_1 \leq c_1$ and $||u||_\infty \leq c_\infty$. Denote by $J(u)$ the total incidence corresponding to intervention strategy $u \in U(c_1, c_\infty)$. Denote

$$J_* \;=\; \inf_{u \in U(c_1, c_\infty)} J(u).$$

Intuition suggests that interventions carried out in a distant future should have a negligible effect on the evolution of the epidemic. Proposition 4 in the appendix confirms this and tells that it is possible to select constants $C, \alpha, T_* > 0$ such that

$$\int_T^\infty I_u(t)\, dt \;\leq\; Ce^{-\alpha T} \tag{5}$$

for all $T \geq T_*$ and all controls bounded by $||u||_1 \leq c_1$. Let us now fix $\epsilon > 0$ and choose $u_1 \in U(c_1, c_\infty)$ such that

$$J(u_1) \leq J_* + \epsilon. \tag{6}$$

(i) *Truncation.* We will approximate u_1 by a control $u_2 = 1_{[0,T]}u_1$ of finite duration, where we choose a large enough $T > T_*$ so that $Ce^{-\alpha T} \leq \epsilon$. Because the trajectories (S_{u_1}, I_{u_1}) and (S_{u_2}, I_{u_2}) coincide up to time T, we see that

$$|J(u_2) - J(u_1)| \;=\; \frac{\gamma}{S(0)} \left| \int_T^\infty I_{u_2}(t)\,dt - \int_T^\infty I_{u_1}(t)\,dt \right|$$

Because $||u_2||_1 \leq ||u_1||_1 \leq c_1$, inequality (5) tells that both integrals on the right are at most $Ce^{-\alpha T}$, and it follows that

$$|J(u_2) - J(u_1)| \;\leq\; 2\epsilon \frac{\gamma}{S(0)}. \tag{7}$$

(ii) *Quantization.* We will next approximate u_2 by a bang–bang control u_3 defined as follows. We divide time into small intervals $I_k = ((k-1)h, kh]$ of length $0 < h \leq \epsilon e^{-(\beta+\gamma)T}$. We quantize the control function u_2 in a frequency modulated fashion so that in each time I_k, first $u_3 = 0$ for $h - \tau_k$ time units, and then $u_3 = c_\infty$ for the remaining τ_k time units, where

$$\tau_k \;=\; b^{-1} \int_{hk-h}^{hk} u_2(t)\,dt$$

is selected so that $\int_{I_k} u_3(t)\,dt = \int_{I_k} u_2(t)\,dt$, see Figure 5. As a consequence of $h \leq \epsilon e^{-(\beta+\gamma)T}$, it follows (details in Proposition 6) that

$$\begin{aligned} \int_0^T |I_{u_3}(t) - I_{u_2}(t)|\,dt \;&\leq\; \int_0^T 3\beta h e^{(\beta+\gamma)t}\,dt \\ &\leq\; 3\beta h(\beta+\gamma)^{-1} e^{(\beta+\gamma)T} \;\leq\; 3\epsilon. \end{aligned}$$

Furthermore,

$$\int_T^\infty |I_{u_3}(t) - I_{u_2}(t)|\,dt \;\leq\; \int_T^\infty I_{u_2}(t)\,dt + \int_T^\infty I_{u_3}(t)\,dt,$$

and by noting that $||u_3||_1 = ||u_2||_1 \leq c_1$ (Lemma 5), inequality (5) again guarantees that both integrals on the right are at most $Ce^{-\alpha T} \leq \epsilon$. Hence it follows that

$$|J(u_3) - J(u_2)| \;\leq\; 5\epsilon \frac{\gamma}{S(0)}. \tag{8}$$

(iii) *Prolongation.* By construction, $u_3 = c_\infty 1_A$ where A is a finite union of intervals of total length $|A|$. Because $||u_3||_1 = c_\infty |A|$ and $||u_3||_1 = ||u_2||_1 \leq c_1$, we find that $|A| \leq c_1/c_\infty$. Intuition suggests that a smaller disease burden will be incurred by replacing A by a larger subset of the time axis, but such monotonicity properties are known to fail in general (see for example [14, 15]). Nevertheless, prolonging the last interval of A from the end does increase the final susceptible share (details in Lemma 8). Let us

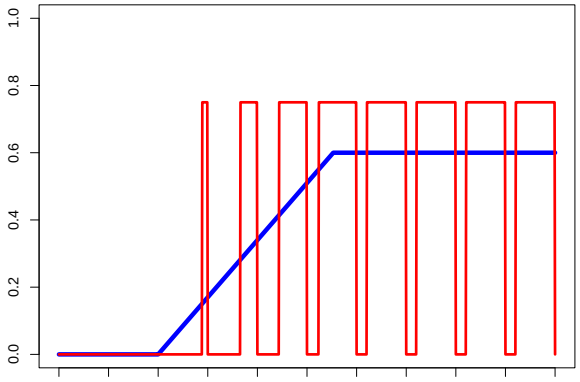

Figure 5: Quantization of a function u (blue) by frequency modulated function $\hat{u}$ with amplitude 0.75 (red).

define $u_4 = c_\infty 1_{\tilde{A}}$ where $\tilde{A}$ is obtained by prolonging the last interval in A from the end so that $|\tilde{A}| = c_1/c_\infty$. Then it follows that $S_{u_4}(\infty) \geq S_{u_3}(\infty)$, and therefore, by recalling the equation $S(0) + I(0) - S_u(\infty) = \gamma||I_u||_1$, we find that

$$J(u_4) \leq J(u_3). \tag{9}$$

(iv) *Merging.* The control u_4 belongs to the set $\tilde{U}$ of all controls of the form $u = c_\infty 1_B$ where $B \subset [0, \infty)$ is a union of finitely many disjoint intervals of total length c_1/c_∞. Among those, the best for minimizing cumulative incidence are those where all disjoint intervals are merged into a single interval of length c_1/c_∞. Especially, it is known [7, Theorem 1.1] that there exists $\sigma \geq 0$ such that the control $u_5 = c_\infty 1_{(\sigma,\sigma+c_1/c_\infty)}$ satisfies $J(u_5) = \inf_{u\in\tilde{U}} J(u)$. Especially,

$$J(u_5) \leq J(u_4). \tag{10}$$

By collecting the inequalities (6)–(10) together, we conclude that

$$J(u_5) \leq J_* + 8\epsilon\frac{\gamma}{S(0)}.$$

To summarize, for any $\epsilon > 0$ there exists $\sigma \geq 0$ such that the single-lockdown control $u_5 = c_\infty 1_{(\sigma,\sigma+c_1/c_\infty)}$ satisfies the above inequality. We conclude that

$$\inf_{u\in U(c_1,c_\infty)} J(u) = \inf_{\sigma\geq 0} J(v_\sigma), \tag{11}$$

where $v_\sigma = c_\infty 1_{(\sigma,\sigma+c_1/c_\infty)}$ denotes a single-lockdown control with start time σ, duration c_1/c_∞, and constant intervention level c_∞.

(v) *Timing.* Finally, we will verify that the infimum in (11) is attained by an optimally chosen start time $\sigma_* \geq 0$. Denote $\hat{U} = \cup_{\sigma\geq 0}\hat{U}_\sigma$ where $\hat{U}_\sigma$

is the set of controls such that $||u||_\infty \leq c_\infty$ and $u = 0$ outside the time interval $[\sigma, \sigma + c_1/c_\infty]$. Then [10, Theorem 1] implies that there exists a unique $\sigma_* \geq 0$ such that

$$\sup_{u \in \hat{U}} S_u(\infty) \;=\; \sup_{\sigma \geq 0} \sup_{u \in \hat{U}_\sigma} S_u(\infty) \;=\; S_{v_{\sigma_*}}(\infty)$$

for $v_{\sigma_*} = c_\infty 1_{(\sigma, \sigma + c_1/c_\infty)}$. Therefore, by recalling the equation $S(0) + I(0) - S_u(\infty) \;=\; \gamma ||I_u||_1.$, it follows that

$$\inf_{u \in \hat{U}} J(u) \;=\; J(v_{\sigma_*}).$$

Because $\hat{U}$ contains all controls of the form $v_\sigma = c_\infty 1_{(\sigma, \sigma + c_1/c_\infty)}$, we conclude that

$$J(v_{\sigma_*}) \;=\; \inf_{\sigma \geq 0} J(v_\sigma).$$

In light of (11), this confirms the statement of the theorem. □

5.2 Proof of Theorem 2

For a standard SIR model with no interventions ($u = 0$), it is well known that the limiting susceptible share is bounded by $S(\infty) \leq \frac{\gamma}{\beta}$. By a careful reasoning (Proposition 3 in the appendix), the same bound extends to epidemics controlled by an intervention with $||u||_1 < \infty$. Therefore, the total incidence is bounded from below by $1 - S(\infty)/S(0) \geq 1 - \gamma/(\beta S(0))$.

For the upper bound, it appears intuitively clear that the limiting susceptible share $S_u(\infty)$ in an epidemic with an arbitrary control u is larger than equal than the corresponding quantity $S_0(\infty)$ in an epidemic with no interventions. Indeed, Chikina and Pegden [14] have shown that a pointwise ordering of interventions $u_1 \leq u_2$ implies $S_{u_1}(\infty) \leq S_{u_2}(\infty)$ under an extra assumption that u_2 is nondecreasing. Straightforward modifications of the analysis in [14] show that the above implication holds also when u_1 instead of u_2 is required to be nondecreasing. By selecting $u_1 = 0$ and letting $u_2 = u$ be an arbitrary piecewise continuous intervention, we find that $S_0(\infty) \leq S_u(\infty)$, and we conclude that the total incidence is bounded from above by $1 - S_u(\infty)/S(0) \leq 1 - S_0(\infty)/S(0)$. □

Acknowledgments The authors are grateful for financial support from project 105572 NordicMathCovid as part of the Nordic Programme on Health and Welfare funded by NordForsk.

References

[1] William Ogilvy Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A*, 115(772):700–721, 1927.

[2] Odo Diekmann, Hans Heesterbeek, and Tom Britton. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, 2013.

[3] Dylan H. Morris, Fernando W. Rossine, Joshua B. Plotkin, and Simon A. Levin. Optimal, near-optimal, and robust epidemic control. *Communications Physics*, 4(1):78, 2021.

[4] James M. Greene and Eduardo D. Sontag. Minimizing the infected peak utilizing a single lockdown: a technical result regarding equal peaks. medRxiv, 2021.

[5] Laurent Miclo, Daniel Spiro, and Jörgen Weibull. Optimal epidemic suppression under an ICU constraint. arXiv, 2020.

[6] Florin Avram, Lorenzo Freddi, and Dan Goreac. Optimal control of a SIR epidemic with ICU constraints and target objectives. *Applied Mathematics and Computation*, 418:126816, 2022.

[7] Yuanyuan Feng, Gautam Iyer, and Lei Li. Scheduling fixed length quarantines to minimize the total number of fatalities during an epidemic. *Journal of Mathematical Biology*, 82(7):69, 2021.

[8] Pierre-Alexandre Bliman, Michel Duprez, Yannick Privat, and Nicolas Vauchelet. Optimal immunity control and final size minimization by social distancing for the SIR epidemic model. *Journal of Optimization Theory and Applications*, 189(2):408–436, 2021.

[9] David I. Ketcheson. Optimal control of an sir epidemic through finite-time non-pharmaceutical intervention. *Journal of Mathematical Biology*, 83(1), Jun 2021.

[10] Pierre-Alexandre Bliman and Michel Duprez. How best can finite-time social distancing reduce epidemic final size? *Journal of Theoretical Biology*, 511:110557, 2021.

[11] Leonardo Cianfanelli, Francesca Parise, Daron Acemoglu, Giacomo Como, and Asuman Ozdaglar. Lockdown interventions in SIR model: Is the reproduction number the right control variable?, 2021. arXiv.

[12] Seth Flaxman, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W. Eaton, Mélodie Monod, Pablo N. Perez-Guzman, Nora Schmit, Lucia Cilloni, Kylie E. C. Ainslie, Marc Baguelin, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, Laura V. Cooper, Zulma Cucunubá, Gina Cuomo-Dannenburg,

Amy Dighe, Bimandra Djaafara, Ilaria Dorigatti, Sabine L. van Elsland, Richard G. FitzJohn, Katy A. M. Gaythorpe, Lily Geidelberg, Nicholas C. Grassly, William D. Green, Timothy Hallett, Arran Hamlet, Wes Hinsley, Ben Jeffrey, Edward Knock, Daniel J. Laydon, Gemma Nedjati-Gilani, Pierre Nouvellet, Kris V. Parag, Igor Siveroni, Hayley A. Thompson, Robert Verity, Erik Volz, Caroline E. Walters, Haowei Wang, Yuanrong Wang, Oliver J. Watson, Peter Winskill, Xiaoyue Xi, Patrick G. T. Walker, Azra C. Ghani, Christl A. Donnelly, Steven Riley, Michaela A. C. Vollmer, Neil M. Ferguson, Lucy C. Okell, Samir Bhatt, and Imperial College COVID-19 Response Team. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020.

[13] Francesco Di Lauro, István Z. Kiss, and Joel C. Miller. Optimal timing of one-shot interventions for epidemic control. *PLOS Computational Biology*, 17(3):1–25, 03 2021.

[14] Maria Chikina and Wesley Pegden. Failure of monotonicity in epidemic models, 2020.

[15] Andreas Handel, Ira M Longini, and Rustom Antia. What is the best control strategy for multiple infectious disease outbreaks? *Proceedings of the Royal Society B: Biological Sciences*, 274(1611):833–837, 2007.

[16] Thomas Kruse and Philipp Strack. Optimal control of an epidemic through social distancing. SSRN Preprint, 2020.

[17] Hassan K. Khalil. *Nonlinear Systems*. Prentice Hall, 1995.

A Truncation

The appendices present technical details needed for proving Theorem 1. Appendix A contains details related to approximating a control of infinite time horizon by truncation. Appendix B contains an analysis of a frequency-modulated quantization operator. Appendix C contains a monotonicity property related to minimizing the disease burden by prolonged interventions.

A.1 Time to reach herd immunity

For an epidemic trajectory (S_u, I_u) controlled by u, we denote the time at which herd immunity is reached by

$$t_H(u) \;=\; \inf\left\{ t \geq 0 : S_u(t) \leq \frac{\gamma}{\beta} \right\}.$$

The following result generalizes [7, Lemma 3.3:(2)] (and corrects a minor mistake in its proof).

Proposition 3. *For any initial state with $S(0) > \frac{\gamma}{\beta}$ and $I(0) > 0$, and any piecewise continuous control such that $||u||_1 < \infty$, the time to reach herd immunity is finite and bounded by*

$$t_H(u) \;\leq\; ||u||_1 + \frac{\log(\frac{\beta}{\gamma} S(0))}{\beta I(0)} e^{\gamma ||u||_1}.$$

Proof. We will analyse the system on the time interval up to $t_H = t_H(u)$. Observe that the logarithmic state variables evolve according to

$$(\log S_u)' \;=\; -\beta(1-u) I_u, \tag{12}$$

$$(\log I_u)' \;=\; \beta(1-u) S_u - \gamma. \tag{13}$$

By definition, $S_u \geq \frac{\gamma}{\beta}$ on $[0, t_H]$. This lower bound combined with (13) implies that $(\log I_u)' \geq -\gamma u$, and therefore

$$I_u(t) \;\geq\; I(0) e^{-\gamma \int_0^t u(s) ds} \;\geq\; I(0) e^{-\gamma ||u||_1} \tag{14}$$

on $[0, t_H]$. By denoting $I_{\min} = I(0) e^{-\gamma ||u||_1}$ and combining the above inequality with (12), it follows that $(\log S_u)' \leq -\beta I_{\min}(1-u)$ on $[0, t_H]$, so that

$$\begin{aligned} \log S_u(t_H) - \log S(0) \;&\leq\; -\beta I_{\min} \int_0^{t_H} (1 - u(t))\, dt \\ &\leq\; \beta I_{\min}(||u||_1 - t_H). \end{aligned}$$

By noting that $S_u(t_H) = \frac{\gamma}{\beta}$ by the continuity of S_u, we see that

$$t_H \;\leq\; ||u||_1 + \frac{\log S(0) - \log \frac{\gamma}{\beta}}{\beta I_{\min}},$$

and the claim follows. □

A.2 Uniform integrability

The following result shows that the collection of infectious trajectories induced by controls bounded by $||u||_1 \leq c_1$ is uniformly integrable.

Proposition 4. *For any $\beta, \gamma > 0$, any initial state with $S(0) > 0$ and $I(0) > 0$, and any $c_1 \geq 0$, there exist constants $\alpha, C, T_* > 0$ such that*

$$\sup_{||u||_1 \leq c_1} \int_T^\infty I_u(t)\, dt \;\leq\; Ce^{-\alpha T} \qquad \textit{for all } T \geq T_*.$$

Proof. (i) Proposition 3 implies that for all controls with $||u||_1 \leq c_1$, the time to reach herd immunity is bounded by the constant

$$t_H^* \;=\; \max\{c_1 + \beta^{-1} I(0)^{-1} \log(\frac{\beta}{\gamma} S(0)) e^{\gamma c_1}, 0\}.$$

Therefore, the susceptible share satisfies $S_u(t) \leq \frac{\gamma}{\beta}$ from time t_H^* onwards. We will choose a slightly larger time horizon $T_* = t_H^* + c_1 + 1$ and show that

$$S_u(t) \;\leq\; (1-\delta)\frac{\gamma}{\beta} \tag{15}$$

for all $t \geq T_*$ and all controls u such that $||u||_1 \leq c_1$, where $\delta = 1 - \exp(-\beta I(0) e^{-\gamma T_*})$. To verify (15), observe first that

$$\int_{t_H^*}^{T_*} (1 - u(t))\, dt \;=\; c_1 + 1 - \int_{t_H^*}^{T_*} u(t)\, dt \;\geq\; c_1 + 1 - ||u||_1 \;\geq\; 1. \tag{16}$$

The crude lower bound $(\log I_u)' = \beta(1-u)S_u - \gamma \geq -\gamma$ implies that $I(t) \geq I(0)e^{-\gamma T_*}$ on $(0, T_*)$. By noting that $(\log S)' = -\beta(1-u)I$, it follows by (16) that

$$\log S_u(T_*) - \log S_u(t_H^*) \;=\; -\beta \int_{t_H^*}^{T_*} (1 - u(t)) I_u(t)\, dt \;\leq\; -\beta I(0) e^{-\gamma T_*}.$$

By noting that the right side above equals $\log(1-\delta)$, and that $S_u(t_H^*) \leq \frac{\gamma}{\beta}$ due to $t_H(u) \leq t_H^*$, we conclude that $\log S_u(T_*) \leq \log \frac{\gamma}{\beta} + \log(1-\delta)$, and that (15) is valid.

(ii) We will next derive an upper bound for the tail integrals of I_u. By applying (15), we see that for all $t > T_*$,

$$(\log I_u)' \;=\; \beta(1-u)S_u - \gamma \;\leq\; \beta(1-\delta)\frac{\gamma}{\beta} - \gamma \;=\; -\gamma\delta,$$

so that

$$I_u(t) \;\leq\; I_u(T_*) e^{-\gamma\delta(t - T_*)} \;\leq\; e^{-\gamma\delta(t-T_*)}.$$

By integrating the above inequality, we see that

$$\int_T^\infty I_u(t)\, dt \;\leq\; \frac{1}{\gamma\delta} e^{-\gamma\delta(T - T_*)} \qquad \text{for all } T \geq T_*.$$

Therefore, the claim holds for $\alpha = \gamma\delta$ and $C = \frac{1}{\gamma\delta} e^{\gamma\delta T_*}$. □

B Quantization

We develop a frequency modulation approach to approximate a general piecewise continuous function with a square waveform having a small wavelength. Denote by $U(b)$ the set of piecewise continuous functions $u : [0, \infty) \to [0, 1]$ such that $||u||_\infty \leq b$. Given an amplitude $b > 0$ and wavelength $h > 0$, we define a quantization operator $Q_{b,h} : U(b) \to U(b)$ by setting $Q_{b,h}u = \hat{u}$ with

$$\hat{u}(t) \;=\; \begin{cases} b & \text{if } t \in I, \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

where $I = \cup_{k\geq 1} I_k$ is a union of disjoint intervals $I_k = [hk - \tau_k, hk)$ having lengths

$$\tau_k \;=\; b^{-1} \int_{hk-h}^{hk} u(t)\, dt.$$

We will show that $\hat{u}$ approximates u well in a weak sense for small wavelengths. We first prove a general approximation property (Lemma 5), and then derive an approximation result of an epidemic trajectory controlled by $\hat{u}$ (Proposition 6).

A function $\phi : [0, \infty) \to \mathbb{R}$ is called locally bounded if $||\phi||_{\infty,t} = \sup_{0\leq s\leq t} |\phi(s)|$ is finite for all t, and locally Lipschitz continuous if $||\phi||_{\text{Lip},t} = \sup_{0\leq t_1<t_2\leq t} \frac{|\phi(t_2)-\phi(t_1)|}{t_2-t_1}$ is finite for all t.

Lemma 5. *For any $b, h > 0$ and $u \in U(b)$, the approximation $\hat{u} = Q_{b,h}u$ defined by (17) satisfies $||\hat{u}||_1 = ||u||_1$, and*

$$\left| \int_0^t \big(\hat{u}(s) - u(s)\big)\phi(s)\, ds \right| \;\leq\; bh\,(||\phi||_{\infty,t} + t||\phi||_{\text{Lip},t})$$

for all $t \geq 0$ and all locally bounded and locally Lipschitz continuous ϕ.

Proof. On each interval $[hk-h, hk)$, the quantized control alternates so that $\hat{u} = 0$ for the first $h - \tau_k$ time units, and then $\hat{u} = b$ for the remaining τ_k time units. The choice of τ_k then implies that

$$\int_{hk-h}^{hk} \hat{u}(s)\, ds \;=\; \int_{hk-h}^{hk} u(s)\, ds \tag{18}$$

for all $k \geq 1$. By summing both sides of the above equality with respect to $k \geq 1$, we find that $||\hat{u}||_1 = ||u||_1$.

Let us now fix $t \geq 0$. Denote $n = \lfloor \frac{t}{h} \rfloor$ and $t_k = kh$, and observe that $\int_0^t \big(\hat{u}(s) - u(s)\big)\phi(s)\, ds = \sum_{k=1}^n A_k + B_n$, where $A_k = \int_{t_{k-1}}^{t_k} \big(\hat{u}(s) - u(s)\big)\phi(s)\, ds$ and $B_n = \int_{t_n}^t \big(\hat{u}(s) - u(s)\big)\phi(s)\, ds$. By applying (18), we find that

$$A_k \;=\; \int_{t_{k-1}}^{t_k} \big(\hat{u}(s) - u(s)\big)\big(\phi(s) - \phi(t_{k-1})\big)\, ds.$$

Because $|\phi(s) - \phi(t_{k-1})| \leq h||\phi||_{\text{Lip},t}$ for all $t_{k-1} < s < t_k$, and $||\hat{u}-u||_\infty \leq b$, it follows that

$$|A_k| \;\leq\; h||\phi||_{\text{Lip},t} \int_{t_{k-1}}^{t_k} \big|\hat{u}(s) - u(s)\big|\, ds \;\leq\; bh^2||\phi||_{\text{Lip},t}.$$

Furthermore, by noting that $t_n \leq t < t_n + h$, it follows that

$$|B_n| \;\leq\; ||\phi||_{\infty,t} \int_{t_n}^{t} \big|\hat{u}(s) - u(s)\big|\, ds \;\leq\; bh||\phi||_{\infty,t}.$$

We conclude that

$$\begin{aligned}\left|\int_0^t \big(\hat{u}(s) - u(s)\big)\phi(s)\, ds\right| \;&\leq\; \sum_{k=1}^{n} |A_k| + |B_n| \\ &\leq\; bnh^2||\phi||_{\text{Lip},t} + bh||\phi||_{\infty,t} \\ &\leq\; bht||\phi||_{\text{Lip},t} + bh||\phi||_{\infty,t}.\end{aligned}$$

□

Proposition 6. *For any $b, h > 0$ and $u \in U(b)$, any $\beta, \gamma > 0$ and any initial state with $S(0), I(0) > 0$, the epidemic trajectories (S_u, I_u) and $(S_{\hat{u}}, I_{\hat{u}})$ associated with u and $\hat{u} = Q_{b,h}u$ satisfy*

$$\max\Big\{|S_{\hat{u}}(t) - S_u(t)|,\, |I_{\hat{u}}(t) - I_u(t)|\Big\} \;\leq\; 3\beta bhe^{(\beta+\gamma)t}$$

for all $t \geq 0$.

Proof. The epidemic trajectory (S_u, I_u) controlled by u is the unique solution to $X'(t) = f(t, X(t))$, $X(0) = (S(0), I(0))$, where $f : [0, \infty) \times \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

$$f(t, X) \;=\; \begin{bmatrix} -\alpha(t)X_1X_2, \\ \alpha(t)X_1X_2 - \gamma X_2 \end{bmatrix} \tag{19}$$

and $\alpha(t) = \beta(1 - u(t))$. The state space of the system is denoted by $\mathcal{X} = \big\{X \in [0,1]^2 : X_1 + X_2 \leq 1\big\}$ and we equip it with max norm $||X|| = \max\{|X_1|, |X_2|\}$. The Jacobian matrix of $X \mapsto f(t, X)$ equals

$$\partial_X f(t, X) \;=\; \begin{bmatrix} -\alpha(t)X_2 & -\alpha(t)X_1 \\ \alpha(t)X_2 & \alpha(t)X_1 - \gamma \end{bmatrix}.$$

The operator norm of the matrix $\partial_X f(t, X)$ induced by the max norm on $\mathbb{R}^2$ is the maximum absolute row sum, which is bounded by

$$\begin{aligned}||\partial_X f(t, X)|| \;&=\; \max\Big\{\alpha(t)(|X_1| + |X_2|),\, \alpha(t)|X_2| + |\alpha(t)X_1 - \gamma|\Big\} \\ &\leq\; \alpha(t)(|X_1| + |X_2|) + \gamma.\end{aligned}$$

Hence $||\partial_X f(t,X)|| \leq \beta+\gamma$ for all $X \in \mathcal{X}$ and $t \geq 0$, it follows [17, Lemma 2.2] that f is Lipschitz continuous according to

$$||f(t,X) - f(t,Y)|| \leq (\beta+\gamma)||X-Y|| \quad \text{for all } X,Y \in \mathcal{X} \text{ and } t \geq 0. \quad (20)$$

Let $Y : [0,\infty) \to \mathcal{X}$ be the unique solution to $Y'(t) = g(t,Y(t))$, $Y(0) = (S(0), I(0))$, where $g(t,X)$ is defined analogously to $f(t,X)$ but with $\alpha(t)$ in (19) replaced by $\hat{\alpha}(t) = \beta(1-\hat{u}(t))$. Then

$$\begin{aligned} X(t) &= X(0) + \int_0^t f(s,X(s))\,ds, \\ Y(t) &= Y(0) + \int_0^t g(s,Y(s))\,ds. \end{aligned}$$

Because $X(0) = Y(0)$, we find that the difference $Z(t) = Y(t) - X(t)$ satisfies

$$\begin{aligned} Z(t) &= \int_0^t \Big(f(s,Y(s)) - f(s,X(s))\Big)\,ds \\ &\quad + \int_0^t \Big(g(s,Y(s)) - f(s,Y(s))\Big)\,ds. \end{aligned}$$

Inequality (20) shows that $||f(s,Y(s)) - f(s,X(s))|| \leq (\beta+\gamma)||z(s)||$ for all s, and we conclude that the max norm of $Z(t)$ is bounded by

$$||Z(t)|| \leq \int_0^t (\beta+\gamma)||Z(s)||\,ds + h(t), \quad (21)$$

where

$$h(t) = \left|\left| \int_0^t \Big(g(s,Y(s)) - f(s,Y(s))\Big)\,ds \right|\right|.$$

Observe next that

$$g(s,Y(s)) - f(s,Y(s)) = \beta(u(s)-\hat{u}(s))\phi(s)\begin{bmatrix}-1\\+1\end{bmatrix},$$

where $\phi(s) = Y_1(s)Y_2(s)$, and hence

$$h(t) = \beta\left|\int_0^t (u(s)-\hat{u}(s))\phi(s)\,ds\right|.$$

We also find that $|Y_1'(t)| = |S_{\hat{u}}(t)| \leq \beta$ and $|Y_2'(t)| = |I_{\hat{u}}(t)| \leq \beta+\gamma$ for all $t \geq 0$. Therefore, $|\phi'(t)| \leq 2(\beta+\gamma)t$, and we conclude that ϕ is globally Lipschitz continuous according to $|\phi(t)-\phi(s)| \leq 2(\beta+\gamma)|t-s|$ for all $s,t \geq 0$. Lemma 5 then implies that

$$h(t) \leq C_1 + C_2 t$$

for $C_1 = \beta bh$ and $C_2 = 2\beta bh(\beta+\gamma)$. By combining this with (21), we conclude that

$$||Z(t)|| \;\leq\; C_1 + C_2 t + \int_0^t (\beta+\gamma)||Z(s)||\, ds,$$

and Grönwall's inequality (e.g. [17, Lemma 2.1]) then implies that

$$||Z(t)|| \;\leq\; C_1 + C_2 t + \int_0^t (C_1 + C_2 s)(\beta+\gamma) e^{(\beta+\gamma)(t-s)} ds.$$

Integration by parts shows that the right side equals $C_1 e^{(\beta+\gamma)t} + \frac{C_2}{\beta+\gamma}\left(e^{(\beta+\gamma)t} - 1\right)$, from which we conclude that

$$||Z(t)|| \;\leq\; \left(C_1 + \frac{C_2}{\beta+\gamma}\right) e^{(\beta+\gamma)t} \;=\; 3\beta bh e^{(\beta+\gamma)t},$$

confirming the claim. □

C Prolongation

C.1 A special function

Lemma 7. *For any $\rho > 0$, the function $f(x) = x - \rho^{-1}\log x$ is strictly decreasing on $(0, \rho^{-1}]$, and strictly increasing on $[\rho^{-1}, \infty)$, and has a unique minimum value $y_0 = \frac{1}{\rho}(1 + \log\rho)$. The restriction of f into $(0, \rho^{-1}]$ is invertible, and the corresponding inverse function $g : [y_0, \infty) \to (0, \rho^{-1}]$ is strictly decreasing.*

Proof. Because $f'(x) = 1 - \rho^{-1}x^{-1}$ we find that $f'(x) < 0$ for $s < \rho^{-1}$, and $f'(x) > 0$ for $x > \rho^{-1}$. Hence the stated monotonicity properties follow. It also follows that f restricted to $(0, \rho^{-1}]$ has a well-defined inverse function g. The inverse function rule $g'(y) = \frac{1}{f'(g(y))}$ shows that $g'(y) < 0$ for all $y > y_0$, and confirms that g is strictly decreasing on $[y_0, \infty)$. □

C.2 Prolonged interventions imply less infections

Lemma 8. *Let (S_1, I_2) be an epidemic trajectory controlled by u_1 such that $u_1 = 0$ outside $[0, T]$. Let (S_2, I_2) be an epidemic trajectory with the same initial state but a modified control $u_2 = u_1 + c1_{[t_1,t_2]}$ with $T \leq t_1 \leq t_2$. Then $S_1(\infty) \leq S_2(\infty)$.*

Proof. The epidemic under a control u_k evolves according to $S_k' = -\beta(1-u_k)S_k I_k$ and $I_k' = \beta(1-u_k)S_k I_k - \gamma I_k$. This implies that $(S_k + I_k)' = -\gamma I_k$ and $(\log S_k)' = -\beta(1-u_k)I_k$. As a consequence, we find that

$$\Big(\beta(1-u_k)(S_k + I_k) - \gamma \log S_k\Big)' \;=\; 0 \tag{22}$$

at every time instant in which $u_k' = 0$.

Let $V_k = S_k + I_k - \frac{\gamma}{\beta}\log S_k$, the *vulnerability* of the population under control u_k. Because both controls vanish on (t_2, ∞) and $I_1(\infty) = I_2(\infty) = 0$, we find from (22) that

$$S_k(\infty) - \frac{\gamma}{\beta}\log S_k(\infty) \;=\; V_k(t_2), \quad k = 1, 2. \tag{23}$$

Because both trajectories are equal up to time t_1, we see that $V_1(t_1) = V_2(t_1)$. Because $u_1 = 0$ on the interval (t_1, t_2), (22) shows that $V_1(t_1) = V_1(t_2)$. Analogously, by noting that $u_2 = c$ on (t_1, t_2), it follows that

$$\begin{aligned}&(1-c)(S_2(t_1) + I_2(t_1)) - \frac{\gamma}{\beta}\log S_2(t_1)\\ &\quad= \;(1-c)(S_2(t_2) + I_2(t_2)) - \frac{\gamma}{\beta}\log S_2(t_2),\end{aligned}$$

which can be rewritten as

$$V_2(t_2) \;=\; V_2(t_1) + c(S_2(t_2) + I_2(t_2) - S_2(t_1) - I_2(t_1)).$$

Because $(S_2 + I_2)' = -\gamma I_2$, we see that $S_2(t) + I_2(t)$ is decreasing, and therefore,

$$V_2(t_2) \;\le\; V_2(t_1) \;=\; V_1(t_1) \;=\; V_2(t_1).$$

Because of (23), it follows (Lemma 7) that $S_1(\infty) \le S_2(\infty)$. □