medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Disease control as an optimization problem

Miguel Navascués,[1] Costantino Budroni,[2,1] and Yelena Guryanova[1]

[1] *Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria*

[2] *Faculty of Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria*

Traditionally, expert epidemiologists devise policies for disease control through a mixture of intuition and brute force. Namely, they use their know-how to narrow down the set of logically conceivable policies to a small family described by a few parameters, following which they conduct a grid search to identify the optimal policy within the set. This scheme is not scalable, in the sense that, when used to optimize over policies which depend on many parameters, it will likely fail to output an optimal disease policy in time for its implementation. In this article, we use techniques from convex optimization theory and machine learning to conduct optimizations over disease policies described by hundreds of parameters. In contrast to past approaches for policy optimization based on control theory, our framework can deal with arbitrary uncertainties on the initial conditions and model parameters controlling the spread of the disease. In addition, our methods allow for optimization over weekly-constant policies, specified by either continuous or discrete government measures (e.g.: lockdown on/off). We illustrate our approach by minimizing the total time required to eradicate COVID-19 within the Susceptible-Exposed-Infected-Recovered (SEIR) model proposed by Kissler *et al.* (March, 2020).

I. INTRODUCTION

At the time of writing, the COVID-19 pandemic had already caused more than half a million deaths worldwide; at the time of re-writing, some weeks later, this number was approaching one million. The effects of the virus have been widespread and substantial, from the collapse of healthcare systems [1–3] to the enforcement of isolation and quarantine. In the case of Nepal, the national lockdown lasted for 120 days uninterrupted [4].

In these circumstances, identifying reliable and effective disease control policies is of utmost importance. Here by "policy" we mean a deliberate intervention intended to mitigate the effects of a disease as it runs its course. In much of the mathematical literature on epidemiology, the process of generating a policy is as follows [5]: (1) based on their intuition, expert epidemiologists propose a number of suitable policies to control the disease; (2) the impact on the population of each of the considered policies is assessed through dynamical models of disease spread; (3) the outcomes of all policies are compared and a decision is taken as to which one is deemed to be the best.

This three-step process has two disadvantages. First of all, the class of policies devised by an expert could well be suboptimal, since the optimal policy (under some figure of merit) could be extremely complicated and counter-intuitive. Second, the method requires one to numerically simulate each policy. Given the exponential growth of the number of policies in the number of control parameters, the number of policies considered may be on the order of billions; hence, this strategy is not guaranteed to identify the optimal disease control policy in time to enforce it.

More sophisticated approaches for disease control rely on optimal control theory to identify a suitable policy (see, e.g., [6–9]). The starting point of all these works is that both the initial conditions (namely, the number of infected, exposed, etc.) and the model parameters (such as the disease's reproduction number) specifying the spread of the disease are known with high precision. This requirement is never met in a real-life epidemic, especially close to the outbreak, when the uncertainty of the disease's reproduction number can be very high [10, 11]. Policies derived through optimal control theory are thus not guaranteed to have the desired effects in practice.

An additional disadvantage of optimal control theory is that government measures in the model cannot be constrained to be discrete, such as a lockdown which is either on or off, a policy that so far has dominated global efforts to control the COVID-19 epidemic [12]. Last but not least, optimal control theory can only handle scenarios where the policies continuously vary over time, thus not making it compatible with observed government measures to control COVID-19, which for the most part have been applied on a weekly basis.

In this paper we introduce a general framework that maps any disease control scenario to an optimization problem. Contrary to the optimal control approach, our framework can accommodate constraints on the disease dynamics which must hold for whole regions of the initial conditions and the disease's model parameters. Our framework can enforce policies to be weekly and/or discrete. Invoking tools from optimization theory and machine learning [13], we propose efficient heuristics to solve the optimization problem and hence identify the government policy that best controls the disease.

To illustrate the power of our approach, we use these optimization techniques to generate long-term plans to fight COVID-19, under the assumption that the disease's dynamics are accurately captured by a variant of the Susceptible-

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Exposed-Infected-Recovered (SEIR) compartmental model [5] proposed in [14]. Our results confirm that optimal policies tend to be too complicated to be devised by a human.

In reality, most epidemiological models only provide short-term approximations to the spread of the disease, with long term projections becoming less and less reliable [15]. In addition, notwithstanding the enormous knowledge gathered since the initial COVID-19 outbreak, many questions remain to be answered regarding the correctness and accuracy of compartmental models such as SEIR: their basic assumptions (e.g., are recovered patients temporarily or permanently immune to the disease?); the actual value of the model's parameters (e.g., the basic reproduction number R_0); and the role of variables not modeled (e.g., age, geographic distribution, contact tracing policies, role of superspreaders).

In this regard, the goal of this work is not to propose a concrete government policy, but rather to present an efficient method to obtain an optimal one, given all the available information. To estimate the effect of our methods in a realistic scenario, we conduct a numerical simulation where we re-calculate the optimal policy plans every month, based on new, incoming data. The very final policy plan that we present requires less stringent physical distancing measures compared with the one which is not re-calculated every month. All the code used in our simulations is freely available at [16].

II. THE FRAMEWORK

Our starting point is an epidemic that affects a closed population. This assumption is not limiting, since a large ensemble of population centres where individuals are free to commute can also be modeled as a closed system [17]. To gain an understanding of how the disease spreads, it is standard to divide the population into different sectors or compartments [5] (see Fig. 1). One can define, e.g., the compartment of all those individuals who are currently infected. This compartment can, in turn, be sub-divided into different compartments, such as symptomatic/asymptomatic. Once the number of relevant compartments is fixed, one can estimate the occupation of each of them and arrange the resulting numbers in a vector $\boldsymbol{x}$. The disease is subsequently analysed by looking at how $\boldsymbol{x}$ changes with time.

In order to control or even extinguish an epidemic, governments can enforce a number of different measures: mass vaccination, physical (or social [18]) distancing measures, or even a full lockdown, are common examples of interventions aimed at fighting the disease. When and to which degree such measures are applied is determined by the *disease control policy*. Consider a disease control policy based on random vaccination campaigns, where the intervention consists of vaccinating a number of individuals at random per day. Call $v(t)$ the number of individuals vaccinated on day t. This function, between the initial and final times t_0 and t_f (i.e. on the interval $[t_0, t_f]$) determines the government's vaccination policy. Similarly, let $s(t)$ take the value 1 if the country is in lockdown on day t and 0 if it is not. Then the government's lockdown policy between times t_0 and t_f corresponds to the function $s(t)$.

$$s(t) = \begin{cases} 0 & \text{lockdown off} \\ 1 & \text{lockdown on} \end{cases} \qquad t \in [t_0, t_f]\,. \tag{1}$$

If the government is intervening both through vaccination and lockdown, then its disease control policy will be identified by *both* the functions $v(t)$ and $s(t)$.

The above are instances of *non-adaptive policies* for disease control, because the functions s, v just depend on the time t, and not, e.g., on the current value of the death toll. A general (adaptive) policy for disease control would take into account the whole past history of data gathered by the government before deciding what to do at each step. Although in the following all our proposed policies are non-adaptive, the formalism we introduce allows one to optimize over adaptive policies as well.

In conclusion, a disease control policy can always be identified with a vector function $\boldsymbol{\alpha}$ of the time t and perhaps some other observed variables $\boldsymbol{o}$, where each vector entry represents a type of government intervention at time t. In turn, we can use a variable vector $\boldsymbol{\mu} \in \mathbb{R}^n$ to parametrize the class of considered policies, that is, $\boldsymbol{\alpha}(t, \boldsymbol{o}) = \boldsymbol{\alpha}(t, \boldsymbol{o}; \boldsymbol{\mu})$. Since $\boldsymbol{\mu}$ completely determines the policy $\boldsymbol{\alpha}$, we can also regard the parameters $\boldsymbol{\mu}$ as the disease policy. We will do so from now on.

The applied policy $\boldsymbol{\mu}$ is assumed to influence the compartment occupation within the time interval $[t_0, t_f]$. That is, $\boldsymbol{x}$ is both a function of t and $\boldsymbol{\mu}$. In this work we are interested in devising policies for disease control which guarantee that the spread of the disease evolves under certain conditions. For example, any country has a fixed number of critical care capacity beds, which we denote by $\mathcal{B}_c$. At each time $t \in [t_0, t_f]$, it is desirable that the number of individuals admitted to critical care in hospitals, $\mathcal{C}(t)$, does not exceed that capacity. That is, we require that

$$\mathcal{C}(t) \leq \mathcal{B}_c, \qquad \text{for } t \in [t_0, t_f]. \tag{2}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

We will call any such condition on the evolution of the disease a *constraint*.

Finally, among all policies satisfying the desired constraints, we typically wish to identify the one that minimizes a certain quantity. For example, for many diseases, a simple physical distancing policy that satisfies the constraint (2) consists of declaring a lockdown throughout the whole time interval $[t_0, t_f]$, i.e., $s(t) = 1$ for $t \in [t_0, t_f]$. This policy is arguably impractical, difficult to enforce and harmful to its citizens' psychological health as well as to the national economy. More rationally, one is interested in finding alternative disease policies which, while respecting the critical care occupation constraint, minimize the number of days of lockdown. Alternatively, one may wish to minimize the total number of deaths during the interval $[t_0, t_f]$, or the number of infected people. In general, the figure of merit, or quantity that we wish to minimize will be a complicated functional f of the considered policy $\boldsymbol{\alpha}$ and $\boldsymbol{x}$. We will call this functional the *objective function*.

III. MODELS AND OPTIMIZATION

The optimization problem sketched above is mathematically ill-defined, unless we specify how $\boldsymbol{x}$ varies with the parameters $\boldsymbol{\mu}$ determining the policy. In order to predict the natural course of a disease or how a given policy might affect its spread, epidemiologists make use of mathematical models. In this paper, we will mainly be concerned with *deterministic compartmental models*. In these models, the whole population is divided into a number of basic compartments and the interactions between those compartments are modeled through a system of ordinary differential equations. Given the occupation $\boldsymbol{x_0}$ of the compartments at time t_0, these models allow us to compute the value of $\boldsymbol{x}$ at any instant $t \in [t_0, t_f]$ as a function of the policy $\boldsymbol{\mu}$. That is, each model provides an implicit functional relation of the form $\boldsymbol{x} = \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x_0})$.

Past literature on disease control has made extensive use of compartmental models to recommend specific strategies for policy-makers in an effort to control the spread of disease. In many cases, suitable policies are devised through a mixture of intuition and grid search, see [14, 19–21]. The starting point is a family of disease control policies with one or two unknowns. For instance, in pulse vaccination [20], those unknowns are the time intervals between two random mass vaccinations and the vaccination rate. In this paper, we propose a scheme that allows for the optimization over policies specified by thousands of parameters in the space of a few hours. Our scheme, detailed in Appendices B and C, is based on a standard tool in optimization theory and machine learning known as *gradient descent* [22, 23]. Starting with a rough guess for the optimal policy $\boldsymbol{\mu}^{(0)}$, gradient descent methods generate a sequence of policies $\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, \ldots$ which typically exhibit increasingly better performance. Although the gradient method is not guaranteed to converge to the optimal policy, after many iterations it generates solutions that are good enough for many practical problems. In fact, gradient descent is the method most commonly used to train deep neural networks [13] and support vector machines [24].

Importantly, the computational resources required to carry out the gradient descent method are comparable to the cost of running a full simulation between times t_0 and t_f with the considered disease model. Furthermore, the necessary computations can be *parallelized* for policies depending on many parameters. Although the focus of this paper is on compartmental disease models, our main ideas can also be used to understand ecological systems undergoing a more complex dynamics [25], see Appendix E. Even in such complicated scenarios, the former scaling laws hold: provided that we can run the considered disease model, we can apply the gradient method to optimize over policies of disease control.

To illustrate our approach for policy design, we will use a variant of the extended "susceptible-exposed-infected-recovered" (SEIR) disease model proposed in [14] to predict the impact of COVID-19 in the USA, a schematic of which is shown in Fig. (1) and the details of which appear in Appendix A. The clinical parameters of the model, such as recovery and hospitalization rates, were estimated in [26, 27], based on early reports from COVID-19 cases in the UK, China and Italy. Following the model in [14], the disease transmission is assumed to be seasonal by analogy with the known behavior of betacoronaviruses such as HCoV-OC43 [27], with a baseline reproduction number between 2 and 2.5, following fits of the early growth-rate of the epidemic in Wuhan [10, 11].

A relevant compartment in this model is $\mathcal{C}(t)$, the population occupying a critical care bed[1] at time t. The patients sent to critical care cannot breathe unassisted, and thus it is fundamental to ensure that such capacity is not surpassed, namely, that the constraint (2) holds. For our simulations, we chose a population size of $N_{pop} = 47$ million and $\mathcal{B}_c = 9.5 \times N_{pop} \times 10^{-5}$. That is, we assumed that the healthcare system provides 95 critical care beds

[1] More precisely, in our model, the relevant compartment is $C_C(t)$, or the *fraction* of the population in critical care at instant t. This has to be compared with B_c, the number of critical care beds per inhabitant. $\mathcal{C}(t), \mathcal{B}_c$ relate to $C_C(t), B_c$ through the expressions $\mathcal{C}(t) = N_{pop} C_C(t), \mathcal{B}_c = N_{pop} B_c$, where N_{pop} is the population size.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

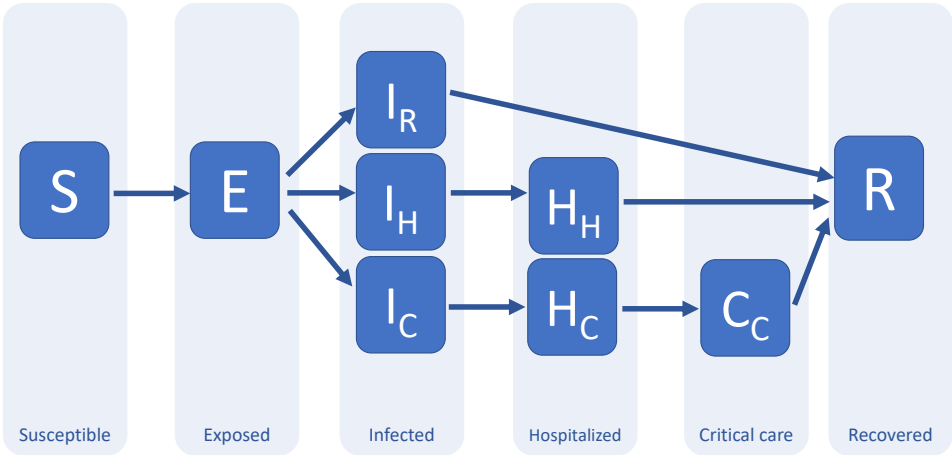

FIG. 1: **A possible compartment model for COVID-19 (adapted from [14]).** The main compartments are: susceptible; exposed; infected; hospitalized; critical and recovered. This compartmental splitting captures different possible evolutions as well as time delays between transitions. The "infected" compartment, for instance, contains those who will recover without hospitalization (I_R); those who will be hospitalized but won't need critical care (I_H); and those who will end up receiving critical care (I_C). The "exposed" compartment is introduced here to model the time delay between the exposure to the disease and the development of symptoms (incubation period), in particular, the possibility of infecting others, which is what is relevant for the model. In this model, the compartment "recovered" includes both dead and alive individuals; in principle, it could be sub-divided further.

.

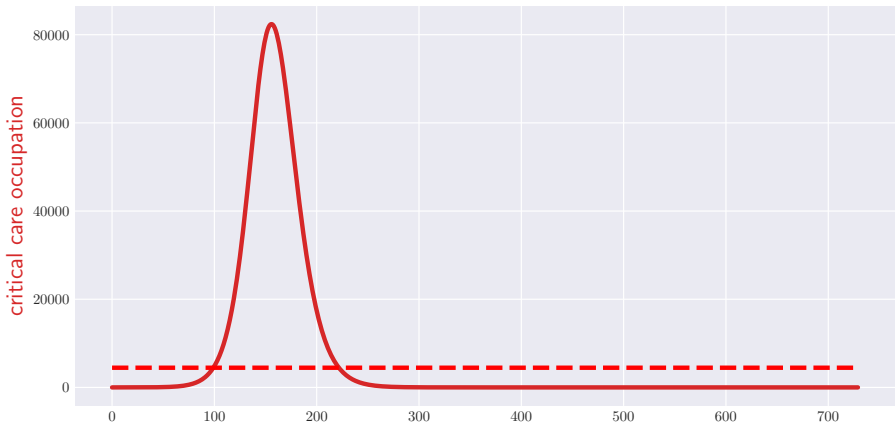

FIG. 2: **Occupation of critical care beds over two years with no policy intervention.** The red dashed line indicates the critical care capacity of the healthcare system.

per million inhabitants. This is a good approximation to the healthcare capacity of many European countries, as well as the USA.

According to the chosen model, without intervention the number of citizens requiring a bed in a critical care unit evolves according to Figure 2 (see Appendix A for the exact initial conditions of our numerical simulations). As the reader can appreciate, between the third and seventh month, the number of people in need of critical care exceeds the capacity of the considered healthcare system by 18 times.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

At the time of this writing, none of the potential vaccines for COVID-19 have passed the necessary clinical trials to be considered safe to administer [28]. In lieu of this, most governments have opted to control the disease via distancing measures and/or lockdowns. The effect of implementing a policy $s(t)$ is to multiply the disease's basic reproduction number R_0 by a factor of r [14, 27], i.e. $R_0 \to rR_0$, where

$$r := (\bar{r} - 1)s + 1. \tag{3}$$

At this point we shall make a distinction: a policy s which outputs a binary value $\{0, 1\}$ shall be known as *discrete* and correspond to situations in which a *lockdown* is either on or off. For discrete $s(t)$ as in Eq. (1), the effect of a lockdown ($s = 1$) results in the reduction of the transmission rate $r = \bar{r}$ in Eq. (3). On the other hand, when the population is free to interact ($s = 0$) then $r = 1$ and there is no change in the disease's basic reproduction number.

On the other hand, a policy s, which outputs a value on the interval $[0, 1]$ will be known as *continuous* and correspond to *physical-distancing measures*. Continuous values of $s(t)$ correspond to intermediate policies (for example mandatory face masks, suspension of sport events, remote working, school closures), the effect of which can be tentatively estimated from available data [26].

If distancing measures are the only type of intervention that a government uses, then a non-adaptive continuous policy is fully determined by the function $s(t; \boldsymbol{\mu}) \in [0, 1]$. To begin with, we will assume that the government can only declare new measures at the beginning of each week, i.e., that $s(t)$ does not vary within the weekly intervals $t \in [7k, 7(k + 1)] =: I_k$, for $k \in \mathbb{Z}$. With these conditions, $s(t)$ can be expressed as

$$s(t) = \sum_k \sigma(\tilde{s}_k)\chi_{I_k}(t). \tag{4}$$

where $\chi_{I_k}(t)$ is the characteristic function of week number k, i.e., $\chi_{I_k}(t)$ equals 1 if t is in the k-th week; and 0, otherwise. The characteristic function, thus, ensures that there is only one policy $s(t)$ per week. $\sigma(y) = \frac{1}{1+\exp(y)}$ denotes the sigmoid function, which guarantees that $s(t) \in [0, 1]$ by continuously mapping the variables $\{\tilde{s}_k\}_k$, such that it is everywhere differentiable, making it amenable to the gradient method. The parameters to optimize over are $\mu \equiv \{\tilde{s}_k\}_k$, since they fully define the government's disease policy.

Having chosen either discrete or continuous measures, one must choose a figure of merit or objective function to optimize over. Our formalism, explained in detail in Appendix C, allows us to minimize expressions of the form

$$A(\boldsymbol{\mu}, \boldsymbol{x_0}) = \int dt\, \mathcal{L}\left(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x_0})\right). \tag{5}$$

In this paper, we consider optimizations that never allow the number of people in the critical care compartment to exceed the maximal occupancy, i.e., we impose the constraint in Eq. (2). Thus we consider optimizations of the form

$$\begin{aligned} \min_{\boldsymbol{\mu}} \quad & \int dt\, \mathcal{L}\left(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x_0})\right) \\ \text{subject to} \quad & \mathcal{C}(t) \leq \mathcal{B}_c\,. \end{aligned} \tag{6}$$

In what follows, we split our analysis by choosing different objective functions $\mathcal{L}\left(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x_0})\right)$ to minimize over, considering policies $s(t; \boldsymbol{\mu})$ which are either continuous, as in Eq. (4), or discrete as in Eq. (1).

A. Objective function 1: minimizing physical distancing and lockdown measures

Suppose, for instance, that one is confident that a vaccine for COVID-19 will be developed within the next two years. In this case, one is interested in minimizing the aggregate economic cost $\mathcal{E}$ associated with the physical-distancing measures implemented by the government over these two years. Thus for the optimization in Eq. (6), we consider Lagrangians (objective functions) of the form

$$\mathcal{L}\left(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x_0})\right) = \mathcal{E}(s(t; \boldsymbol{\mu}))\,. \tag{7}$$

Figure 3 shows the result of applying the gradient method to minimize this functional for the cost function $\mathcal{E}(s) = s$, under continuous policies $s(t)$ of the form (4) satisfying the critical care capacity constraint (2). The plot shows both the critical care occupancy $\mathcal{C}(t)$ and the physical-distancing measure $s(t)$ between times t_0 and $t_f = t_0 + 2 \times 365$. The aggregate cost of the optimal policy is equal to the economic cost of sustaining a full lockdown for 294 days. As anticipated, the optimal policy found by the computer is very complicated. Notice that the critical care occupancy

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

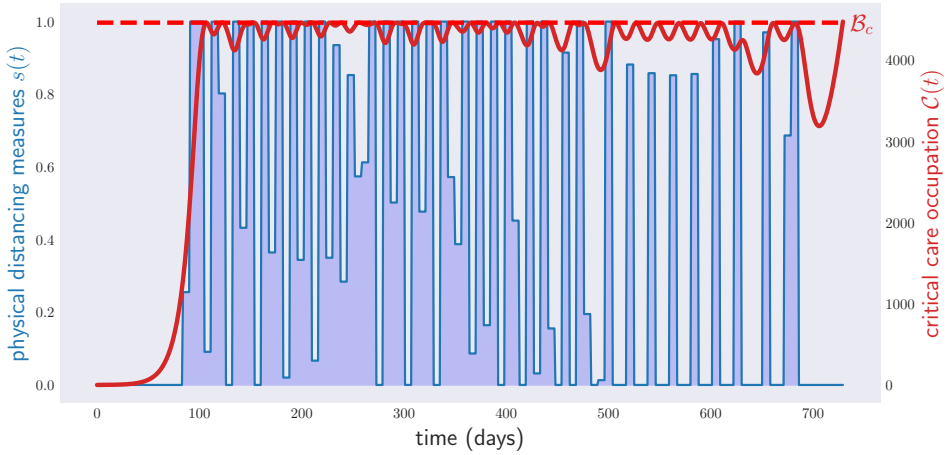

FIG. 3: **Occupation of critical care beds (red) and physical-distancing measures (blue) for a period of two years.** The optimization has been performed over continuous weekly policies, 104 continuous parameters, i.e., any value of physical-distancing measure s between 0 and 1 is accepted. The algorithm, however, tends to prefer 0/1 configurations, i.e., full lockdown or no lockdown in most instances.

grows quickly towards the end of the plot. The reason for this is that we asked the computer to minimize the total time in lockdown over a fixed period of two years, which is exactly what it did: it minimized the physical-distancing measures in the first two years with complete disregard for what could happen next. To overcome such a pathological case, one can introduce additional terms into Eq. (7) that would constrain the final slope of the curve $\mathcal{C}(t)$, to make it less steep, or even decreasing.

In some circumstances, the only distancing measures considered by governments are discrete: lockdown on or off as in Eq. (1). These kinds of optimizations over discrete variables cannot be carried out directly with the gradient method. In Appendix D, we propose two different ways to tackle this problem.

One way consists of using stochastic gradient descent to obtain an optimal non-deterministic weekly policy that is later "binned" to arrive at a deterministic policy for disease control, see Appendix D. The results of this first method are shown in Figure 4. This time the critical care occupancy curve touches the critical care capacity just after the end of year 1. The reason for this is that we demanded lockdowns to last exactly one week: had we allowed the government to declare a lockdown on any day of the week, the computer would have found a tighter solution, with every peak of the red curve touching the dashed line. Even under this discrete weekly simplification, the solution found by the computer is non-trivial: it requires the government to declare a lockdown 27 *times*. The total length of the lockdown in the course of two years is 371 days.

Our second discretization method consists in parametrizing the disease control policy through the specific times where lockdown is declared. In other words, $\boldsymbol{\mu} = (t_1, t_2, ..., t_{2N})$, and lockdown is assumed to take place within the time intervals $[t_1, t_2], [t_3, t_4], ...$ In this parametrization, lockdowns can be declared or lifted at arbitrary times within $[t_0, t_f]$, and not only on Mondays, like in the first method. This second discretization method has the advantage of allowing one to set the maximum number N of lockdowns throughout the period $[t_0, t_f]$. For $N = 9$, the corresponding critical care occupation and lockdown graphs are shown in Figure 5. The total length of the lockdown is 338 days.

B. Objective function 2: achieving herd immunity

One problem with the policies above is that they require one to implement lockdown measures over long periods of time. Another strategy to fight COVID-19 consists of steering the population towards herd immunity, without violating the constraint on the critical care capacity. This is captured by using a functional of the form

$$\mathcal{L}\left(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x_0})\right) = |S(t) - S_h|, \tag{8}$$

in Eq. (6), where S (as depicted in Fig. 1) is the component of $\boldsymbol{x}$ that denotes the proportion of individuals in the population who are susceptible to the disease. $S_h = \frac{1}{R_0}$ is the proportion of susceptible individuals required for

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

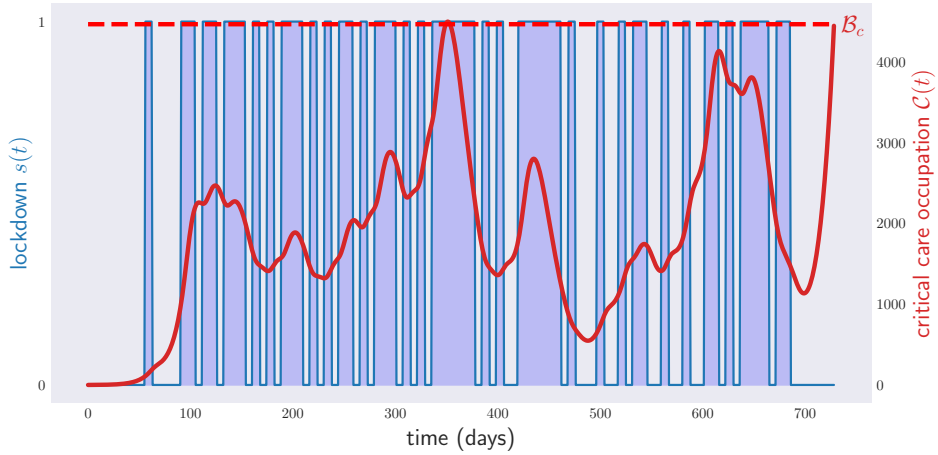

FIG. 4: **Occupation of critical care beds (red) and lockdown (blue) for a period of two years.** The plot shows the result of the optimization over probabilistic policies via gradient descent over a period of two years.

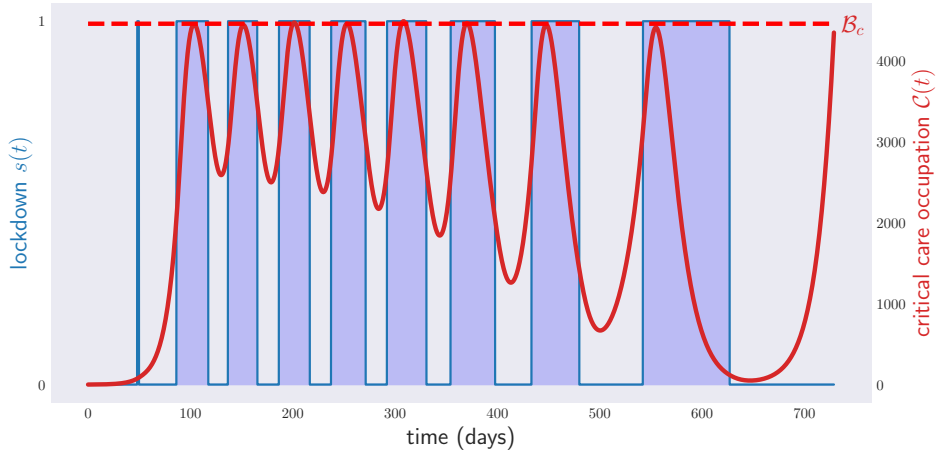

FIG. 5: **Occupation of critical care beds (red) and lockdown measures (blue) for a period of two years.** Optimization over deterministic policies with arbitrary initial and final times for each lockdown period. A total of 9 lockdown periods has been fixed prior to the optimization. Notice that the first lockdown period, around day 60, has been basically removed by the optimization procedure.

herd immunity to be guaranteed; thus, once $S(t) < S_h$ although people will continue to become infected, the natural evolution of the disease will be such that the rate and number of infected quickly dies out.

Fig. 6 illustrates the results of optimizing the objective function in Eq. (8) for continuous physical-distancing measures. The number of susceptible individuals reaches S_h on day 864, after which the disease can be considered extinct.

Notwithstanding the unprecedented velocity in the development of a COVID-19 vaccine, with many pharmaceutical companies and research institutions that already reached advanced stages in clinical trials on humans, as of now, no vaccine has completed all testing phases before being approved [28]. One may wonder what will happen if the situation were different with a possible vaccine not available in a time span of months.

A possible policy to reach faster herd immunity, while not saturating the critical care capacity, could be to deliberately infect a fraction of the population at the same time as imposing physical distance measures. It is important to remark that, even allowing this procedure on a voluntary basis among the population which are at the lowest health risk (e.g., young adults in good health), this practice poses high ethical concerns; as such, we remark that *we do*

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

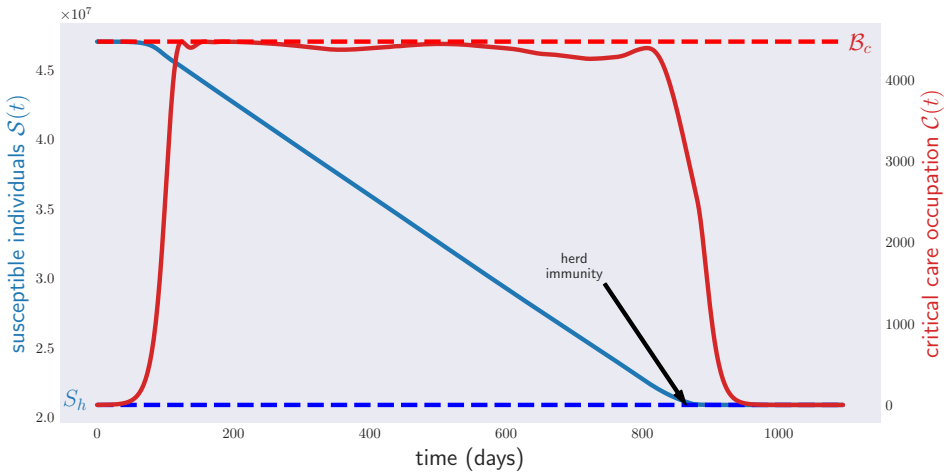

FIG. 6: **Occupation of critical care beds (red) and population of susceptible individuals (blue) for a period of three years.** The blue line corresponds to the level of susceptibles guaranteeing herd immunity over the whole year.

not endorse applying such "infection policies" on a human population. We just consider them in this paper for the academic purpose of exploring the power of the gradient method to devise complex policies for disease control. In fact, the results of our simulations indicate that the advantage with respect to simple physical distancing policies are minimal (cf. Fig. 7).

That being said, consider the following model of such an *unethical policy*. Let $\lambda(t)$ be the rate at which the government intentionally infects the general population on day t. Then, a lockdown-infection policy is identified with the functions $s(t)$ and $\lambda(t)$. As with s, one can parametrize the function λ to change once per week. Denoting by Λ the maximum rate of this deliberate infection, a suitable parametrization for $\lambda(t)$ is

$$\lambda(t) = \Lambda \sum_k \sigma(\lambda_k)\chi_{I_k}(t). \tag{9}$$

The disease model used to predict the impact of this policy will depend on how the government decides to deal with those infected: (a) the infected are quarantined until they overcome the disease; (b) the infected are allowed to mingle with the general population. In either compartmental model, one can compute the functions $\boldsymbol{x}(t;\boldsymbol{\mu},\boldsymbol{x_0})$, with $\boldsymbol{\mu} = \{\tilde{s}_k\}_k \cup \{\lambda_k\}_k$. We are interested in minimizing the functional given by (8) for the problem in Eq. (6).

The results for both quarantined/non-quarantined policies for $\Lambda = \frac{10^4}{N_{pop}}$ are shown in Figure 7, where we included the curves of the no-infection policy for comparison. As can be seen, the unethical policy of intentional infection brings the population to herd immunity, and thus the disease eventually to extinction, just slightly earlier than the optimal policy based on physical-distancing measures only.

Note how tight the constraint on critical care bed occupation becomes for policy (b). Figure 8 shows the strategy achieving this result. As in the previous cases, the optimal policy is extremely complicated in terms of physical distancing measures. The infection policy is, however, very simple: except at the beginning of the outbreak and after the disease's extinction, infection occurs at the maximum rate. Notice also that, once the disease is extinct, the computer recommends a permanent lockdown, the reason being that the objective function in Eq. (8), contrary to Eq. (7), does not penalize the overzealous use of physical distancing measures.

IV. MODELS VS. REALITY

In practice, the predictions of any mathematical model for a physical system will not be perfect for a number of reasons. First, basic parameters of the model, such as the transmission rate or the initial occupation $\boldsymbol{x_0}$, are only known up to approximations. Even if reality were exactly described by a particular mathematical model, small errors in such parameters would accumulate in the long run, making long-term predictions unreliable. Second, reality is never exactly described by mathematical models: on the contrary, any tractable disease model is, at best, a rough approximation to reality. Consequently, even the most successful disease models in the market cease to deliver solid predictions beyond 4 weeks [15].

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

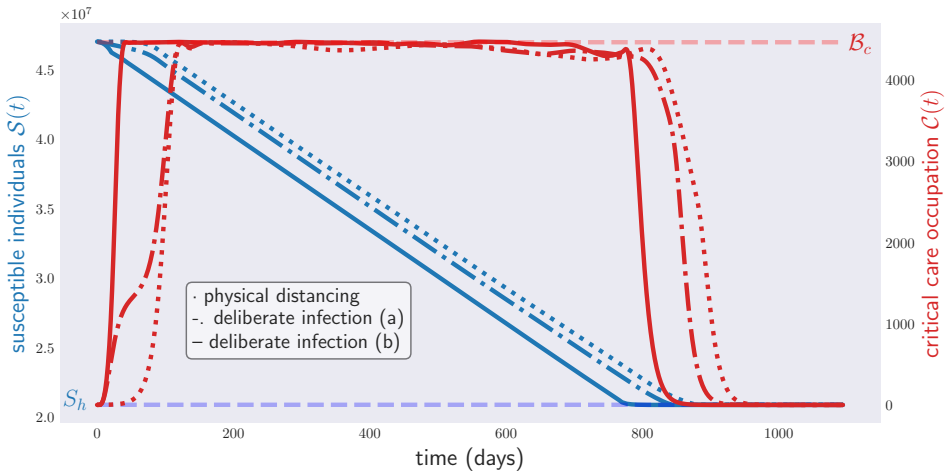

FIG. 7: **Occupation of critical care beds (red) and population of susceptible individuals (blue) for a period of three years under different policies for disease control.** The dotted lines represent the effects a physical distancing measures-only policy (dotted lines), infection policy (a) (dashed-dotted lines) and infection policy (b) (solid line).

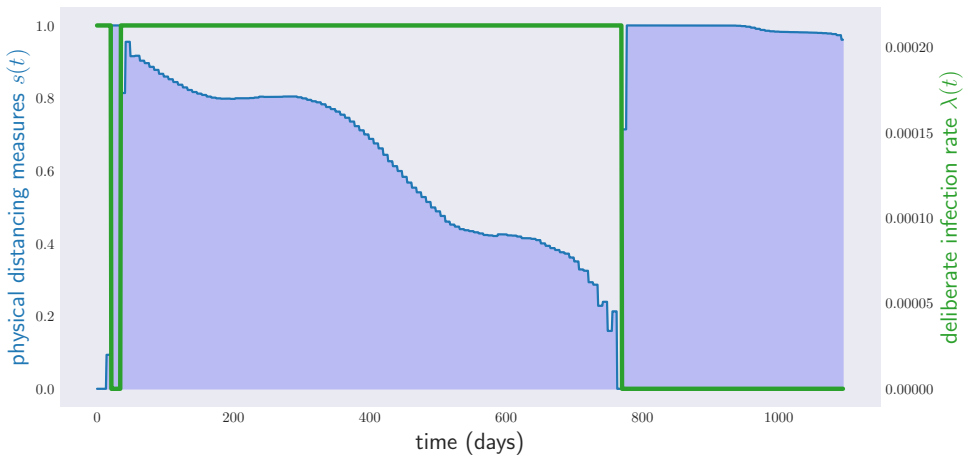

FIG. 8: **Government infection (green) and physical distancing measures (blue) under infection policy (b).** The algorithm recommend an almost constant infection rate, except for a brief initial period, which goes to zero after reaching herd immunity. Notice that, since we are minimizing the distance from the herd immunity threshold (cf. Eq (8)), the algorithm does not care about minimizing physical distancing measures. This is the reason why after reaching herd immunity, a full lockdown is suggested. This type of behavior of the algorithm can be easily corrected by hand after the solution is obtained.

These considerations make us question how practical a two-year disease control policy really is. Consider the policy depicted in Figure 3, which was obtained by applying the gradient method to the SEIR model in [14]. Here, the model parameters $\boldsymbol{\nu}$ correspond to the average values $\bar{\boldsymbol{\nu}}$ of the parameter ranges in Table A of Appendix A. In reality of course, the values of the parameters are never all equal to their averages, so we proceed to generate sets of parameters with some fluctuations. Let $\Delta\boldsymbol{\nu}$ be the vector with entries given by the difference between the upper and lower bounds of all the entries of the table, and suppose that the actual parameters of "reality" are unknown and uniformly distributed in the region of values $\mathcal{N}_a = \{\nu : -a\frac{\Delta\boldsymbol{\nu}}{2} \leq \boldsymbol{\nu} - \bar{\boldsymbol{\nu}} \leq a\frac{\Delta\boldsymbol{\nu}}{2}\}$, where a can be interpreted as the amount of noise or uncertainty. How robust is a policy to uncertainty in the initial parameters?

Fig. 9 shows the result of generating 1000 independent parameter samples from the region $\mathcal{N}_{0.05}$, corresponding to a 5% uncertainty, with respect to the given interval of values, and running the corresponding models for the optimal physical distancing policy in Fig. 3. As one can see, for some sampled values of parameters, the critical care capacity

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

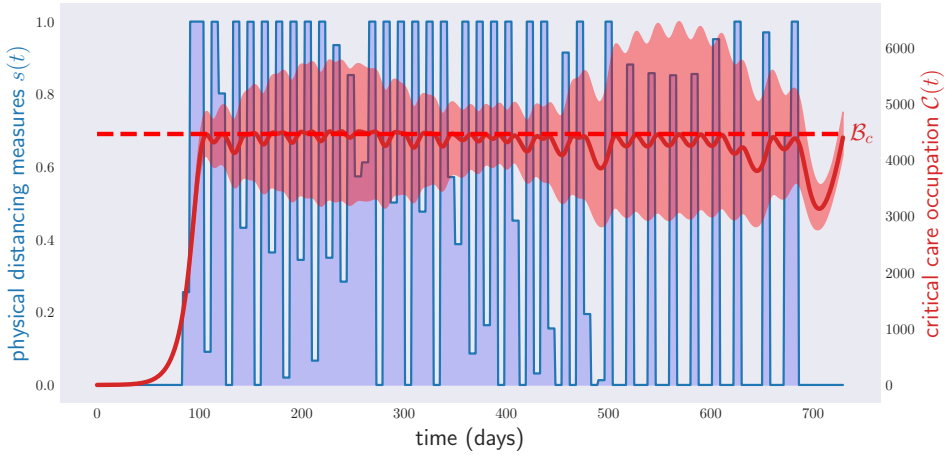

FIG. 9: **Occupation of critical care beds (red) and (unoptimized) physical distancing measures (blue) for a period of two years under random parameters in $\mathcal{N}_{0.05}$.** The region in red is obtained by sampling 1000 times from the region of model parameters $\mathcal{N}_{0.05}$ and evolving the corresponding models with the physical distancing policy optimized over the model with average-value parameters (as in Fig. 3). More precisely, the red region is the one delimited by the minimum and the maximum critical care occupation for all the 1000 models, at each time. The red line represents the average critical care occupation in all those simulations.

of the healthcare system is exceeded. This is not surprising, since the policy depicted in Fig. 3 was devised to perform well under the assumption that $\boldsymbol{\nu} = \bar{\boldsymbol{\nu}}$ and not $\boldsymbol{\nu} \in \mathcal{N}_{0.05}$, i.e., for the model parameters corresponding to their average value.

In order to tame the behavior in the plot in Fig. 9, one can employ a variant of the gradient method called *stochastic gradient descent* (see Appendix B). This allows one to optimize over long-term policies without violating the constraints of the problem for a range of values for the parameters of the model as well as the initial conditions $\boldsymbol{x_0}$[2]. Fig. 10 shows a lockdown policy minimizing the physical distancing measures under the condition that constraint (2) holds for different values of $\nu \in \mathcal{N}_{0.05}$, i.e., the critical care capacity is not exceeded. This time, the violation of condition (2) is neither so extreme nor so frequent. This comes, however, at the cost of enforcing physical distancing measures with a cost equivalent to 331 days of lockdown. Repeating the optimization for $\mathcal{N}_{0.25}$, Fig. 11, we see that the critical care capacity is rarely surpassed. However, this time the total cost is equivalent to 414 days of lockdown.

This result is what one would have expected. As time goes by, the predictions of the disease model for different values of $\boldsymbol{\nu}$ diverge: any policy which aims to satisfy constraint (2) for large ranges of these parameters will necessarily require extensive physical distancing measures.

In practical policy-making, graphs such as Figs. 10 and 11 should not be understood to represent the actual physical distancing policy, but rather to provide a provisional *policy plan*. A policy plan gives a recommendation for action for the immediate future, given the current knowledge of the disease. In Fig. 10, the policy plan is advising not to declare physical distancing measures in the first weeks. That is the measure that the government should adopt then. After a first time period, say four weeks, more data will have been gathered: this will allow us to obtain a better estimate of the parameters $\boldsymbol{\nu}$, and then re-run the models for another two years ahead. The measure to enforce should then be whatever the new policy plan recommends for the following four-week time period. The process is then repeated.

To test how this idea would perform in practice, we consider a scenario where the parameters defining the disease model are unknown, but the region in parameter space in which they live shrinks every month (to be precise, we used a 28-day period, corresponding to four weeks). That is, at month k, the government is informed that the parameters $\boldsymbol{\nu}$ satisfy $\nu \in \mathcal{N}_{0.25/\sqrt{k}}$. Every four weeks, the policy is recalculated to minimize the physical distancing measures for the rest of the two years ahead, using the range $\boldsymbol{\nu} \in \mathcal{N}_{0.25/\sqrt{k}}$. The final curves for the critical care occupation and

[2] Rigorously speaking, stochastic gradient descent just minimizes the average value of the objective function A. By adding to A sufficiently strong penalties to the violation of each optimization constraint, we make sure that such constraints will hold with high probability when $\boldsymbol{x_0}$ and $\boldsymbol{\nu}$ are sampled from the training distribution.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

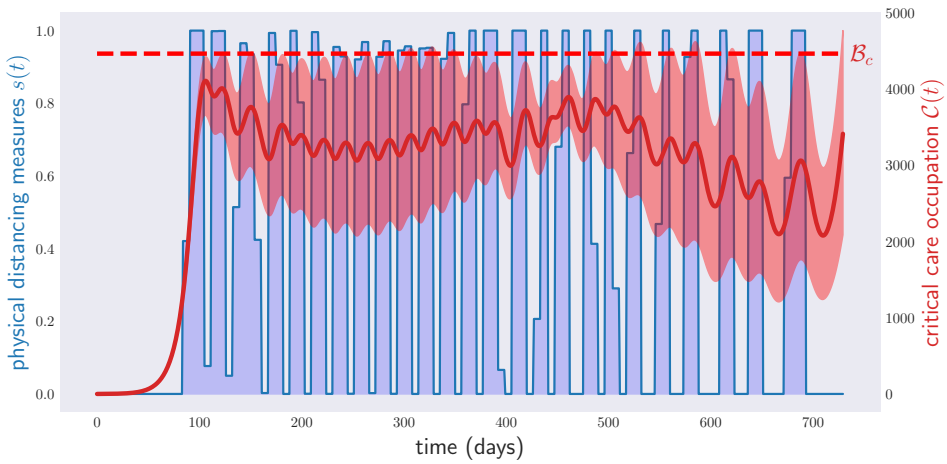

FIG. 10: **Occupation of critical care beds (red) and (optimized) physical distancing measures (blue) for a period of two years under random parameters in $\mathcal{N}_{0.05}$.** The disease control policy was optimized to respect condition (2) over the whole range of parameters $\mathcal{N}_{0.05}$. As in Fig. 9, the region in red depicts again the range of critical care occupations observed in a sample of 1000 model parameters in $\mathcal{N}_{0.05}$, and the red line the average critical care occupation.

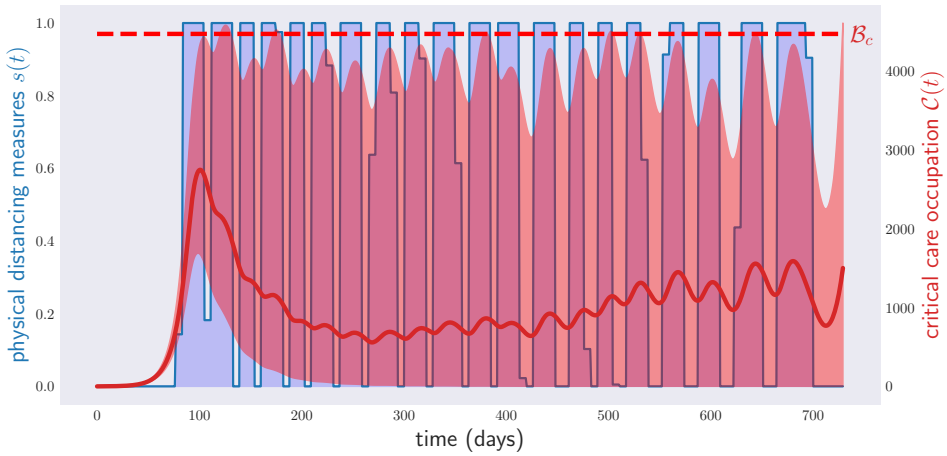

FIG. 11: **Occupation of critical care beds (red) and (optimized) physical distancing measures (blue) for a period of two years under random parameters in $\mathcal{N}_{0.25}$.** The disease control policy was optimized to respect condition (2) over the whole range of parameters $\mathcal{N}_{0.05}$. As in Fig. 9, the region in red depicts again the range of critical care occupations observed in a sample of 1000 model parameters in $\mathcal{N}_{0.25}$, and the red line the average critical care occupation.

the physical distancing measures are shown in Fig. 12 for $\boldsymbol{\nu}$, in a sequence of shrinking regions $\mathcal{N}_{0.25/\sqrt{k}}$, for each month k (red region) and in the case of a fixed uncertainty region corresponding to the last month, i.e., $\mathcal{N}_{0.049}$ (inner dark blue region). The total cost of the physical distancing measures is equivalent to 358 lockdown days. This has to be compared with the cost of 414 days predicted by the initial policy plan under the assumption $\nu \in \mathcal{N}_{0.25}$.

In principle, one could further decrease the total planned cost by devising *adaptive policy plans*, where the measure to be taken at each moment depends not only on the current time t, but also on the past history of physical distancing measures and their observed effects. In fact, we tried optimizing over adaptive policies described by a continuous version of a neural network architecture known as Long Short-Term Memory (LSTM) [29]. In all our numerical experiments, such simple LSTM architectures could not improve the performance of non-adaptive strategies, but this could be due to ineffective training on our side.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

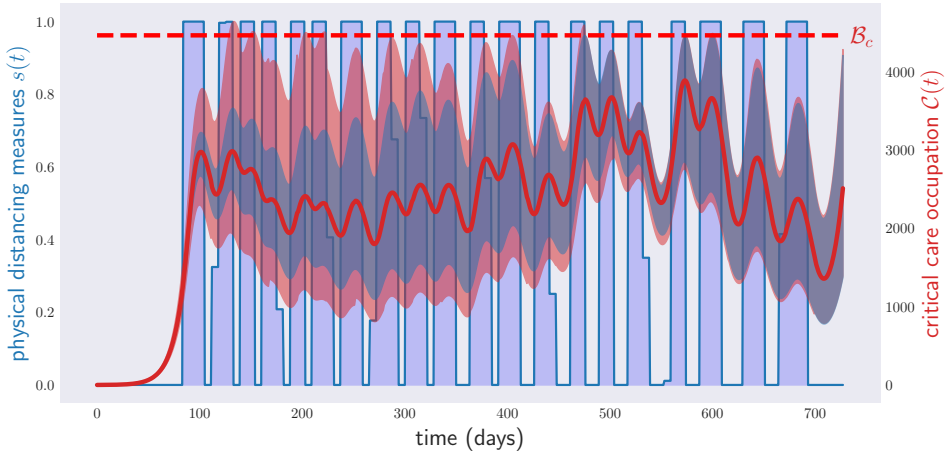

FIG. 12: **Occupation of critical care beds (red) and physical distancing measures (blue) for a period of two years under random parameters and monthly noise decrease.** The disease control policy was optimized to respect the condition in Eq. (2), i.e., critical care capacity not exceeded, starting with the parameter region $\mathcal{N}_{0.25}$ and with a monthly noise decrease of $\frac{1}{\sqrt{k}}$, i.e., in the k-th month the noise is equal to $0.25/\sqrt{k}$. The region in red depicts the range of critical care occupations observed in a sample of 1000 model parameters in a sequence of shrinking regions $\mathcal{N}_{0.25/\sqrt{k}}$; more precisely, for each month k the red region is obtained by evolving, from the initial time to month k (included), 1000 different models with parameters sampled from the region $\mathcal{N}_{0.25/\sqrt{k}}$. The final plot is obtained by joining the plots for each month k. The inner region in dark blue depicts the range of critical care occupations corresponding to the uncertainty in the final month, i.e., obtained with 1000 models with parameters sampled in $\mathcal{N}_{0.049}$. The red line represents the average critical care occupation, obtained by joining the average of the simulations with decreased uncertainty for each month k. Despite the initial uncertainty on the parameters ν of the disease model, the final lockdown time is much lower, due to monthly revisions of the original policy plan.

V. CONCLUSION

In this paper, we have applied standard tools from optimization theory and machine learning to identify optimal disease control policies, given an epidemiological model. This is in stark contrast to standard practice in mathematical epidemiology, where human intuition is used to narrow down the considered set of policies to a uni-parametric family. We saw that the optimal solutions found by our algorithms are highly counter-intuitive, and thus unlikely to be identified by a human. This supports the idea that policies for disease control should be based on a combination of both human expertise and machine learning.

Compared to previous approaches that tried to identify suitable disease control policies through optimal control theory, our framework allows one to devise policies that satisfy arbitrary constraints under arbitrary uncertainties in the initial conditions and model parameters that determine the disease's dynamics. Our methods, in addition, allow one to optimize over discrete policies, as well as policies that can just vary at certain fixed times.

To illustrate our ideas, we studied a scenario in which a computer is tasked with outputting the minimal amount of physical distancing measures for an epidemic in a hypothetical country, in such a way as to never exceed the critical care bed capacity. We looked at situations in which these measures were continuous (recommendations on the interval $[0, 1]$) as well as discrete (either 0 or 1) – a lockdown that is off or on, respectively for periods of 2 years. We experimented with measures which are just allowed to change weekly, as well as those in which there is a maximum number of lockdowns that is allowed to be declared. When considering unethical interventions, in which parts of the population are deliberately infected, we found that there is little difference in the time for the disease to become extinct, compared with the natural evolution of the disease.

We examined the problems that one may encounter when applying these techniques to scenarios where the model parameters are not known with high accuracy, which led us to propose practical policy plans which must be continually revisited, to account for our ever-changing and ever-growing knowledge in an epidemic. We tested the viability of this approach by simulating a scenario where the uncertainty on the disease model parameters decreases with time. As expected, the final policy implemented was safe for the final range of parameters and required considerably less physical distancing measures than the initial policy plan hinted.

In this last regard, an interesting problem for future research is how to devise adaptive policy plans for disease

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

control, where the actual measure at each time depends on the whole history of disease indicators accessible to the government. In theory, such plans should predict lower values of the average objective function in scenarios where the model parameters are unknown. In our experience, though, gradient descent alone seems to be unable to beat the non-adaptive score.

Finally, we would like to remark once more that, since the optimization problems we dealt with in this paper are non-convex, the gradient method is not guaranteed to converge to the minimum of the (average) objective function. While conducting this research, in order to convince ourselves that the solutions found by our numerical methods were close to optimal, we had to repeat our optimizations several times, with different initial policies $\boldsymbol{\mu}^{(0)}$ and learning rates ϵ. Such a redundant use of computational resources would have been entirely avoidable if we had had some rough approximation to the exact solution of the problem. Hence we conclude this paper with a challenge for the operations research community: develop mathematical tools which allow one to *lower bound* the solution of minimization problems involving ordinary differential equations.

Acknowledgments

We thank Luca Gerardo-Giorda and Mario Budroni for useful discussions. C.B. and Y.G. acknowledge funding from the Austrian Science Fund (FWF) through the Zukunftskolleg ZK03.

---

[1] U. Nations, *A un framework for the immediate socio-economic response to covid-19* (2020), URL https://unsdg.un.org/resources/un-framework-immediate-socio-economic-response-covid-19.

[2] Medecins sans Frontieres (2020), URL https://www.msf.org/covid-19-has-made-yemen-health-system-collapse-complete.

[3] Medecins sans Frontieres (2020), URL https://www.msf.org/health-system-close-collapse-amid-covid-19-el-salvador.

[4] T. R. Pradhan, The Kathmandu Post (2020), URL https://kathmandupost.com/national/2020/07/21/government-decides-to-lift-the-four-month-long-coronavirus-lockdown-but-with-conditions.

[5] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, 2008), ISBN 9780691116174, URL http://www.jstor.org/stable/j.ctvcm4gk0.

[6] H. Gaff and E. Schaefer, Mathematical Biosciences and Engineering **6**, 469 (2009), ISSN 1551-0018, URL http://aimsciences.org//article/id/c4a66425-a177-45d3-89b6-6b6caa97965d.

[7] M. d. R. de Pinho, I. Kornienko, and H. Maurer, in *CONTROLO'2014 – Proceedings of the 11th Portuguese Conference on Automatic Control*, edited by A. P. Moreira, A. Matos, and G. Veiga (Springer International Publishing, Cham, 2015), pp. 135–145.

[8] L. L. Obsu and S. F. Balcha, Journal of Biological Dynamics **14**, 590 (2020), pMID: 32696723, https://doi.org/10.1080/17513758.2020.1788182, URL https://doi.org/10.1080/17513758.2020.1788182.

[9] M. Zamir, Z. Shah, F. Nadeem, A. Memood, H. Alrabaiah, and P. Kumam, Computer Methods and Programs in Biomedicine **196**, 105642 (2020), ISSN 0169-2607, URL http://www.sciencedirect.com/science/article/pii/S0169260720314759.

[10] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al., New England Journal of Medicine **382**, 1199 (2020), pMID: 31995857, URL https://doi.org/10.1056/NEJMoa2001316.

[11] J. Riou and C. L. Althaus, Eurosurveillance **25**, 2000058 (2020).

[12] G. Volpicelli, Wired (2020), URL https://www.wired.co.uk/article/china-coronavirus-local-lockdowns.

[13] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, USA, 2016), URL http://www.deeplearningbook.org.

[14] S. Kissler, C. Tedijanto, M. Lipsitch, and Y. Grad, *Social distancing strategies for curbing the COVID-19 epidemic* (2020), URL https://dash.harvard.edu/handle/1/42638988.

[15] N. G. Reich, L. C. Brooks, S. J. Fox, S. Kandula, C. J. McGowan, E. Moore, D. Osthus, E. L. Ray, A. Tushar, T. K. Yamana, et al., Proceedings of the National Academy of Sciences **116**, 3146 (2019), ISSN 0027-8424, URL https://www.pnas.org/content/116/8/3146.

[16] *Online repository with the code used for the simulations, optimizations, and plots presented*, URL https://github.com/costantinobudroni/opt-disease-control/.

[17] M. J. Keeling and P. Rohani, Ecology Letters **5**, 20 (2002), URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1461-0248.2002.00268.x.

[18] *World Health Organization, Press briefing, March 20th 2020*, URL https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-press-conference-full-20mar2020.pdf?sfvrsn=1eafbff_0.

[19] A. R. McLean and S. M. Blower, Proceedings of the Royal Society of London. Series B: Biological Sciences **253**, 9 (1993), URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.1993.0075.

[20] Z. Agur, L. Cojocaru, G. Mazor, R. M. Anderson, and Y. L. Danon, Proceedings of the National Academy of Sciences **90**, 11698 (1993), ISSN 0027-8424, URL https://www.pnas.org/content/90/24/11698.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[21] M. J. Keeling, M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell, Science **294**, 813 (2001), ISSN 0036-8075, URL https://science.sciencemag.org/content/294/5543/813.

[22] S. Boyd, L. Xiao, and A. Mutapcic, lecture notes of EE392o, Stanford University, Autumn Quarter (2004), URL https://web.stanford.edu/class/ee392o/.

[23] J. Duchi, `https://web.stanford.edu/class/ee364b/lectures.html` (2018), URL https://web.stanford.edu/class/ee364b/lectures.html.

[24] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (Cambridge University Press, 2000).

[25] J. P. Keller, L. Gerardo-Giorda, and A. Veneziani, Journal of biological dynamics **7 Suppl 1**, 31—46 (2013), ISSN 1751-3758, URL https://europepmc.org/articles/PMC3957468.

[26] F. et al., Imperial College (2020), URL https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/.

[27] S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, and M. Lipsitch, Science **368**, 860 (2020), URL https://science.sciencemag.org/content/368/6493/860/tab-pdf.

[28] *Coronavirus Vaccine Tracker*, URL https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html.

[29] S. Hochreiter and J. Schmidhuber, Neural computation **9**, 1735 (1997).

[30] D. P. Kingma and J. Ba, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun (2015), URL http://arxiv.org/abs/1412.6980.

[31] *Front Matter* (John Wiley & Sons, Ltd, 2016), ISBN 9781119121534, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119121534.fmatter.

[32] S. A. Cook, in *Proceedings of the Third Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 1971), STOC '71, p. 151–158, ISBN 9781450374644, URL https://doi.org/10.1145/800157.805047.

[33] D. L. Logan, *A First Course in the Finite Element Method Using Algor* (Brooks/Cole Publishing Co., USA, 2000), 2nd ed., ISBN 0534380689.

APPENDIX A: MODELS FOR THE SPREAD OF COVID-19

In all our numerical simulations, we will assume that the dynamics of the COVID-19 are well approximated by a compartmental model of the SEIR type. When the government policy reduces to enforcing physical distance measures, we will adopt a simplified version of the model used in [14]. This model divides those infected with COVID-19 into three different compartments: I_R or those who recover by themselves from the disease; I_H, those who require hospitalization but do not enter a critical care unit; and I_C, those who are both hospitalized and visit a critical care unit before recovery. The dynamics of the model are governed by the system of ordinary differential equations below, see the diagram in Figure 13:

$$\begin{aligned}
\frac{dS}{dt} &= -r(t)\beta(t)S(I_R + I_H + I_C),\\
\frac{dE}{dt} &= r(t)\beta(t)S(I_R + I_H + I_C) - \nu E,\\
\frac{dI_R}{dt} &= p_R\nu E - \gamma I_R,\\
\frac{dI_H}{dt} &= p_H\nu E - \gamma I_H,\\
\frac{dI_C}{dt} &= p_C\nu E - \gamma I_C,\\
\frac{dH_H}{dt} &= \gamma I_H - \delta_H H_H,\\
\frac{dH_C}{dt} &= \gamma I_C - \delta_C H_C,\\
\frac{dC}{dt} &= \delta_C H_C - \xi C,\\
\frac{dR}{dt} &= \gamma I_C + \delta_H H_H + \xi C.
\end{aligned} \tag{A1}$$

Here

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

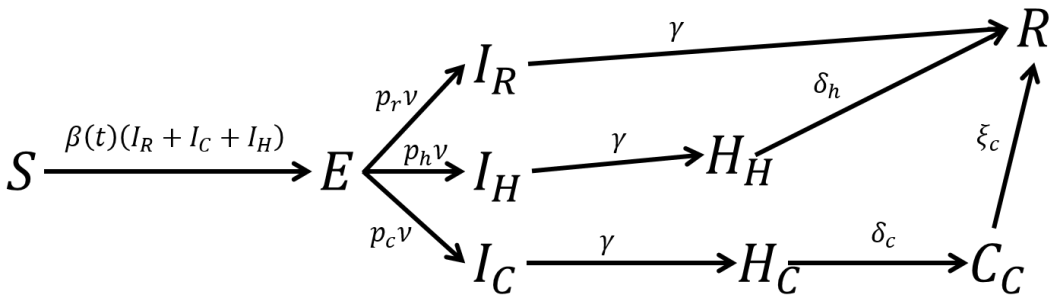

FIG. 13: **COVID-19 model for disease policies based on physical distance measures.**

parameter	value	units
$\bar{r}$	$[0, 0.6]$	none
Δ	$[0.7, 1.0]$	none
γ	$1/5$	days^{-1}
ν	$1/4.6$	days^{-1}
p_R	0.9596	none
p_H	0.0308	none
p_C	0.0132	none
δ_H	$1/8$	days^{-1}
δ_C	$1/6$	days^{-1}
ξ	$1/10$	days^{-1}
ϕ	-7×3.8	days
R_0	[2, 2.5]	None

TABLE I: Parameter ranges for the compartmental model for COVID-19 proposed in [14].

$$\beta(t) = \gamma R_0 \left(\frac{1+\Delta}{2} + \frac{1-\Delta}{2} \cos\left(\frac{2\pi(t+\phi)}{7 \times 52} \right) \right) \tag{A2}$$

denotes the virus' transmitivity, that is subject to seasonal variability. $r(t) = (\bar{r} - 1)s(t) + 1$ models the effect of a government-mandated lockdown $s(t) \in \{0, 1\}$ on the virus' transmission rate. The values of the remaining parameters are taken from [14], and appear in Table A.

If the government has the option of deliberately infecting random samples of the population with COVID-19, then the model gets slightly more complicated. If those infected are quarantined until they overcome the disease, the new model, depicted in Figure 14, is described by the system of equations:

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

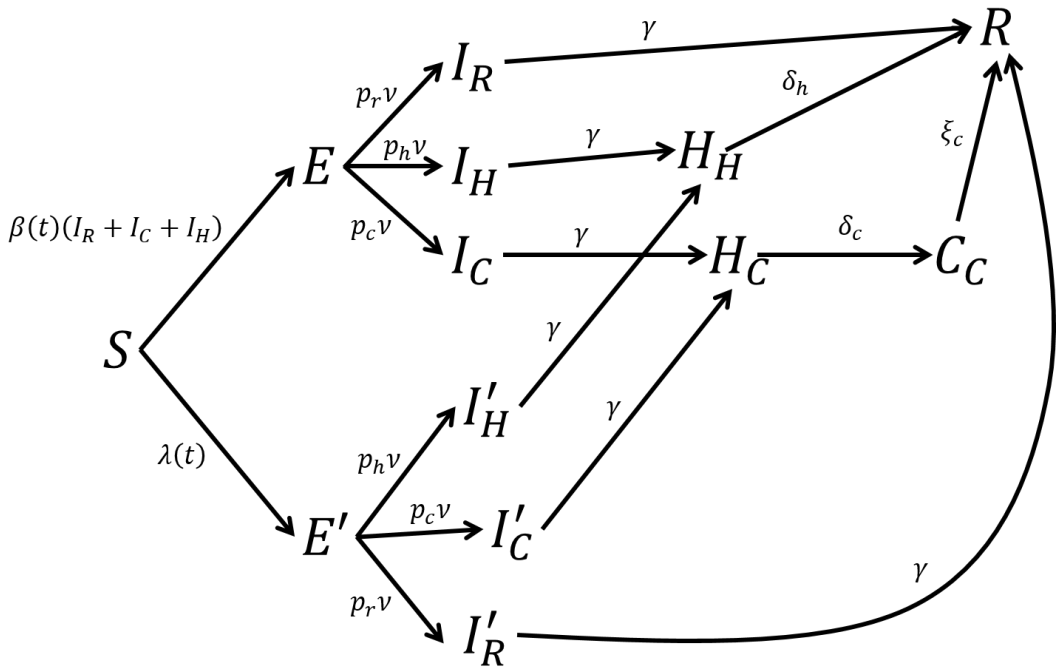

FIG. 14: **COVID-19 model for disease policies based on physical distance measures and deliberate inoculation.**

$$
\begin{aligned}
\frac{dS}{dt} &= -r\beta(t)S(I_R + I_H + I_C) - \lambda(t)S, \\
\frac{dE}{dt} &= r\beta(t)S(I_R + I_H + I_C) - \nu E, \\
\frac{dI_R}{dt} &= p_R\nu E - \gamma I_R, \\
\frac{dI_H}{dt} &= p_H\nu E - \gamma I_H, \\
\frac{dI_C}{dt} &= p_C\nu E - \gamma I_C, \\
\frac{dE'}{dt} &= \lambda(t)S - \nu E', \\
\frac{dI'_R}{dt} &= p_R\nu E' - \gamma I'_R, \\
\frac{dI'_H}{dt} &= p_H\nu E' - \gamma I'_H, \\
\frac{dI'_C}{dt} &= p_C\nu E' - \gamma I'_C, \\
\frac{dH_H}{dt} &= \gamma(I_H + I'_H) - \delta_H H_H, \\
\frac{dH_C}{dt} &= \gamma(I_C + I'_C) - \delta_C H_C, \\
\frac{dC}{dt} &= \delta_C H_C - \xi C, \\
\frac{dR}{dt} &= \gamma I_C + \delta_H H_H + \xi C.
\end{aligned}
\tag{A3}
$$

If, on the contrary, the government allows those deliberately infected to mingle with the general population, the equations above simplify considerably: it suffices to replace all instances of the term $r\beta(t)S(I_R + I_H + I_C)$ in eqs. (A1) by $r\beta(t)S(I_R + I_H + I_C) + \lambda(t)S$.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

In all our numerical simulations, we take the total population to be 47 million; the critical care bed capacity per inhabitant C_c is also taken to be 9.5×10^{-5}. We assume that the government starts its intervention on day $t_0 = 60$, 30 days after the outbreak of the disease. We model the disease outbreak by assuming that, at time $t_{out} = 30$, there are 10 individuals in compartment E. By default, the values of the disease parameters are taken to be the arithmetic means of the intervals shown in Table A. We assume that the government can deliberately infect 10^4 individuals per day. This sets a value for Λ of $\frac{10^{-2}}{47}$.

APPENDIX B: THE GRADIENT METHOD

Given functions $f, g_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, ..., K$, consider the following optimization problem:

$$\min_{\boldsymbol{\mu}} f(\boldsymbol{\mu}) \quad \text{such that } g_i(\boldsymbol{\mu}) \leq h_i, \text{ for } i = 1, \ldots, K. \tag{B1}$$

Each of the conditions $g_i(\boldsymbol{\mu}) \leq h_i$ is called a *constraint*.

Define $M = \{\boldsymbol{\mu} : g_i(\boldsymbol{\mu}) \leq h_i, i = 1, ..., K\}$. If f is sub-differentiable, a simple heuristic to solve this problem is the projected gradient method [22]. Call $\boldsymbol{\mu}^\star$ the solution of the problem. Starting from an initial guess $\boldsymbol{\mu}^{(0)}$, the gradient method generates a sequence of values $(\boldsymbol{\mu}^{(k)})_k$ with the property that $\lim_{k\to\infty} \boldsymbol{\mu}^{(k)} = \boldsymbol{\mu}^\star$, provided that M, f are, respectively, a convex set and a convex function [22]. The sequence $(\boldsymbol{\mu}^{(k)})_k$ is generated recursively via the relation

$$\boldsymbol{\mu}^{(k)} = \pi_M \left(\boldsymbol{\mu}^{(k-1)} - \epsilon \nabla_{\boldsymbol{\mu}} f(\boldsymbol{\mu}^{(k-1)}) \right), \tag{B2}$$

where, for any set $A \in \mathbb{R}^n$, $\pi_A(\boldsymbol{z})$ denotes the point $\boldsymbol{y} \in A$ that minimizes the Euclidean distance, i.e., $\min_{\boldsymbol{y}\in A} \|\boldsymbol{y} - \boldsymbol{x}\|_2$.

Unfortunately, in the problems we encounter in the main text, f is not convex and, sometimes, neither is M. This implies that the sequence output by the projected gradient method is not guaranteed to converge to the solution of the problem, but to a local minimum thereof.

In machine learning, optimization problems with non-convex objective function f and $M = \mathbb{R}^n$ are legion. To solve them, deep learning practitioners typically use variants of the gradient method sketched above. One of these variants, Adam [30], is extensively used to train neural networks.

Adam works as follows. Starting with the null vectors $\boldsymbol{m}^{(0)}, \boldsymbol{v}^{(0)} \in \mathbb{R}^n$, vector sequences are generated according to the following iteration rule:

$$\begin{aligned}
\boldsymbol{G}^{(k)} &= \nabla_{\boldsymbol{\mu}} f(\boldsymbol{\mu}^{(k-1)}) \\
\boldsymbol{m}^{(k)} &= b_1 \boldsymbol{m}^{(k-1)} + (1 - b_1) \boldsymbol{G}^{(k)} \\
\boldsymbol{v}^{(k)} &= b_2 \boldsymbol{v}^{(k-1)} + (1 - b_2) \boldsymbol{G}^{(k)} \odot \boldsymbol{G}^{(k)} \\
\hat{\boldsymbol{m}}^{(k)} &= \frac{\boldsymbol{m}^{(k)}}{1 - b_1^k} \\
\hat{\boldsymbol{v}}^{(k)} &= \frac{\boldsymbol{v}^{(k)}}{1 - b_2^k} \\
\boldsymbol{\mu}^{(k)} &= \boldsymbol{\mu}^{(k-1)} - \epsilon \frac{\hat{\boldsymbol{m}}^{(k)}}{\sqrt{\hat{\boldsymbol{v}}^{(k)}} + \delta},
\end{aligned} \tag{B3}$$

where $G^{(k)} \odot G^{(k)}$ denotes the vector of the element-wise product; similarly, the fraction and square root in the definition of $\boldsymbol{\mu}^{(k)}$ are defined element-wise. Recommended values for the free parameters $\epsilon, b_1, b_2, \delta$ are $\epsilon = 0.001$, $b_1 = 0.9$, $b_2 = 0.999$ and $\delta = 10^{-8}$ [30].

Like the projected gradient method, Adam is not guaranteed to converge to the optimal solution of the problem. However, provided that the initial conditions and the learning rate ϵ are chosen with care, Adam has been observed to typically output a local minimum that is "good enough".

Note that, by taking $M = \mathbb{R}^n$, it is not clear how to enforce that the solution satisfies constraints of the form $g_i(\boldsymbol{\mu}) \leq h_i$. The answer is to include those constraints as penalties in the objective function. That is, rather than minimizing f, we apply Adam to minimize the function

$$f(\boldsymbol{\mu}) + \sum_{i=1}^{K} \nu_i (h_i - g_i(\boldsymbol{\mu}))^+, \tag{B4}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

where $\nu_i \gg 1$ and z^+ denotes the positive part of z, i.e., $z^+ = z$ for $z > 0$; otherwise, $z^+ = 0$. For high enough values of $\{\nu_i\}_i$, the solution of the problem will just violate the constraints slightly, i.e., $h_i - g(\boldsymbol{\mu}^\star) \in [-\delta, \infty)$, for $\delta \ll 1$. If no violation whatsoever is desired then one can instead optimize over a function of the form

$$f(\boldsymbol{\mu}) + \sum_{i=1}^{K} \nu_i (h_i - \delta' - g_i(\boldsymbol{\mu}))^+, \tag{B5}$$

with $\delta' > 0$.

In some situations, the objective function f will be complicated to the point that computing its exact gradient is an intractable problem. It might be possible, though, to generate a random vector $\tilde{\nabla}_{\boldsymbol{\mu}} f \in \mathbb{R}^n$ with the property

$$\left\langle \tilde{\nabla}_{\boldsymbol{\mu}} f \right\rangle = \nabla_{\boldsymbol{\mu}} f. \tag{B6}$$

In such a predicament, we can solve the original optimization problem (B1) through stochastic gradient descent methods [23]. Stochastic gradient descent consists in applying the considered gradient method, with the difference that, every time that the method requires the gradient of f, we input the random variable $\tilde{\nabla}_{\boldsymbol{\mu}} f \in \mathbb{R}^n$ instead. Namely, it suffices to replace $\nabla_{\boldsymbol{\mu}} f(\boldsymbol{\mu}^{(k-1)})$ by $\tilde{\nabla}_{\boldsymbol{\mu}} f(\boldsymbol{\mu}^{(k-1)})$ in the iterative equations (B2), (B3). As before, if both M and f are convex, stochastic gradient descent methods are guaranteed to converge to the optimal solution of problem (B1) [23].

APPENDIX C: OPTIMIZATION OVER CONTINUOUS POLICIES FOR DISEASE CONTROL

Our starting point is an ordinary differential equation of the form

$$\frac{dx^i}{dt} = G^i(t, \boldsymbol{x}; \boldsymbol{\mu}); i = 1, ..., m. \tag{C1}$$

The entries of vector x represent the occupations of the different compartments of a disease model[3]. $\boldsymbol{\mu} \in \mathbb{R}^n$ represents a parametrization of the effects of a given policy. In the examples of the main text, $\boldsymbol{\mu} = \{\tilde{s}_k\}_k$ or $\boldsymbol{\mu} = \{\tilde{s}_k\}_k \cup \{\lambda_k\}_k$. Call $\boldsymbol{x}(t; \bar{\boldsymbol{\mu}}, \boldsymbol{x}_0)$ the solution of Eq. (C1) with initial conditions $\boldsymbol{x}(0) = \boldsymbol{x}_0$ and $\boldsymbol{\mu} = \bar{\boldsymbol{\mu}}$.

Given the set $M \subset \mathbb{R}^n$, we consider the problem of finding the parameters $\boldsymbol{\mu}^\star \in M$ such that $\boldsymbol{x}(t; \boldsymbol{\mu}^\star, \boldsymbol{x}_0)$ minimizes a given functional A. This functional defines how we wish to control the disease and what for: it might represent the number and duration of lockdown, etc. For the time being, let us assume this functional to be of the form

$$A(\boldsymbol{\mu}, \boldsymbol{x}_0) = \int_{t_0}^{t_f} dt \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0)). \tag{C2}$$

Note that we are assuming to know the initial conditions $\boldsymbol{x}_0$ with precision. We will relax this requirement by the end of the section.

From the discussion in section B, functional (C2) might also contain constraints which we wish the solution $\boldsymbol{x}(t; \boldsymbol{\mu}^\star, \boldsymbol{x}_0)$ to satisfy. For instance, if we want Eq. (2) to hold, then one of the terms in $\mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0))$ could be

$$\rho(t) \left(C(t; \boldsymbol{\mu}, \boldsymbol{x}_0) - B_c\right)^+, \tag{C3}$$

with $\rho(t) \gg 1$. In the optimizations presented in the paper, we did just this, with $\rho(t) = \frac{500}{B_c}$ (figures 12, 11), $\rho(t) = \frac{100}{B_c}$ (figure 10) or $\rho(t) = \frac{10}{B_c}$ (rest of the figures).

As we saw in section B, optimizations over a large number of parameters $\boldsymbol{\mu} \in \mathbb{R}^n$ are usually conducted via gradient descent methods, such as Adam [30]. Such methods require us to compute $\nabla_{\boldsymbol{\mu}} A$. For functionals of the form (C2), we have that

$$\frac{\partial A}{\partial \mu_j} = \int_{t_0}^{t_f} dt \left(\frac{\partial \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0))}{\partial \mu_j} + \sum_i \frac{\partial \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0))}{\partial x^i} \frac{\partial x^i}{\partial \mu_j} \right). \tag{C4}$$

[3] In the case of adaptive policies, some of such entries might also represent the components of the *cell state* $\boldsymbol{\theta}$ [29], i.e., the internal variables used by the government to keep track of the evolution of the disease and guide future government interventions.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The next question is thus how to compute the derivatives $\frac{\partial x^i}{\partial \mu_j}$. To this aim, define the variables $y^i_j(t;\boldsymbol{\mu},\boldsymbol{x}_0) \equiv \frac{\partial x^i(t;\boldsymbol{\mu},\boldsymbol{x}_0)}{\partial \mu_j}$. Differentiating equation (C1) by μ_j, we have that

$$\frac{\partial y^i_j}{\partial t} = \frac{\partial G^i(t,\boldsymbol{x};\boldsymbol{\mu})}{\partial \mu_j} + \sum_l \frac{\partial G^i(t,\boldsymbol{x};\boldsymbol{\mu})}{\partial x^l} y^l_j; i = 1,...,m, j = 1,...,n. \tag{C5}$$

In order to obtain $\{y^i_j(t)\}$ for each time t, it hence suffices to solve the system of coupled differential equations given by (C1), (C5) with initial conditions $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$, $y^i_j(t_0) = 0$. This can be achieved numerically through several different methods, depending on the desired accuracy. In all our numerical simulations, we used the Euler explicit method [31]. Namely, for $\delta > 0$, we regarded time as a discrete variable of the form $t_k = t_0 + \delta k$, for $k = 0,...,\lceil \frac{t_f - t_0}{\delta} \rceil$. We obtain the quantities $\{x^i(t_k), y^i_j(t_k) : k\}$ by recursively applying the relations

$$\begin{aligned} x^i(t_{k+1}) &= x^i(t_k) + \delta G^i(t_k, \boldsymbol{x}(t_k);\boldsymbol{\mu}), \\ y^i_j(t_{k+1}) &= y^i_j(t_k) + \delta \left(\frac{\partial G^i(t_k,\boldsymbol{x}(t_k);\boldsymbol{\mu})}{\partial \mu_j} + \sum_l \frac{\partial G^i(t_k,\boldsymbol{x}(t_k);\boldsymbol{\mu})}{\partial x^l} y^l_j(t_k) \right). \end{aligned} \tag{C6}$$

We chose $\delta = 1.0$ to generate all our plots, with the exception of Figure 5, where $\delta = 0.1$.

In some situations in the main text, our functional A is more complicated than (C2). Some parameters $\boldsymbol{\zeta}$ (not policy parameters) regulating the evolution (C1), such as the disease's basic reproduction number, might be unknown, or perhaps the initial conditions $\boldsymbol{x}_0$ are just known within some bounds. In such cases, the problem's objective function A might adopt the form

$$A(\boldsymbol{\mu}) = \int p(\boldsymbol{\zeta},\boldsymbol{x}_0) d\boldsymbol{\zeta} d\boldsymbol{x}_0 \int dt \mathcal{L}(t,\boldsymbol{\mu},\boldsymbol{x}(t;\boldsymbol{\mu},\boldsymbol{\zeta},\boldsymbol{x}_0)). \tag{C7}$$

Again, we wish to minimize A over $\lambda \in M$. As explained in section B, this can be achieved via stochastic gradient descent methods [22]: all we need is an unbiased estimator $\tilde{\nabla}_{\boldsymbol{\mu}} A$ of $\nabla_{\boldsymbol{\mu}} A$. We obtain this estimator by taking N independent samples $(\boldsymbol{\zeta}^{(j)}, \boldsymbol{x}_0^{(j)})_{j=1}^N$ from the measure $p(\boldsymbol{\zeta},\boldsymbol{x}_0) d\boldsymbol{\zeta} d\boldsymbol{x}_0$ and using them to compute the quantity

$$\tilde{\nabla}_{\boldsymbol{\mu}} A = \frac{1}{N} \sum_{j=1}^N \nabla_{\boldsymbol{\mu}} \int dt \mathcal{L}(t,\boldsymbol{\mu},\boldsymbol{x}(t;\boldsymbol{\mu},\boldsymbol{\zeta}^{(j)},\boldsymbol{x}_0^{(j)})). \tag{C8}$$

The optimal value of N depends on the variance of the specific estimator, and we used different values for different levels of noise, defining the set $\mathcal{N}$ in the main text.

APPENDIX D: OPTIMIZATION OVER DISCRETE POLICIES OF DISEASE CONTROL

The above section explains how to conduct optimizations over disease control policies, as long as the parameters $\boldsymbol{\mu}$ defining the policy are allowed to vary all over $\mathbb{R}^n$. Some policies, though, are by their very nature, discrete. For instance, on day t we can either declare a lockdown ($s = 1$) or not declare a lockdown ($s = 0$). The effect of the policy measure over the disease's transmission rate will be to multiply the latter by the amounts $\bar{r}$ and 1, respectively.

In principle, we could model this situation via a continuous variable $\lambda \in \mathbb{R}$ and write the effect r on the transmitivity by means of a piece-wise continuous function of λ, e.g.: $r = \Theta(-\lambda)\bar{r} + (1 - \Theta(-\lambda))1$, where $\Theta(z)$ denotes the Heaviside function [4].

In that case, however, the gradient method would not work. Note that the Heaviside function has zero derivative everywhere except at 0. In every iteration of Adam, $\nabla_\lambda A$ would be null, and so $\lambda^{(k)} = \lambda^{(0)}$ for all k.

Optimizing over discrete variables is, in general, a very difficult endeavor: apparently simple problems can be argued not to have an efficient solution [32]. In this paper, we propose a simple heuristic to attack this problem in this particular situation.

---

[4] $\Theta(z)$ equals 1 for $z \geq 0$, or 0, otherwise.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Suppose, for the time being, that our lockdown policy were probabilistic, i.e., at each week k, we declare a lockdown with probability $p_k(1) = \sigma(\tilde{s}_k)$; otherwise, with probability $p_k(0) = 1 - \sigma(\tilde{s}_k)$, we let the population roam freely. We wish to minimize our *average* objective function, that is, the expression

$$\bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}) = \sum_{c_1,\ldots,c_q=0,1} \prod_{k=1}^{q} p_k(c_k) \int dt \mathcal{L}\left(t, \boldsymbol{\mu}, \boldsymbol{x}(t; c, \boldsymbol{\mu}, \boldsymbol{x}_0)\right), \tag{D1}$$

where c is the whole vector of weekly lockdowns, and $\boldsymbol{\mu}$ corresponds to the continuous parameters of the policy, e.g.: vaccination rates.

In principle, we could apply gradient descent to minimize (D1). Estimating the exact gradient of the above expression is, however, unrealistic, as it involves summing a number of terms exponential in the number of weeks q. Instead, we will produce a random unbiased estimate of the gradient and invoke stochastic gradient descent methods, see Appendix B.

Let us first differentiate Eq. (D1) with respect to the continuous variables $\boldsymbol{\mu}$. The result is

$$\frac{\partial \bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu})}{\partial \mu_j} = \sum_{c_1,\ldots,c_q=0,1} \prod_{k=1}^{q} p_k(c_k) v_j(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c) = \langle v_j(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c)\rangle_c, \tag{D2}$$

where

$$v_j(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c) = \int dt \frac{\partial \mathcal{L}\left(t, \boldsymbol{\mu}, \boldsymbol{x}(t; c, \boldsymbol{\mu}, \boldsymbol{x}_0)\right)}{\partial \mu_j} \tag{D3}$$

and the components of the random variable $c \in \{0,1\}^q$ are generated by sequentially sampling from the Bernouilli distributions $(1 - \sigma(\tilde{s}_k), \sigma(\tilde{s}_k))_k$. Note that the expression in the integrand of (D3) can be computed using the techniques discussed in Appendix C.

Differentiating Eq. (D1) with respect to $\tilde{s}_k$ we find that

$$\frac{\partial \bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu})}{\partial \tilde{s}_k} = \sum_{a=0,1} \left\langle w_k(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c^{(k,a)})\right\rangle_{c^{(k,a)}}, \tag{D4}$$

with

$$w_k(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c^{(k,a)}) = \frac{\partial p_k(a)}{\partial \tilde{s}_k} \int dt \mathcal{L}\left(t, \boldsymbol{\mu}, \boldsymbol{x}(t; c^{(k,a)}, \boldsymbol{\mu}, \boldsymbol{x}_0)\right), \text{ for } a = 0, 1 \tag{D5}$$

and the average $\left\langle w_k(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c^{(k,a)})\right\rangle_{c^{(k,a)}}$ is obtained via sampling over the product of Bernoulli distributions for $c_1^{(k,a)}, \ldots, c_{k-1}^{(k,a)}, c_{k+1}^{(k,a)}, \ldots, c_q^{(k,a)}$ and fixing $c_k^{(k,a)} = a$.

Putting this all together, we have that the random vectors v, w satisfy

$$\begin{aligned} \langle v\rangle &= \nabla_{\boldsymbol{\mu}} \bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}) \\ \langle w\rangle &= \nabla_s \bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}). \end{aligned} \tag{D6}$$

Since both vectors can be sampled efficiently, we can use them (and their averages) to optimize over $\bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu})$ via stochastic gradient descent.

At this point, the reader might object that our original goal was to minimize (C2) over policies with *deterministic* lockdown. Very conveniently, independently of the initial values of $s, \boldsymbol{\mu}$ the stochastic gradient method will converge to a policy $p^\star, \boldsymbol{\mu}^\star$ such that the deterministic policy with the same continuous parameters $\boldsymbol{\mu}^\star$ and lockdown given by

$$c_k^\star = \begin{cases} 0 & \text{for } p_k^\star(0) > 1/2 \\ 1 & \text{otherwise} \end{cases} \tag{D7}$$

has the same objective value.

Indeed, for $k \in \{1, ..., n\}$, fix $\{s_j^\star : j \neq k\}$. Then,

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

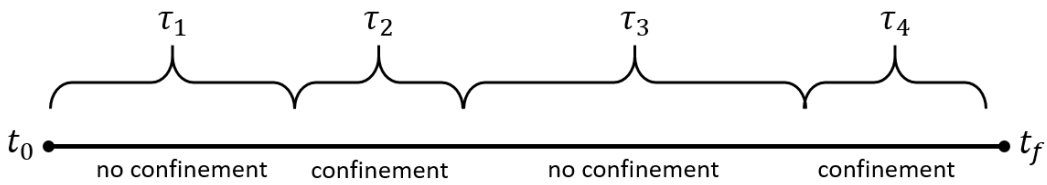

FIG. 15: **Parametrization of continuous lockdown times.**

$$\bar{A}(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star) = \sum_{a=0,1} p_k^\star(a) A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = a), \tag{D8}$$

with

$$A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = a) \equiv \sum_{c_1,\dots c_{k-1},c_{k+1},\dots,c_q=0,1} \prod_{j\neq k}^{q} p_j^\star(c_j) \int dt \mathcal{L}\left(t, \boldsymbol{\mu}^\star, c, \boldsymbol{x}(t; c_1, ..., c_{k-1}, a, c_{k+1}, ..., c_q, \boldsymbol{\mu}^\star, \boldsymbol{x}_0)\right). \tag{D9}$$

Since $p^\star, \boldsymbol{\mu}^\star$ is a local minimum of $\bar{A}(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star)$, it follows that, either $p_k^\star(0)p_k^\star(1) = 0$ or

$$A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = 0) = A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = 1) = \bar{A}(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star). \tag{D10}$$

In either case, fixing c_k through the procedure (D7) cannot increase the average value of the objective function. Iterating over $k = 1, ..., q$, we prove the claim.

Note that this method easily extends to optimizations over discrete adaptive policies. In that case, one can model the probability p_k of lockdown on week k as a function of both the current input $\tilde{x}$ and the cell state $\boldsymbol{\theta}$, i.e., $p_k(a) = p_k(a|\tilde{x}, \boldsymbol{\theta})$. Let us assume that the family of functions $p_k(a|\tilde{x}, \boldsymbol{\theta})$ available is rich enough to regard $p_k(a|\tilde{x}, \boldsymbol{\theta}), p_k(a|\tilde{x}', \boldsymbol{\theta}')$ as independent. Then one can argue as before and conclude that a deterministic policy can be inferred from the limiting probabilistic policy generated by stochastic gradient descent.

1. Optimization over discrete policies with continuous lockdown times

In this section, we explain how to optimize over lockdown policies of the following form:

$$s(t) = \begin{cases} 0 & \text{if } t - t_0 \leq \tau_1 \text{ or } \sum_{k=1}^{2i} \tau_k \leq t - t_0 \leq \sum_{k=1}^{2i+1} \tau_k, \text{ for some } i, \\ 1 & \text{otherwise.} \end{cases} \tag{D11}$$

Here t_0 is fixed and the variables $\{\tau_k\}_{k=1}^n$ are assumed to be non-negative and add up to $t_f - t_0$; this policy can hence be parametrized by a vector $\boldsymbol{\mu} \in \mathbb{R}^n$, with $\boldsymbol{\tau} = (t_f - t_0)\text{softmax}(\boldsymbol{\mu})$ [5]. Intuitively, $\{\tau_k\}_{i=1}^n$ divide the interval $[t_0, t_f]$ into n different parts. In each part, lockdown is alternatively declared ($s = 1$) or suspended ($s = 0$), see Figure 15.

At time t, the disease's basic reproduction number is given by (3), with $s(t)$ defined as above. To find out $y_j^i \equiv \frac{\partial x^i(t;\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j}$, we invoke Eq. (C5). In computing the term

$$\frac{\partial G^i}{\partial \mu_j} = \sum_k \frac{\partial G^i}{\partial \tau_k} \frac{\partial \tau_k}{\partial \mu_j}, \tag{D12}$$

[5] The function softmax$(\boldsymbol{\mu})$ of a vector $\boldsymbol{\mu}$ is a vector $\boldsymbol{\nu}$ of the same dimension with components $\nu_i = \frac{\exp(\mu_i)}{\sum_j \exp(\mu_j)}$.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

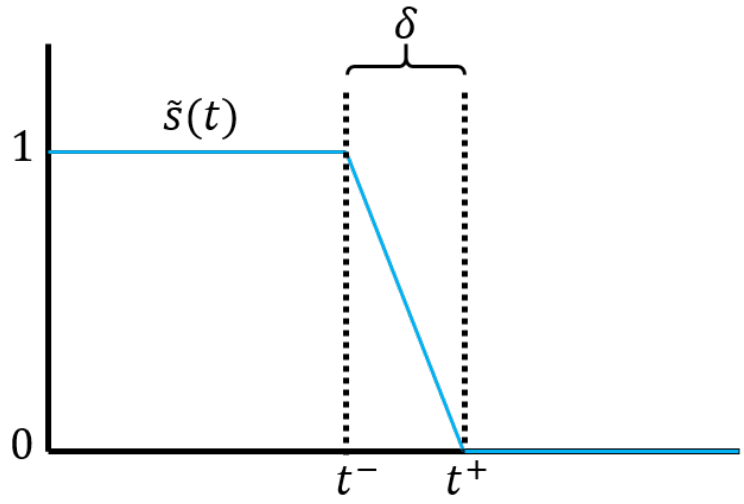

FIG. 16: **Modified continuous function** $\tilde{s}(t)$.

we have the problem that, due to (D11), G is not continuous or derivable. To work our way out, we approximate $s(t)$ by a piece-wise continuous function with bounded derivative that transitions from 0 to 1 (or viceversa) linearly and in time $\delta \ll 1$, see Figure 16; later we will take the limit $\delta \to 0$.

The new function $\tilde{s}(t)$ has zero derivative with respect to μ_i, except for t satisfying

$$t^- := t_0 + \sum_{k=1}^{u} \tau_k \leq t \leq t_0 + \sum_{k=1}^{u} \tau_k + \delta =: t^+. \tag{D13}$$

In that case, the derivative of $\hat{s}$ with respect to τ_j, with $j \leq u$, will (approximately) be

$$\frac{\partial}{\partial \tau_j} \hat{s}(t) = \frac{(-1)^u}{\delta}. \tag{D14}$$

The derivative with respect to any of the variables $\{\tau_j : j > u\}$ is zero.

The dominant term on the right-hand side of (C5) for $t \in [t^-, t^+]$ is therefore

$$\begin{aligned} \frac{\partial G^i}{\partial \mu_j} &= \frac{\partial G^i(t^-, \boldsymbol{x}(t^-))}{\partial \tilde{s}} \sum_{k=1}^{u} \frac{\partial \tilde{s}}{\partial \tau_k} \frac{\partial \tau_k}{\partial \mu_j} \\ &= (t_f - t_0) \frac{(-1)^u}{\delta} \frac{\partial G^i(t^-, \boldsymbol{x}(t^-))}{\partial \tilde{s}} \sum_{k=1}^{u} \frac{\partial \mathrm{softmax}(\boldsymbol{\mu})_k}{\partial \mu_j} =: \frac{K^i_j(t^-, \boldsymbol{x}(t^-), \boldsymbol{\mu}, u)}{\delta}. \end{aligned} \tag{D15}$$

Since the evolution takes place for time $\delta = t^+ - t^-$, we have that $y^i_j(t^+) \approx y^i_j(t^-) + K^i_j(t^-, \boldsymbol{x}(t^-), \boldsymbol{\mu}, m)$. Taking the limit $\delta \to 0$, we have that the evolution of y^i_j is determined by the following prescription:

1. $y^i_j(t_0) = 0$.
2. Let $t = t_0 + \sum_{1=0}^{u} \tau_k$, for some u. Then y^i_j is updated by the rule

$$y^i_j(t) \to y^i_j(t) + K^i_j(t, \boldsymbol{x}, \boldsymbol{\mu}, u). \tag{D16}$$

3. For all other values of t, y^i_j continuously evolves via the equation

$$\frac{dy^i_j}{dt} = \sum_l \frac{\partial G^i(t, \boldsymbol{x}; \boldsymbol{\mu})}{\partial x^l} y^l_j; i = 1, ..., m, j = 1, ..., n. \tag{D17}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.09.15.20194811; this version posted March 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

APPENDIX E: CONTINUOUS-SPACE POPULATION MODELS

Even though the focus of this article is that of compartmental models of the form (C1), one can also apply the principles of gradient descent for policy optimizations on dynamical systems governed by partial differential equations. Consider, e.g., the scenario studied in [25], where the authors model the spread of rabies in raccoons across a realistic landscape $\Omega \subset \mathbb{R}^2$ through a system of reaction-diffusion equations of the form

$$\frac{\partial}{\partial t}\boldsymbol{u} - \operatorname{div}(\nu \nabla \boldsymbol{u}) = A(\boldsymbol{u})\boldsymbol{u}. \tag{E1}$$

In this equation, the three entries of the vector field $u(t, X, Y) \in \mathbb{R}^3$ respectively denote the number of susceptible, exposed and infected individuals at time t in position X, Y. ν, A are 3×3 matrices that, in principle, might depend on some controllable parameters $\boldsymbol{\mu}$. This equation is to be solved under the initial conditions $u(0, X, Y) = u_0(X, Y)$ and the homogeneous von Neumann boundary conditions

$$\nabla \boldsymbol{u}(t, X, Y) \cdot \boldsymbol{n}(X, Y) = 0, \text{ for } (X, Y) \in \partial\Omega, \tag{E2}$$

where $\boldsymbol{n}(X, Y) \in \mathbb{R}^2$ denotes the vector normal to the contour $\partial\Omega$ at location (X, Y). The authors of [25] solve this equation numerically via the Finite Element Method (FEM) [33].

Suppose that we wished to optimize the policy parameters $\boldsymbol{\mu} \in \mathbb{R}^n$ over some functional A depending on $\boldsymbol{u}(t, X, Y; \boldsymbol{\mu}, u_0)$ (instead of $\boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0)$) via the gradient method. Then at some point we would need to compute the quantities $v_j^i(t, X, Y; \boldsymbol{\mu}, u_0) \equiv \frac{\partial u^i(t,X,Y;\boldsymbol{\mu},u_0)}{\partial \mu_j}$. Call $\boldsymbol{v_j} \in \mathbb{R}^3$ the vector with components v_j^i and differentiate both (E1) and (E2) with respect to μ_i. This results in the equation

$$\begin{aligned} &\frac{\partial}{\partial t}\boldsymbol{v_j} - \operatorname{div}\left(\frac{\partial \nu}{\partial \mu_j}\nabla \boldsymbol{u} + \nu \nabla \boldsymbol{v_j}\right) = \left(\frac{\partial A}{\partial \mu_j} + \frac{\partial A}{\partial u_i} v_j^i\right)\boldsymbol{u} + A(\boldsymbol{u})\boldsymbol{v_j} \\ &\nabla \boldsymbol{v_j}(t, X, Y) \cdot \boldsymbol{n} = 0, \text{ for } (X, Y) \in \partial\Omega. \end{aligned} \tag{E3}$$

Since $\boldsymbol{u}(0, X, Y; \boldsymbol{\mu}, \boldsymbol{u_0})$ does not depend on $\boldsymbol{\mu}$, this new diffusion equation must be solved for the initial conditions $\boldsymbol{v_j}(X, Y, 0) = 0$. This can be achieved numerically in the same way that the authors of [25] solved Eq. (E1), that is, via the FEM.