RESEARCH ARTICLE

Disease control as an optimization problem

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

1 Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Vienna, Austria, **2** Faculty of Physics, University of Vienna, Vienna, Austria

* miguel.navascues@oeaw.ac.at

Abstract

In the context of epidemiology, policies for disease control are often devised through a mixture of intuition and brute-force, whereby the set of logically conceivable policies is narrowed down to a small family described by a few parameters, following which linearization or grid search is used to identify the optimal policy within the set. This scheme runs the risk of leaving out more complex (and perhaps counter-intuitive) policies for disease control that could tackle the disease more efficiently. 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, and stochastic models. In addition, our methods allow for optimization over policies which remain constant over weekly periods, specified by either continuous or discrete (e.g.: lockdown on/off) government measures. 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).

1 Introduction

The COVID-19 pandemic has already caused over four million deaths worldwide. 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 expert intuition, a number of suitable policies to control the disease are proposed; (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.

The main advantage of this three-step process is that the final recommended policy is comprehensible, i.e., it can be interpreted and explained. Moreover, for simple on/off policies, one

can sometimes derive analytic results [6]. The process has, nonetheless, two disadvantages. First of all, the class of policies devised by an human could well be suboptimal, since the optimal policy (under some figure of merit) could be extremely complicated and counter-intuitive. Second, the method generally requires one to numerically simulate each policy, and so is inapplicable when the considered class of policies depends on many control parameters: due to the exponentially large number of conceivable policies, by the time one finds the optimal disease control policy, it would be too late to enforce it.

Other approaches for disease control rely on optimal control theory to identify a suitable policy (see, e.g., [7–10]). The starting point of all these works is that both the initial conditions (namely, the number of individuals 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 in the disease's reproduction number can be very high [11, 12]. 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. It is true that some optimal control problems admit a discrete solution, a so-called bang-bang control, but one cannot enforce this property on the solution *a priori*. This makes optimal control theory all the more impractical, for discrete measures, such as a lockdown that is either on or off, have so far dominated global efforts to control the COVID-19 epidemic [13]. Last but not least, optimal control theory can only handle scenarios where the policies vary over time continuously, thus not making it compatible with observed government measures to control COVID-19, which for the most part have been applied on a discrete, 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 [14], 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-Exposed-Infected-Recovered (SEIR) compartmental model [5] proposed in [15]. 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 [16]. 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 [17].

2 The framework in a nutshell

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 [18]. 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 x. The disease is subsequently analysed by looking at how 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 [19]) distancing measures, or even a full lockdown, are common examples of interventions aimed to fight 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 vaccination campaigns, where the intervention consists of vaccinating a number of individuals per day, across all compartments, i.e., without distinguishing between susceptible, exposed, recovered, and so on. Let $v(t)$ be the fraction of the total population 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
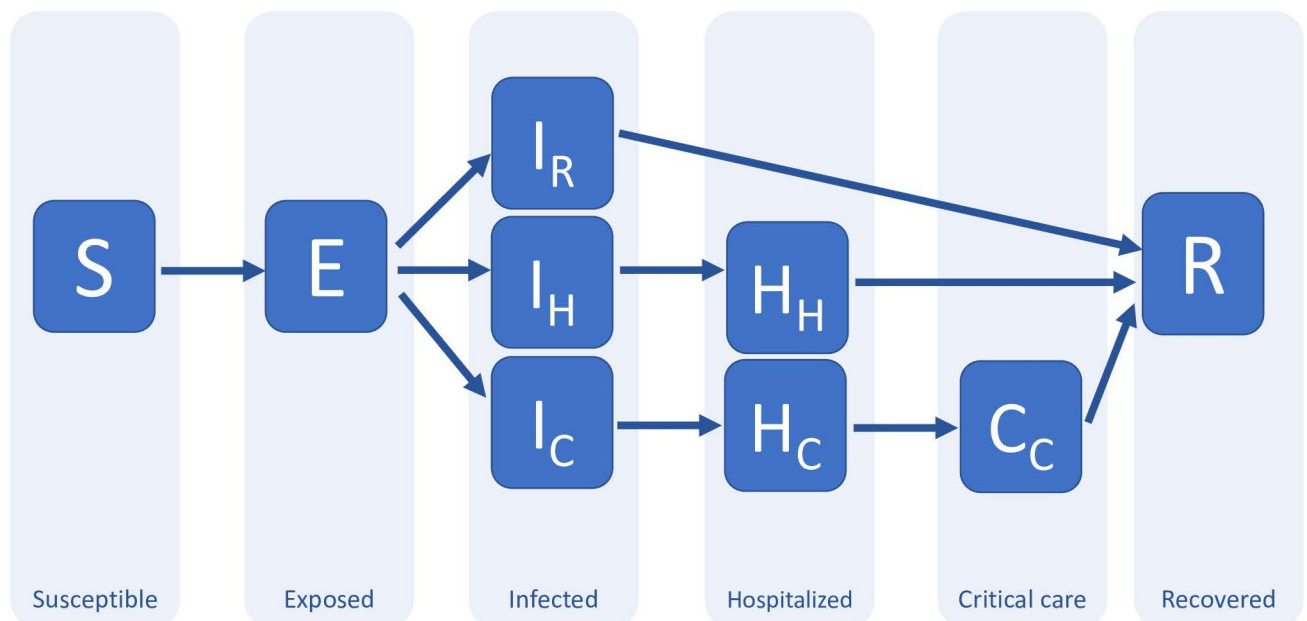

Fig 1. A possible compartment model for COVID-19 (adapted from [15]). The main compartments are: susceptible; exposed; infected; hospitalized; critical and recovered. This compartmental splitting captures different possible evolutions as well as time delays between transitions. The "infected" compartment, 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.

https://doi.org/10.1371/journal.pone.0257958.g001

lockdown policy between times t_0 and t_f corresponds to the function $s(t)$.

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

If the government is intervening both through vaccination and lockdown, then its disease control policy will be identified by *both* functions $v(t)$ and $s(t)$. Note that, whereas $v(t)$ can take a continuum of values, $s(t)$ can only have finitely many. In the following, we will call the first class of interventions *continuous*; and the second, *discrete*. In general, a policy for disease control will combine both kinds of measures, but, as we will see, optimizing over one class or the other requires very different techniques.

The above are instances of *non-adaptive policies* for disease control, because the functions s, v only depend on the time t, and not, e.g., on the current death rate. 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 associated to a time-dependent vector function $\boldsymbol{\alpha}$ 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 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

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

We will call any such condition on the policy, or on its effect on the evolution of the disease, a *constraint*. Further examples of relevant constraints are the requirement that the number of planned vaccinations does not exceed the government's total supply, or that the death rate does not surpass a given threshold.

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 of the considered policy $\boldsymbol{\alpha}$ and $\boldsymbol{x}$. We will call this functional the *objective function*. A sufficiently complex objective function, together with the appropriate optimization constraints, can meet any conceivable set of societal demands.

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, experts make use of mathematical models. In this paper, we will mainly be concerned with *deterministic compartmental models*, but we will show also how our method extends to stochastic models. In these models, the whole population is divided into a number of basic compartments and the interactions between those compartments are modeled, in the deterministic case, through a system of ordinary differential equations. Given the occupation x_0 of the compartments at time t_0, these models allow us to compute the value of x at any instant $t \in [t_0, t_f]$ as a function of the policy μ. That is, each model provides an implicit functional relation of the form

$$x = x(t; \mu, x_0). \tag{3}$$

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 a disease. In many cases, suitable policies are devised through a mixture of intuition and grid search, see [15, 20–22]. The starting point is a family of disease control policies with one or two unknowns. For instance, in pulse vaccination [21], those unknowns are the time intervals between two 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 the following sections, is based on a standard tool in optimization theory and machine learning known as *gradient descent* [23, 24]. Starting with a rough guess for the optimal policy $\mu^{(0)}$, gradient descent methods generate a sequence of policies $\mu^{(1)}$, $\mu^{(2)}$, ... 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 purposes. In fact, gradient descent is the method most commonly used to train deep neural networks [14] and support vector machines [25].

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 more complex dynamics [26], see Appendix D. 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.

3 The gradient method

We next introduce the gradient method and show how one can use it to tackle an abstract optimization problem. Given functions $f, g_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \ldots, K$, consider the following task:

$$\min_{\mu} f(\mu)$$
$$\text{such that} \quad g_i(\mu) \leq h_i, \text{ for } i = 1, \ldots, K. \tag{4}$$

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

Define $M = \{\mu: g_i(\mu) \leq h_i, i = 1, \ldots, K\}$. If f is sub-differentiable, a simple heuristic to solve this problem is the projected gradient method [23]. Call $\mu^\star$ the solution of the problem. Starting from an initial guess $\mu^{(0)}$, the gradient method generates a sequence of values $(\mu^{(k)})_k$ with the property that $\lim_{k \to \infty} \mu^{(k)} = \mu^\star$, provided that M, f are, respectively, a convex set and a convex function [23]. The sequence $(\mu^{(k)})_k$ is generated recursively via the relation

$$\mu^{(k)} = \pi_M(\mu^{(k-1)} - \epsilon \nabla_\mu f(\mu^{(k-1)})), \tag{5}$$

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 this paper, 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 [27], 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:

$$
\begin{aligned}
\boldsymbol{G}^{(k)} &= \nabla_{\boldsymbol{\mu}} f(\boldsymbol{\mu}^{(k-1)}) \\
\boldsymbol{m}^{(k)} &= b_1 \boldsymbol{m}^{(k-1)} + (1 - b_1)\boldsymbol{G}^{(k)} \\
\boldsymbol{v}^{(k)} &= b_2 \boldsymbol{v}^{(k-1)} + (1 - b_2)\boldsymbol{G}^{(k)} \odot \boldsymbol{G}^{(k)} \\
\hat{\boldsymbol{m}}^{(k)} &= \frac{\boldsymbol{m}^{(k)}}{1 - b_1^k} \\
\hat{\boldsymbol{v}}^{(k)} &= \frac{\boldsymbol{v}^{(k)}}{1 - b_2^k} \\
\boldsymbol{\mu}^{(k)} &= \boldsymbol{\mu}^{(k-1)} - \epsilon \frac{\hat{\boldsymbol{m}}^{(k)}}{\sqrt{\hat{\boldsymbol{v}}^{(k)}} + \delta},
\end{aligned}
\tag{6}
$$

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 ϵ, b_1, b_2, δ are $\epsilon = 0.001$, $b_1 = 0.9$, $b_2 = 0.999$ and $\delta = 10^{-8}$ [27].

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

$$
f(\boldsymbol{\mu}) + \sum_{i=1}^{K} v_i (g_i(\boldsymbol{\mu}) - h_i)^+,
\tag{7}
$$

where $v_i \gg 1$ and z^+ denotes the positive part of z, i.e., $z^+ = z$ for $z > 0$; otherwise, $z^+ = 0$. For high enough values of $\{v_i\}_i$, 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

$$
f(\boldsymbol{\mu}) + \sum_{i=1}^{K} v_i (g_i(\boldsymbol{\mu}) - h_i + \delta')^+,
\tag{8}
$$

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

$$\langle \tilde{\nabla}_{\boldsymbol{\mu}} f \rangle = \nabla_{\boldsymbol{\mu}} f. \tag{9}$$

In such a predicament, we can solve the original optimization problem (4) through stochastic gradient descent methods [24]. Stochastic gradient descent consists of 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 Eqs (5) and (6). As before, if both M and f are convex, stochastic gradient descent methods are guaranteed to converge to the optimal solution of problem (4) [24].

4 Overview of techniques

Our novel contribution is to show how to apply the gradient method to any disease control scenario and optimize over *discrete* or *continuous* response policies. In both cases the task is first to identify a suitable functional A, that is to be minimised, and then to show how to compute the gradient of it, which appears in the first line of the Adam algorithm in Eq (6), such that the algorithm can step to the next iteration.

In a discrete policy, the parameters describing the government intervention can only take a finite number of values. Eq (1) is an example of a discrete policy, because on day t the government can either declare a lockdown ($s(t) = 1$) or declare no lockdown ($s(t) = 0$): we allow for no values of $s(t)$ inbetween.

On the other hand, a continuous policy is one in which the parameters that define the policy, which in general we denote by the vector $\boldsymbol{\mu}$, are allowed to vary over a continuum. For example, the fraction $v(t)$ of the population vaccinated on day t can take any value between 0 and 1.

Regardless of whether we chose to optimise over discrete or continuous policies, our starting point is an ordinary differential equation of the form

$$\frac{dx^i}{dt} = G^i(t, \boldsymbol{x}; \boldsymbol{\mu}); \quad i = 1, \ldots, m, \tag{10}$$

where the entries of the vector $\boldsymbol{x}$ represent the occupations of the different compartments of a disease model and $\boldsymbol{\mu}$ represents a continuous or discrete parametrization of the effects of a given policy. Let $\boldsymbol{x}(t; \bar{\boldsymbol{\mu}}, \boldsymbol{x}_0)$ be the solution of Eq (10) with initial conditions $\boldsymbol{x}(0) = \boldsymbol{x}_0$ and $\boldsymbol{\mu} = \bar{\boldsymbol{\mu}}$, where the initial policy condition is typically taken to be 'lockdown off'. We consider the problem of finding the parameters $\boldsymbol{\mu}^\star$ 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 periods, etc. If the initial conditions $\boldsymbol{x}_0$ are known, then it makes sense to consider functionals of the form

$$A(\boldsymbol{\mu}, \boldsymbol{x}_0) = \int_{t_0}^{t_f} dt \; \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0)) + \hat{A}(\boldsymbol{\mu}), \tag{11}$$

where, as discussed in Section 3, constraints which we wish the solution $\boldsymbol{x}(t; \boldsymbol{\mu}^\star, \boldsymbol{x}_0)$ to satisfy are incorporated into $\mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0))$. For instance, if we want that the number of patients in critical care does not exceed the number of beds, i.e. Eq (2) to hold, then one of the terms in $\mathcal{L}$ would be

$$\rho(t)(C(t; \boldsymbol{\mu}, \boldsymbol{x}_0) - B_c)^+, \tag{12}$$

with $\rho(t) \gg 1$. Similarly, to model a constraint on the vaccine supply of the form $\int dt\, v(t;\boldsymbol{\mu}) \leq V$, we would add the term

$$\lambda \left(\int dt\, v(t;\boldsymbol{\mu}) - V \right)^{+}, \tag{13}$$

with $\lambda \gg 1$.

As it turns out, minimizing functional (11) (or more complicated ones, see Appendix B) requires different techniques depending on whether the desired policy is continuous or discrete.

4.1 Continuous policies

If all the parameters $\boldsymbol{\mu}$ defining the policy are allowed to take arbitrary values in $\mathbb{R}^n$ and their effect on the disease's equations of motion (26) is differentiable, then one can simply apply gradient descent methods to minimize $A(\boldsymbol{\mu}, \boldsymbol{x}_0)$. The only difficulty stems in computing $\nabla_{\boldsymbol{\mu}} A$ (i.e., the first line of the Adam algorithm in Eq (6)). For functionals of the form (11), we have

$$\frac{\partial A}{\partial \mu_j} = \int_{t_0}^{t_f} dt \left(\frac{\partial \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t;\boldsymbol{\mu}, \boldsymbol{x}_0))}{\partial \mu_j} + \sum_i \frac{\partial \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t;\boldsymbol{\mu}, \boldsymbol{x}_0))}{\partial x^i} \frac{\partial x^i}{\partial \mu_j} \right) + \frac{\partial \hat{A}(\boldsymbol{\mu})}{\partial \mu_j}. \tag{14}$$

In order to evaluate this quantity to step the algorithm, the challenge is to compute the derivatives $\frac{\partial x^i}{\partial \mu_j}$. In Appendix B we show that these derivatives arise as the solution of a system of ordinary differential equations, of complexity comparable to that of Eq (10). We also explain how to use stochastic gradient descent to deal with scenarios where the model parameters (e.g., the basic reproduction number R_0) and/or the initial conditions $\boldsymbol{x}_0$ are only known to lie within some bounds, or when the evolution is stochastic. In either case, once a method to compute (14) is established (or, in the case of uncertainty in the initial conditions, some unbiased statistical estimator thereof (see Appendix B)), one can simply run the vanilla gradient descent (5) or the Adam algorithm (6) to arrive at a quasi-optimal policy.

4.2 Discrete policies

When some or all the entries of $\boldsymbol{\mu}$ are restricted to take values on a finite, fixed set, it is clear that one cannot apply gradient descent methods straightforwardly in order to minimize the objective function. In fact, optimizations over discrete variables are, in general, a very difficult endeavor: it can be argued that problems which appear simple do not have an efficient solution [28]. Nonetheless, in Appendix C we present two heuristics to tackle such optimizations in the context of policies for disease control.

The first heuristic consists of mapping the original minimization problem to an optimization over *probabilistic* policies, whereby the government decides how to intervene by sampling a discrete probability distribution, which is continuously dependent on a number of auxiliary (continuous) parameters. By applying stochastic gradient descent methods over such auxiliary parameters, we can find the probabilistic policy that minimizes the average value of the objective function. As shown in Appendix C, independently of the initial conditions, the gradient method is guaranteed to converge to a deterministic policy.

The second heuristic is tailor-made for optimizations over lockdown (discrete) policies, although it can be easily extended to deal with arbitrary discrete policies. Consider a scenario in which the government is allowed to declare or lift a lockdown at any time. In this case, we can parametrize the policy by the duration of each lockdown phase and the pauses in between, and use gradient descent methods to arrive at the optimal times. In Appendix C we show that

estimating the correspondent gradients only requires a minor modification of the methods developed for continuous policies.

5 Application: COVID-19

To illustrate the use of gradient descent for policy design, we consider a variant of the extended "susceptible-exposed-infected-recovered" (SEIR) disease model proposed in [15] 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. Each state variable X in the diagram corresponds to the *fraction* of the whole population $\mathcal{N}_{pop}$ in the corresponding diagram. From now on, we denote intensive quantities like X with a normal mathematical font, while extensive quantities $\mathcal{X} = X \times \mathcal{N}_{pop}$ are to be represented with calligraphic letters. The clinical parameters of the model in Fig 1, such as recovery and hospitalization rates, were estimated in [29, 30], based on early reports from COVID-19 cases in the UK, China and Italy. Following the model in [15], the disease transmission is assumed to be seasonal by analogy with the known behavior of betacoronaviruses such as HCoV-OC43 [30], with a baseline reproduction number between 2 and 2.5, following fits of the early growth-rate of the epidemic in Wuhan [11, 12].

A relevant compartment in this model is $\mathcal{C}(t) \equiv C_C(t) \times \mathcal{N}_{pop}$, the population occupying a critical care bed 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 $\mathcal{N}_{pop} = 47$ million and $\mathcal{B}_c = 9.5 \times 10^{-5} \times \mathcal{N}_{pop}$. That is, we assumed that the healthcare system provides 95 critical care beds 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 Fig 2 (see Appendix A for the exact initial conditions of our numerical simulations). As the reader can appreciate, between the third and seventh months, the number of people in need of critical care exceeds the capacity of the considered healthcare system by 18 times.

Vaccines against COVID-19 were not available during the first year of the pandemic. Hence initially most governments 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 [15, 30], i.e. $R_0 \rightarrow rR_0$, where

$$r := (\bar{r} - 1)s + 1. \tag{15}$$

Depending on the discrete or continuous nature of the intervention, we shall consider two kinds of such non-pharmaceutical policies. On one hand, we will speak of *lockdown policies* when s is only allowed to take the values $\{0, 1\}$; those 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 (15). On the contrary, 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, we will speak of *physical-distancing policies* when s is allowed to take any value in the interval $[0, 1]$. Continuous values of $s(t)$ correspond to intermediate measures (for example mandatory face masks, suspension of sport events, remote working, school closures), the effect of which can be tentatively estimated from available data [29].

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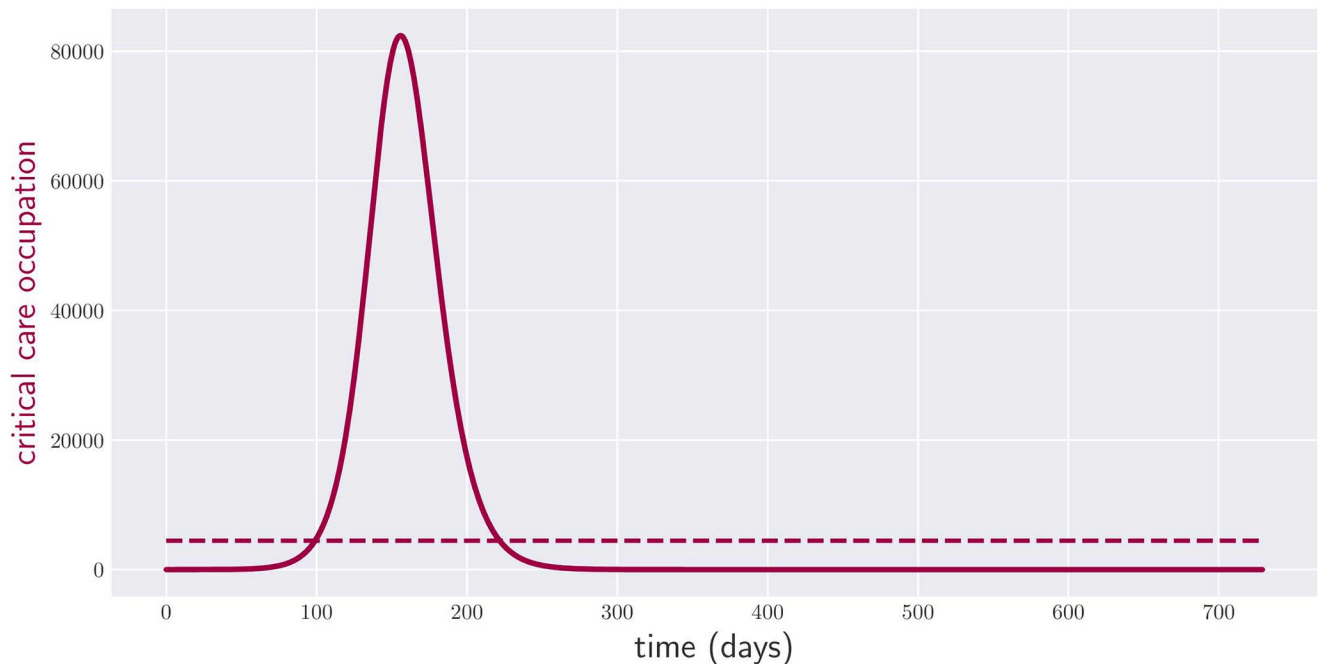
Fig 2. Occupation of critical care beds over two years with no policy intervention. The red dashed line indicates the critical care capacity of the healthcare system.

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

$$s(t) = \sum_k \sigma(\tilde{s}_k)\chi_{I_k}(t). \tag{16}$$

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, for $t \in [t_0, t_f]$, $s(t; \{s_k\})$ is everywhere differentiable, making it amenable to the gradient method. The parameters to optimize over are $\boldsymbol{\mu} \equiv \{\tilde{s}_k\}_k$, since they fully define the government's disease policy.

In 2021, several vaccines against COVID-19 successfully passed clinical trials, and nowadays many governments are fighting the disease through national vaccination campaigns. We model the effect of a COVID-19 vaccination campaign by assuming that: a) the government vaccinates the population across all compartments, i.e., without distinguishing between susceptible, exposed, recovered, etc.; b) just susceptible individuals benefit from the vaccine; c) all such vaccinated individuals become immune to the disease. The reader can find the explicit model in Appendix A. Obviously, this model is a gross simplification of the effect of COVID-19 vaccines currently in the market, which on one hand varies depending on the particular vaccine and the prior exposure to COVID-19, and on the other hand does not seem to always confer full immunity to the disease [31]. More complicated vaccination models can be built to properly assess a government on vaccination policies, but, for the sake of illustration of our techniques, this simplified model will suffice.

We place a cap of $\Lambda = 0.0011 \, days^{-1}$ on the fraction of the population that can be vaccinated per day. We also assume that changes in the vaccination rate can only be made weekly. This brings us to parametrize the vaccination rate via the function

$$v(t) = \Lambda \sum_k \sigma(v_k) \chi_{I_k}(t). \tag{17}$$

Finally, we assume that the government holds a supply of vaccines for just a third of the total population. This means that, together with (2), we need to enforce the extra constraint

$$\int_{t_0}^{t_f} dt v(t) \leq \frac{1}{3}. \tag{18}$$

We achieve this by adding the term (13) to the objective function, with $V = \frac{1}{3}, \lambda = 10^4$. Without additional constraints, this model represent the case of a vaccine with 100% efficacy. A reduced efficacy, let us denote it by $e < 1$, can be straightforwardly modeled simply by multiplying the vaccination rate $v(t)$ and the vaccine supply V by e.

In all our examples, we 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). To enforce it, we add the term (12) to the objective function, with $\rho(t) = \frac{100}{B_c}$.

Finally, in order to solve the differential (29), we use the Euler explicit method (30) with step size $\delta = 1.0$ to generate all our plots, apart from in Figs 6 and 7, where we use $\delta = 0.1$.

Next, we use the gradient method to derive the optimal interventions to control COVID-19 in a number of disease scenarios.

5.1 Fighting COVID-19 via physical distancing measures

We first consider policies exclusively based on non-pharmaceutical interventions; more concretely, continuous weekly policies $s(t)$ of the form (16). For starters, we tackle the problem of quickly steering the population towards herd immunity, all the while respecting the constraint (2) on the critical care capacity. Our goal can be captured by a functional of the form (11), with $\hat{A}(\boldsymbol{\mu}) = 0$ and

$$\mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x_0})) = |S(t) - S_h|, \tag{19}$$

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 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 3 illustrates the results of optimizing the objective function in Eq (19) for continuous physical-distancing measures. The plot shows both the critical care occupancy $\mathcal{C}(t)$ and the total number $\mathcal{S}(t)$ of susceptible individuals between times t_0 and $t_f = t_0 + 3 \times 365$. The number of susceptible individuals reaches S_h on day 864, after which the disease can be considered extinct. As the reader can appreciate, the critical care occupancy curve never surpasses the critical value $\mathcal{B}_c$.

The results in Fig 3 are achieved at the cost of enforcing constant physical-distancing measures throughout the considered time frame. Naturally, the next problem we consider is that of minimizing the aggregate economic cost $\mathcal{E}$ associated with the physical-distancing measures implemented by the government over the first two years of the disease (while respecting the critical care capacity constraint (2)). Thus we take an objective function of the form (11), with
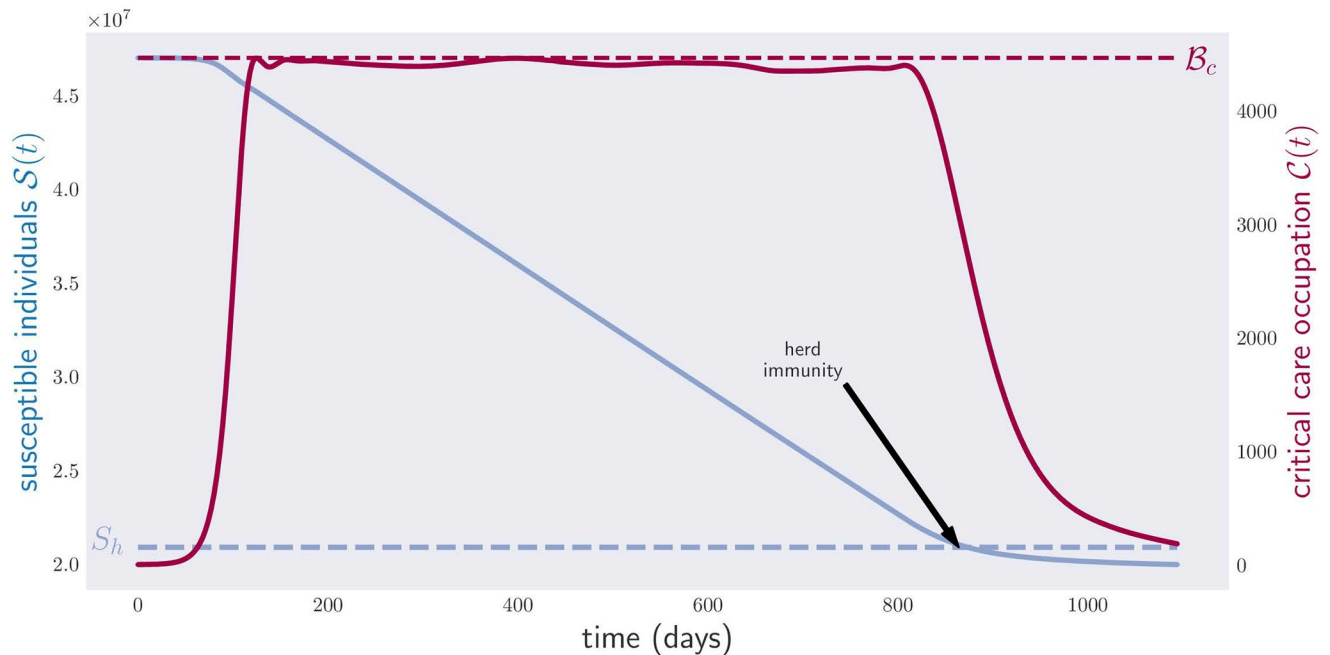
Fig 3. Occupation of critical care beds (red) and population of susceptible individuals (blue) for a period of three years. The blue line corresponds to the level of susceptibles guaranteeing herd immunity over the whole year.

$\hat{A}(\boldsymbol{\mu}) = 0$ and

$$\mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x_0})) = \mathcal{E}(s(t; \boldsymbol{\mu})) \, . \tag{20}$$

Fig 4 shows the result of applying the gradient method to minimize this functional for the cost function $\mathcal{E}(s) = s$.

This plot shows the critical care occupancy $\mathcal{C}(t)$, and also 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 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 (20) to make the final slope of the curve $\mathcal{C}(t)$ less steep, or even decreasing.

5.2 Fighting COVID-19 via lockdowns

In some circumstances, the only distancing measures considered by governments are discrete: lockdown on or off, as in Eq (1). Let the objective function be given by Eq (20), with $\mathcal{E}(s) = s$, i.e., we wish to minimize the total time under lockdown. As stated earlier, optimizations over discrete government interventions cannot be carried out directly with the gradient method. We conduct them instead with the two heuristics proposed in Appendix C.

The first heuristic allows optimizing over weekly on/off confinements, and its results are shown in Fig 5. This time the critical care occupancy curve touches the critical care capacity just once, 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
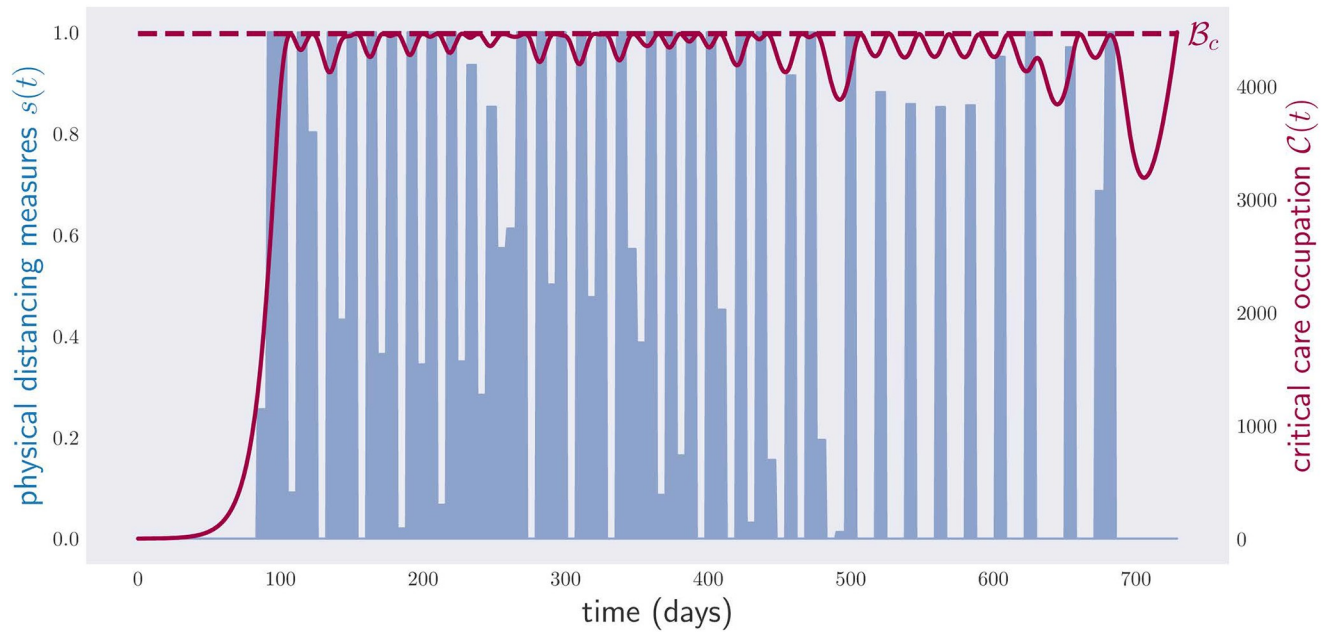
Fig 4. Occupation of critical care beds (red) and physical-distancing measures (blue) for a period of two years. The optimization has been performed over continuous weekly policies, 104 continuous parameters, i.e., any value of physical-distancing measure *s* between 0 and 1 is accepted. The algorithm, however, tends to prefer 0/1 configurations, i.e., full lockdown or no lockdown in most instances.

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* (as we will soon see, one can limit the total number of lockdowns in the final policy by adding
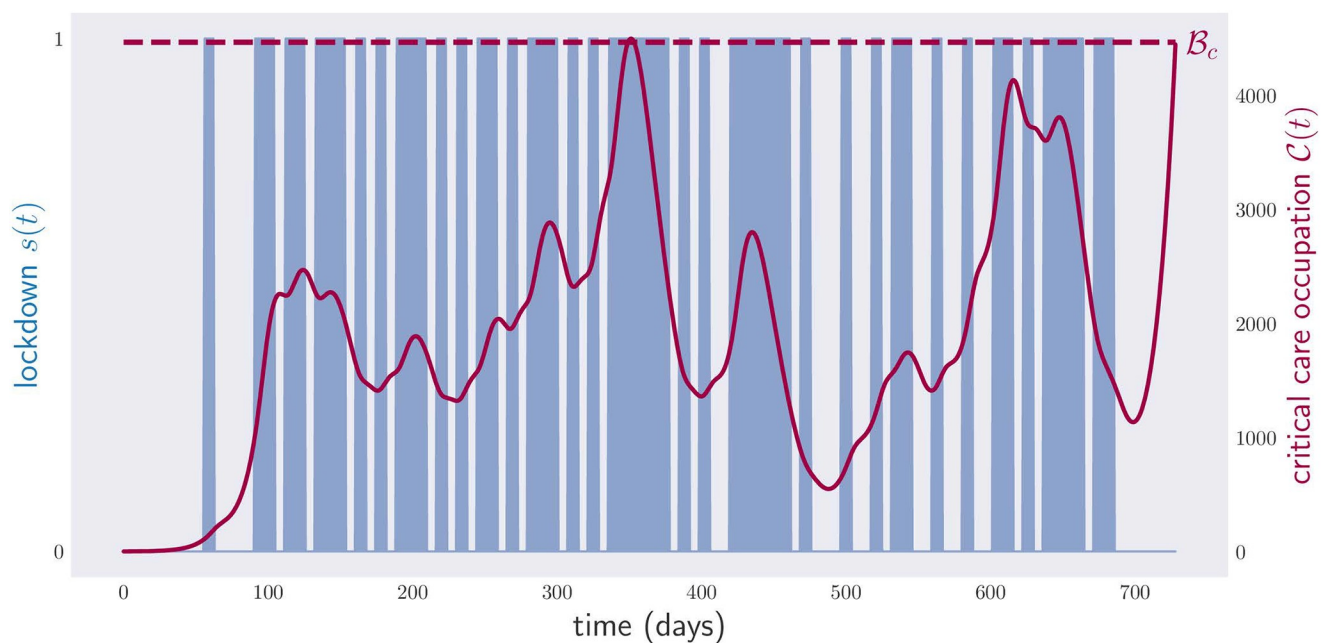

Fig 5. Occupation of critical care beds (red) and lockdown (blue) for a period of two years. The plot shows the result of the optimization over probabilistic policies via gradient descent over a period of two years.
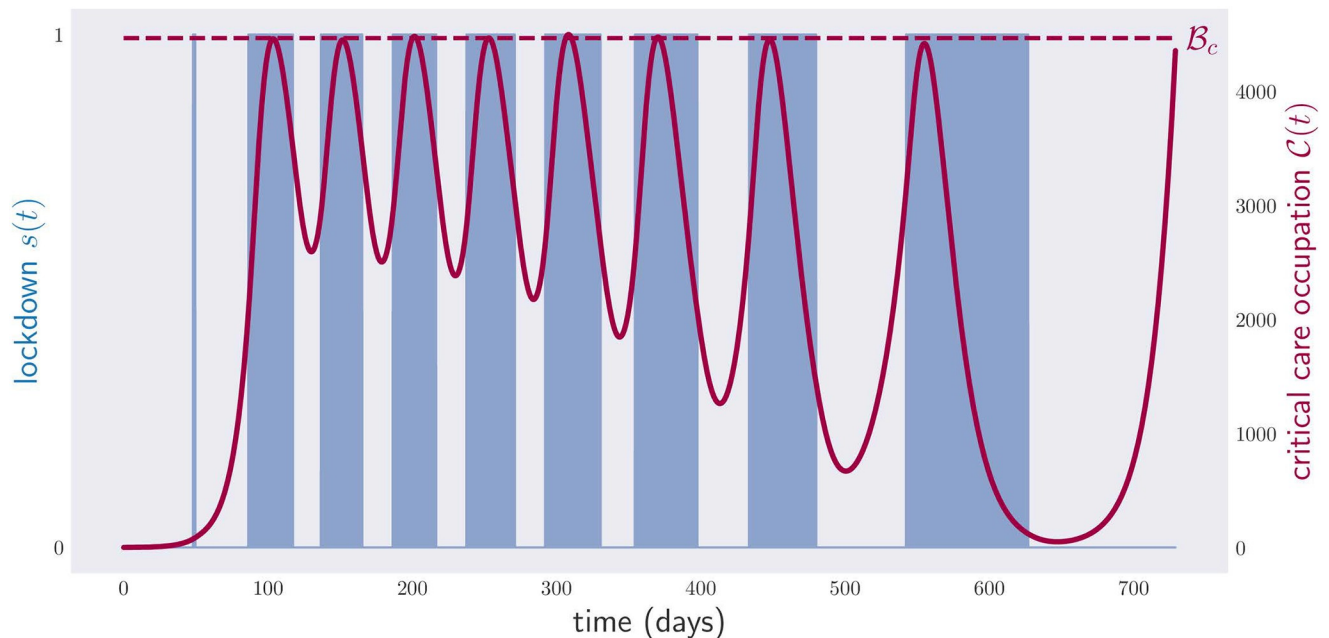
Fig 6. Occupation of critical care beds (red) and lockdown measures (blue) for a period of two years. Optimization over deterministic policies with arbitrary initial and final times for each lockdown period. A total of 9 lockdown periods has been fixed prior to the optimization. Notice that the first lockdown period, around day 60, has been basically removed by the optimization procedure.

constraints to the optimization problem). The total length of the lockdown in the course of two years is 371 days.

In the second discretization method, the disease control policy is continuously parametrized through the vector $\boldsymbol{\mu} = (t_1, t_2, \ldots, t_{2N})$, 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 heuristic. 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 Fig 6. The total length of the lockdown is 338 days.

5.2.1 Fighting COVID-19 via limited lockdowns and vaccination. We next study the effect of vaccination campaigns in reducing the total time spent in lockdown. To ensure that the final recommended policies are easy to implement, we place the limit $N = 5$ on the maximum number of lockdowns.

If we allow the government to declare or lift lockdowns at any point in time, the second heuristic outputs the policy depicted in Fig 7. Note that the optimal policy does not require 5 lockdowns, but 4. The policy involves vaccinating the population at maximum rate since the very beginning of the government intervention until about day 200, after which the vaccination rate gradually drops to zero. The total span of the required lockdown is 186 days. This must be compared with the 338-days lockdown required when we allow for $N = 10$ lockdowns, but no vaccines are available. Even at such a slow vaccination pace, and under a shortage of supplies, the effects of a vaccination campaign prove to be very impressive.

Let us now see how the picture changes when the government is not allowed to enforce an intervention in the middle of the week. Enforcing a maximum number N of lockdowns is not automatic when we deal with the first heuristic; it requires us to add an extra constraint to the
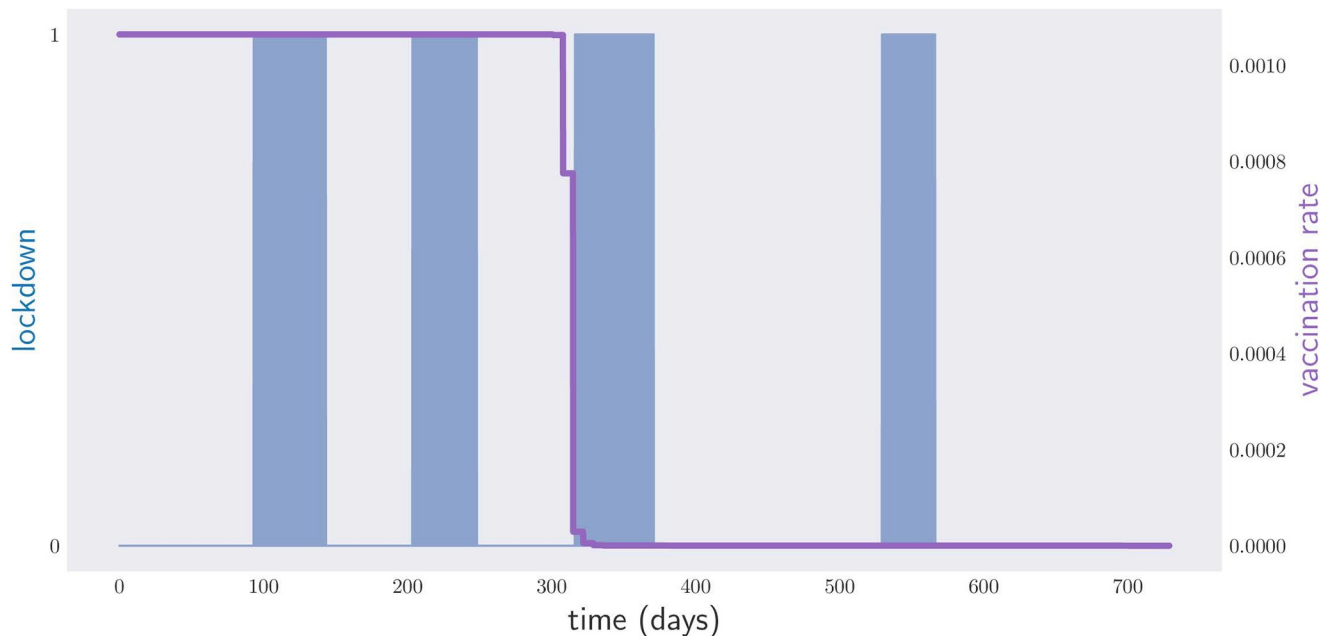
Fig 7. Lockdown times (blue) and vaccination rates (magenta) for a period of two years.

optimization over non-deterministic policies. A possibility is to include the term

$$\Omega \left\langle \left(\sum_j |c_j - c_{j+1}| - 2N \right)^+ \right\rangle \tag{21}$$

in the objective function. Here the random variable c_k equals 1 if there was a lockdown on week k or 0 otherwise, see Appendix C. The penalty for changing the government response by $2N + x$ times within the period $[t_0, t_f]$ is therefore $x \times \Omega$. Choosing $\Omega = 10^2$, we arrive, through stochastic gradient descent, at the deterministic policy depicted in Fig 8. It demands a total lockdown time of 231 days (compared to the 371 days of lockdown required by the corresponding non-pharmaceutical policy, depicted in Fig 5). As the reader can appreciate, the time spent in lockdown is considerably higher when we require lockdowns to be enforced or lifted on Mondays than when we allow the government to intervene at arbitrary times.

5.3 Dealing with uncertainty

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 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 [16].

5.3.1 Uncertainty in the parameters. These considerations make us question how practical a two-year disease control policy really is. Consider the physical-distancing policy depicted in Fig 4, which was obtained by applying the gradient method to the SEIR model in [15]. Here,
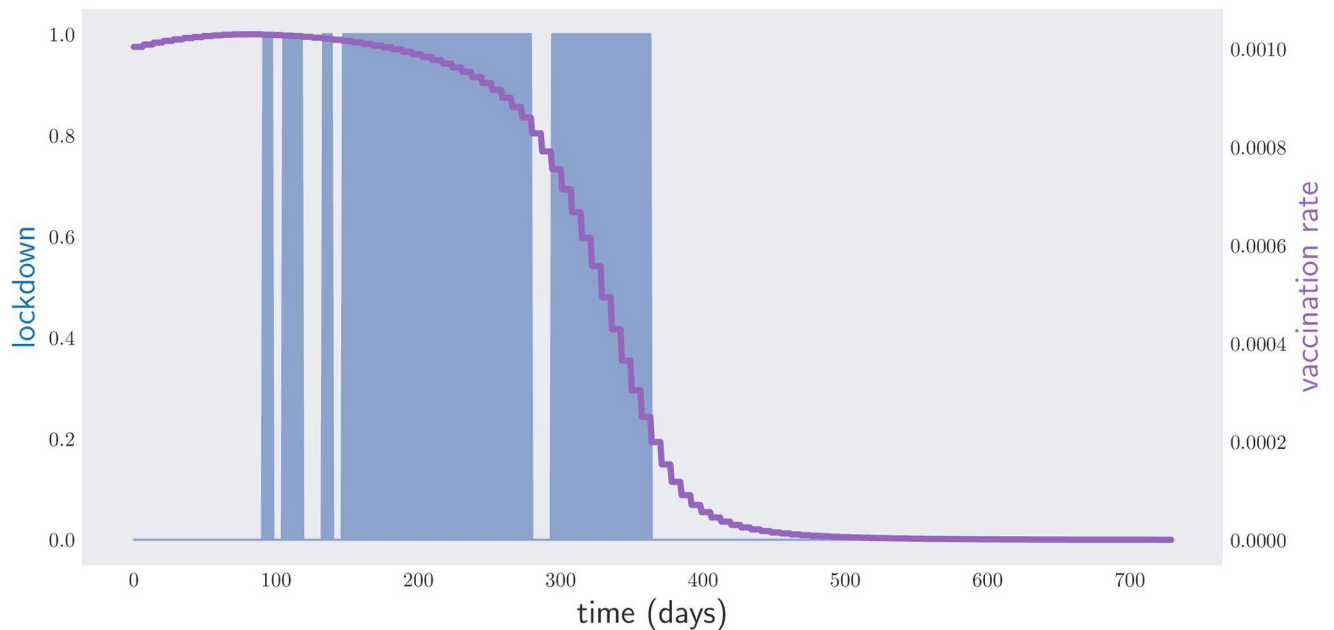
Fig 8. Weekly lockdown times (blue) and vaccination rates (magenta) for a period of two years.

the model parameters v correspond to the average values $\bar{v}$ of the parameter ranges in Table 1 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 Δv 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 $\mathcal{N}_a = \left\{ v : -a\frac{\Delta v}{2} \leq v - \bar{v} \leq a\frac{\Delta v}{2} \right\}$, where a can be interpreted as the amount of noise or uncertainty. How robust is the afore-mentioned 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 4. As one can see, for some sampled values of parameters, the critical care capacity of the healthcare system is exceeded. This is not surprising, since the policy depicted in Fig 4 was devised to perform well under the assumption that $v = \bar{v}$ and not $v \in \mathcal{N}_{0.05}$.

In order to tame the behavior in the plot in Fig 9, we use stochastic gradient descent to minimize the average value of the objective function A assuming a uniform distribution of v over $\mathcal{N}_a$, as explained at the end of Appendix B. By adding to A sufficiently strong penalties for the violation of each optimization constraint, we make sure that such constraints will approximately hold for most of the points in $\mathcal{N}_a$.

Fig 10 shows a lockdown policy minimizing the physical distancing measures under the condition that constraint (2) holds for different values of $v \in \mathcal{N}_{0.05}$, 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.
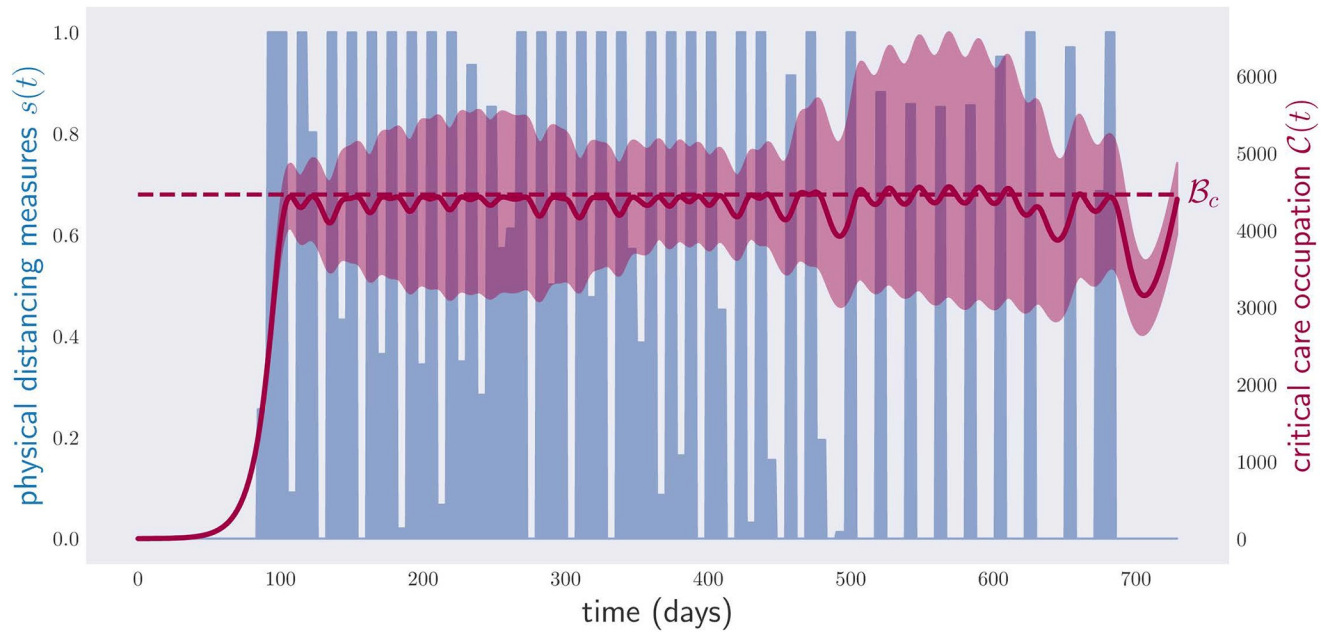
Fig 9. Occupation of critical care beds (red) and (unoptimized) physical distancing measures (blue) for a period of two years under random parameters in $\mathcal{N}_{0.05}$. The region in red is obtained by sampling 1000 times from the region of model parameters $\mathcal{N}_{0.05}$ and evolving the corresponding models with the physical distancing policy optimized over the model with average-value parameters (as in Fig 4). 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.

https://doi.org/10.1371/journal.pone.0257958.g009

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.

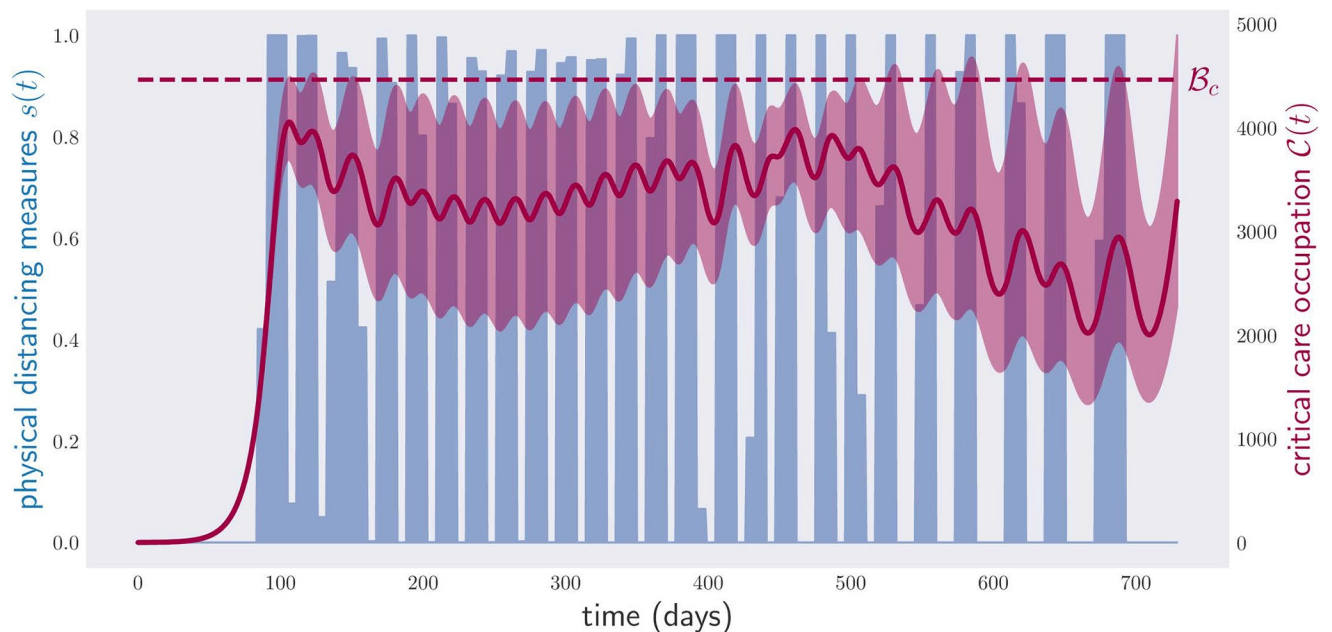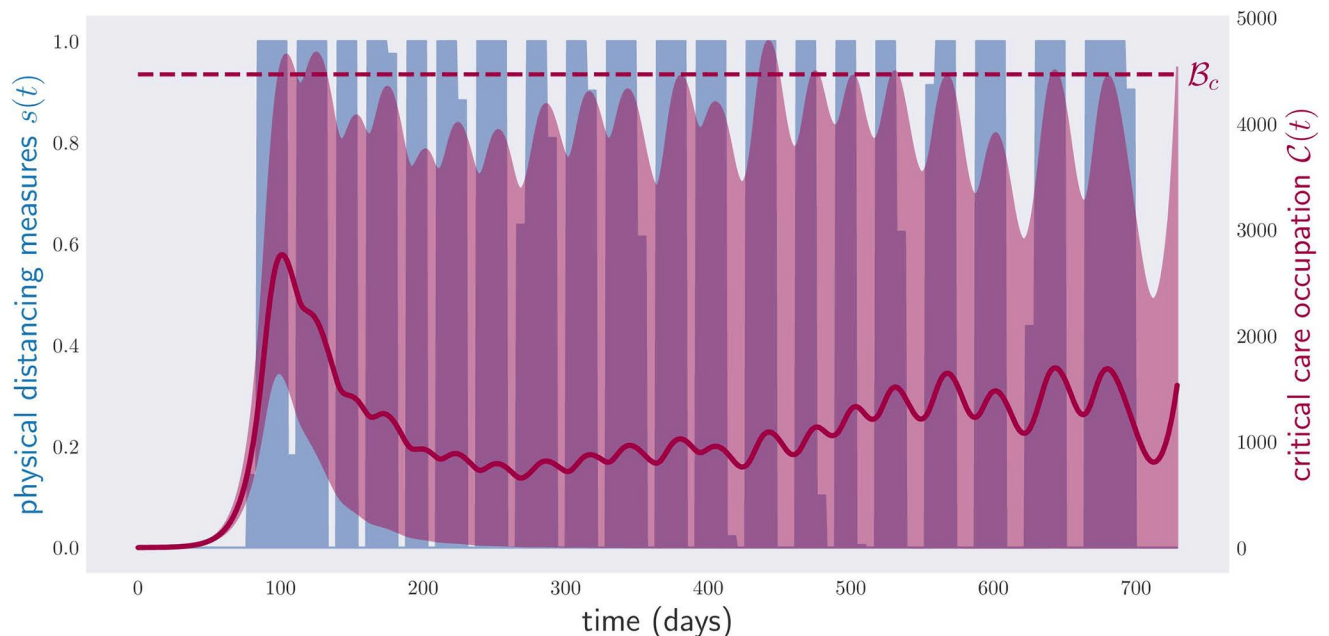https://doi.org/10.1371/journal.pone.0257958.g010

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 $\mathcal{N}_{0.25}$. 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.

https://doi.org/10.1371/journal.pone.0257958.g011

This result is what one would have expected. As time goes by, the predictions of the disease model for different values of $\mathbf{v}$ diverge: any policy that 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 $\mathbf{v}$, 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 $\mathbf{v}$ satisfy $v \in \mathcal{N}_{0.25/\sqrt{k}}$. Every four weeks, the policy is recalculated to minimize the physical distancing measures for the rest of the two years ahead, using the range $v \in \mathcal{N}_{0.25/\sqrt{k}}$. The final curves for the critical care occupation and the physical distancing measures are shown in Fig 12 for $\mathbf{v}$, 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 $v \in \mathcal{N}_{0.25}$.
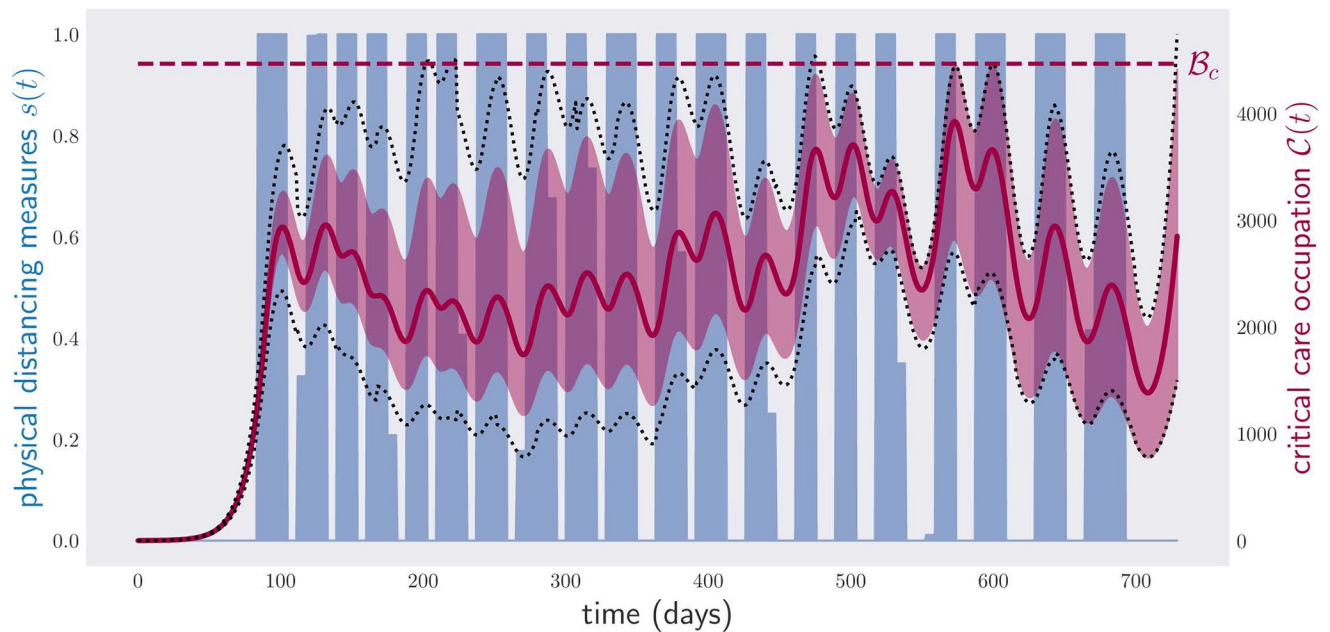
Fig 12. Occupation of critical care beds (red) and physical distancing measures (blue) for a period of two years under random parameters and monthly noise decrease. The disease control policy was optimized to respect the condition in Eq (2), i.e., critical care capacity not exceeded, starting with the 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 black-dotted region 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 red 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 v of the disease model, the final lockdown time is much lower, due to monthly revisions of the original policy plan.

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 generic adaptive policies described by a continuous version of a neural network architecture known as Long Short-Term Memory (LSTM) [32]. 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.

5.3.2 Nondeterministic models. So far, all models considered in the optimization are deterministic, namely, given the model parameters and initial conditions, the compartments evolve along a unique trajectory. As a more general case, we can consider the one in which the disease's dynamics is governed, rather than by Eq (23), by a stochastic differential equation. More concretely, consider the model that results when we replace the first and second lines of Eq (23) by

$$
\frac{dS}{dt} = -(1 + \xi)r(t)\beta(t)S(I_R + I_H + I_C),
$$
$$
\frac{dE}{dt} = (1 + \xi)r(t)\beta(t)S(I_R + I_H + I_C) - vE,
$$

(22)

where ξ represents a Gaussian noise term reflecting stochastic fluctuations on the virus' transmissivity [5]. Solving this equation by discretization, each occurrence of ξ at time $t_k = t_0 + k\delta$ is
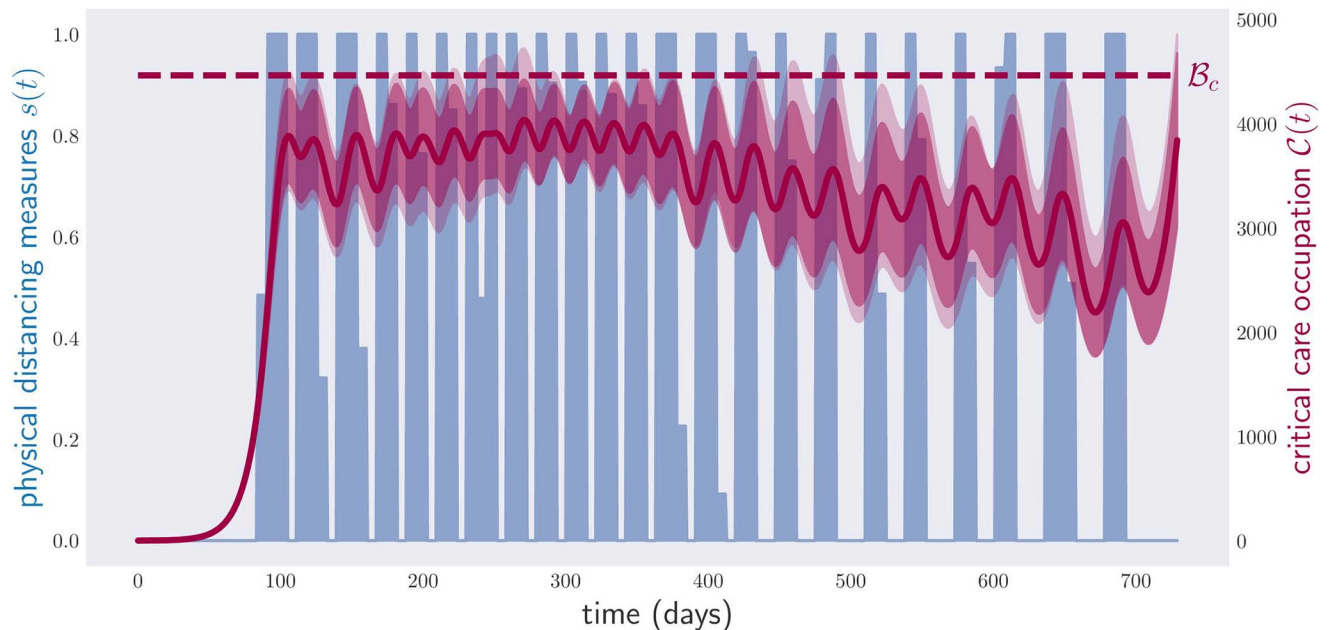
Fig 13. Occupation of critical care beds (red) and physical distancing measures (blue) for a period of two years for stochastic evolution with noise-level $\eta = 0.05$. The disease control policy was optimized to respect the condition in Eq (2), i.e., critical care capacity not exceeded. The region in light red depicts the range of critical care occupations observed in a sample of $1000 \times \eta$ different evolutions, with a stochastic noise $\eta = 0.05$; the red line, the average critical care occupation; and the darker red region, the trajectories that lie within two standard deviations from the average value. The policy corresponds to a total of 321 lockdown days.

replaced by a quantity ξ_k sampled from a Gaussian distribution of zero mean and standard deviation $\eta\sqrt{\delta}$, where η denotes the magnitude of the noise term.

For our simulation, we consider the noise $\eta = 0.05$. As before, the method of stochastic gradient descent is used for obtaining a noise-robust policy. At each time step, we average the gradient over $1024 \times \eta$ different stochastic evolutions. In contrast to the case of noisy parameters, for a stochastic evolution the sampling has to be repeated at each time step, see Appendix B for more details. The results of our simulations are shown in Fig 13. Notice that, in contrast to Fig 9, the noise is not uniformly distributed in a limited interval, but is given by a Gaussian, which allows for (rare) events far away from the mean. This justifies the different representations of uncertainty in these plots. Despite the presence of relatively high noise in the evolution, as shown by the wide fluctuations in Fig 13, the algorithm is able to devise strong policies that maintain the critical care occupation almost always, i.e., with a high probability, below its capacity.

6 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, and stochastic evolution. 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 non-pharmaceutical interventions 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 that 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. We also showed how vaccine supplies, even when meager and poorly distributed, can dramatically shorten the total confinement required to keep the disease under control.

Admittedly, some of our computer-generated policies are too complicated to be implemented in practice. Our formalism allows, however, to limit the complexity of the found solutions by adding extra constraints to the original optimization problem. To highlight this feature, we generated a near-optimal weekly policy plan of vaccinations and on/off weekly confinement measures with a cap on the total number of lockdowns.

We examined the problems that one may encounter when applying our techniques to scenarios where the model parameters are not known with high accuracy, or the model evolves stochastically. This 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 devising a gradient-friendly ansatz for adaptive policy plans for disease 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 applied to the standard LSTM architecture 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 $\mu^{(0)}$ 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.

7 Appendix

A Models for the spread of COVID-19

In all our numerical simulations, we 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 adopt a simplified version of the model
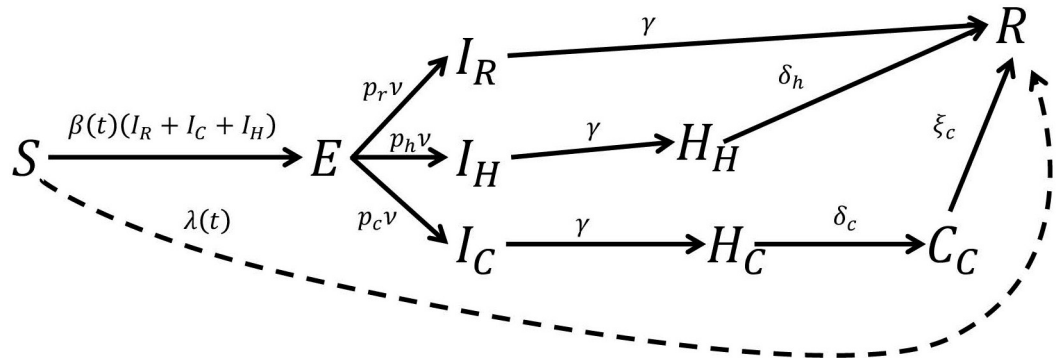
Fig 14. Two COVID-19 compartmental models for disease policies based on physical distancing measures. The solid lines show the model first presented in [30], where the coefficients connecting the compartments are interpreted couplings, population fractions, time delays and so forth. The addition of the dotted line connecting the 'susceptible' and 'recovered' compartments creates another model where a vaccination campaign is taken into account.

used in [15]. 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 Fig 14:

$$
\begin{aligned}
\frac{dS}{dt} &= -r(t)\beta(t)S(I_R + I_H + I_C), \\
\frac{dE}{dt} &= r(t)\beta(t)S(I_R + I_H + I_C) - vE, \\
\frac{dI_R}{dt} &= p_R vE - \gamma I_R, \\
\frac{dI_H}{dt} &= p_H vE - \gamma I_H, \\
\frac{dI_C}{dt} &= p_C vE - \gamma I_C, \\
\frac{dH_H}{dt} &= \gamma I_H - \delta_H H_H, \\
\frac{dH_C}{dt} &= \gamma I_C - \delta_C H_C, \\
\frac{dC}{dt} &= \delta_C H_C - \xi C, \\
\frac{dR}{dt} &= \gamma I_C + \delta_H H_H + \xi C.
\end{aligned}
\tag{23}
$$

Here

$$
\beta(t) = \gamma R_0 \left(\frac{1+\Delta}{2} + \frac{1-\Delta}{2} \cos\left(\frac{2\pi(t+\phi)}{7 \times 52} \right) \right)
\tag{24}
$$

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 [15], and appear in Table 1.

Table 1. Parameter ranges for the compartmental model for COVID-19 proposed in [15].

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

https://doi.org/10.1371/journal.pone.0257958.t001

If the government has a vaccine available, the model changes. In that case, the first and last lines of Eq (23) shall be replaced by

$$\frac{dS}{dt} = -r(t)\beta(t)S(I_R + I_H + I_C) - v(t)S,$$

$$\frac{dR}{dt} = \gamma I_C + \delta_H H_H + \xi C + v(t)S,$$

(25)

where $v(t)$ denotes the vaccination rate.

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 1. We assume that the government can vaccinate at most 50, 000 individuals per day. This means that, at any given time t, $v(t) \leq \Lambda := \frac{50,000}{47,000,000} \approx 0.00106$. In addition, we assume that the total supply of vaccines can just cover a third of the total population. In other words, $\int dt v(t) \leq \frac{1}{3}$.

B Optimization over continuous measures for disease control

In this section, we explain how to apply the gradient method to optimize over continuous classes of government interventions. Our starting point is an ordinary differential equation of the form

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

(26)

The entries of vector x represent the occupations of the different compartments of a disease model. In the case of adaptive policies, some of such entries might also represent the components of the *cell state* $\boldsymbol{\theta}$ [32], i.e., the internal variables used by the government to keep track of the evolution of the disease and guide future government interventions. $\boldsymbol{\mu} \in \mathbb{R}^n$ represents a parametrization of the effects of a given policy. Let $\boldsymbol{x}(t; \bar{\boldsymbol{\mu}}, \boldsymbol{x}_0)$ be the solution of Eq (26) 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 the means by which we wish to control the disease: it might represent the number and duration of lockdown, etc. For the time being, let us assume this functional to be of the form

$$A(\boldsymbol{\mu}, \boldsymbol{x}_0) = \int_{t_0}^{t_f} dt \, \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0)) + \hat{A}(\boldsymbol{\mu}) \, , \tag{27}$$

From the discussion in Section 4, the functional in Eq (27) might also contain constraints which we wish the solution $\boldsymbol{x}(t; \boldsymbol{\mu}^\star, \boldsymbol{x}_0)$ to satisfy, such as (12) and (13). 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.

To minimize $A(\boldsymbol{\mu}, \boldsymbol{x}_0)$ via gradient descent, we need to compute $\nabla_{\boldsymbol{\mu}} A$. For functionals of the form (27), we have that

$$\frac{\partial A}{\partial \mu_j} = \int_{t_0}^{t_f} dt \left(\frac{\partial \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0))}{\partial \mu_j} + \sum_i \frac{\partial \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \boldsymbol{x}_0))}{\partial x^i} \frac{\partial x^i}{\partial \mu_j} \right) + \frac{\partial \hat{A}(\boldsymbol{\mu})}{\partial \mu_j} \, . \tag{28}$$

The next question is thus how to compute the derivatives $\frac{\partial x^i}{\partial \mu_j}$. To this aim, define the variables $y_j^i(t; \boldsymbol{\mu}, \boldsymbol{x}_0) \equiv \frac{\partial x^i(t; \boldsymbol{\mu}, \boldsymbol{x}_0)}{\partial \mu_j}$. Differentiating Eq (26) by μ_j, we have that

$$\frac{\partial y_j^i}{\partial t} = \frac{\partial G^i(t, \boldsymbol{x}; \boldsymbol{\mu})}{\partial \mu_j} + \sum_l \frac{\partial G^i(t, \boldsymbol{x}; \boldsymbol{\mu})}{\partial x^l} y_j^l; i = 1, ..., m, j = 1, ..., n. \tag{29}$$

In order to obtain $\{y_j^i(t)\}$ for each time t, it hence suffices to solve the system of coupled differential equations given by (26) and (29) with initial conditions $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$, $y_j^i(t_0) = 0$. This can be achieved numerically through several different methods, depending on the desired accuracy. The simplest such method is called *Euler explicit* [33]: for some $\delta > 0$, it consists of regarding time as a discrete variable of the form $t_k = t_0 + \delta k$, for $k = 0, ..., \lceil \frac{t_f - t_0}{\delta} \rceil$. The quantities $\{x^i(t_k), y_j^i(t_k) : k\}$ are then obtained by recursively applying the relations

$$x^i(t_{k+1}) = x^i(t_k) + \delta G^i(t_k, \boldsymbol{x}(t_k); \boldsymbol{\mu}),$$

$$y_j^i(t_{k+1}) = y_j^i(t_k) + \delta \left(\frac{\partial G^i(t_k, \boldsymbol{x}(t_k); \boldsymbol{\mu})}{\partial \mu_j} + \sum_l \frac{\partial G^i(t_k, \boldsymbol{x}(t_k); \boldsymbol{\mu})}{\partial x^l} y_j^l(t_k) \right). \tag{30}$$

We will also encounter situations where our functional A is more complicated than (27). Some parameters ζ (not policy parameters) regulating the evolution (26), such as the basic reproduction number of the disease, 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

$$A(\boldsymbol{\mu}) = \int p(\zeta, \boldsymbol{x}_0) d\zeta d\boldsymbol{x}_0 \int dt \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \zeta, \boldsymbol{x}_0)) + \hat{A}(\mu), \tag{31}$$

for some probability measure $p(\zeta, \boldsymbol{x}_0) d\zeta d\boldsymbol{x}_0$. Again, we wish to minimize A over $\boldsymbol{\mu}$. As explained in section 3, this can be achieved via stochastic gradient descent methods [23]: 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 $(\zeta^{(j)}, \boldsymbol{x}_0^{(j)})_{j=1}^N$ from the measure $p(\zeta, \boldsymbol{x}_0) d\zeta d\boldsymbol{x}_0$ and using them to compute the quantity

$$\tilde{\nabla}_{\boldsymbol{\mu}} A = \nabla_{\boldsymbol{\mu}} \hat{A}(\mu) + \frac{1}{N} \sum_{j=1}^N \nabla_{\boldsymbol{\mu}} \int dt \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; \boldsymbol{\mu}, \zeta^{(j)}, \boldsymbol{x}_0^{(j)})). \tag{32}$$

The prescription above also allows optimizing over control policies in scenarios where the disease's equations of motion are not deterministic, but probabilistic. Suppose, for instance, that the disease's dynamics are governed by a stochastic differential equation

$$\frac{dx^i}{dt} = G^i(t, \boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\xi}); i = 1, ..., m, \tag{33}$$

where the entries of $\boldsymbol{\xi} \in \mathbb{R}^s$ represent Gaussian white noise. When we solve this equation by discretization, each occurrence of $\boldsymbol{\xi}$ at time $t_k = t_0 + k\delta$ is to be replaced by an s-dimensional vector $\boldsymbol{\xi}^k$ of independent Gaussian variables of 0 mean and standard deviation $\sqrt{\delta}$. Call $M = \lceil \frac{t_f - t_0}{\delta} \rceil$. If we aim to minimize the expectation value of the objective function, we can estimate the gradient of the average through Eq (32), with $\zeta = (\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^M)$ and

$$p(\zeta) = \prod_{k=1}^M \left(\frac{1}{\sqrt{2\pi\delta}} \right)^s e^{-\left(\frac{\xi^{(k)}}{\sqrt{\delta}} \right)^2}, \tag{34}$$

and use stochastic gradient descent to find the minimum. In the specific case we considered in our simulations, the noise simply affected the "susceptible" and "exposed" compartments, where the usual infection rate is multiplied by a factor $(1 + \xi)$, thus modelling a stochastic fluctuation of the infection rate. This is arguably the most interesting term to study stochastic fluctuations, since the product SI appearing in the differential equation is at the origin of the nonlinearity of the SEIR model.

In the next section, we will use the same idea to carry out policy optimizations in scenarios where the disease's evolution is influenced by a finite number of discrete random variables.

C Optimization over discrete policies of disease control

The section above 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 $(s(t) = 1)$ or not declare a lockdown $(s(t) = 0)$. As we discuss in section 5 in the main text, the effect of a lockdown policy in the evolution of the disease can be modeled by introducing a term in Eq (26) that is proportional to $s(t)$. Namely, $G^i(t, \boldsymbol{x}; \boldsymbol{\mu}) = \hat{G}^i(t, \boldsymbol{x}; \boldsymbol{\mu}) + s(t) F^i(\boldsymbol{x})$.

To optimize over such discrete policies via gradient descent methods, one could think of introducing a continuous variable $\lambda \in \mathbb{R}$ and writing its effect on the disease's equations of motion by means of a piece-wise continuous function of λ, e.g.: $s(t) = \Theta(\lambda)$, for $t \in [t_1, t_2]$. Here $\Theta(z)$ denotes the Heaviside function (i.e., $\Theta(z)$ equals 1 for $z \geq 0$, or 0, otherwise). In this case, however, the gradient method would not work, since 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.

Applying the gradient method to optimize over policies for disease control involving discrete government interventions is therefore not straightforward. In the following, we propose two heuristics to tackle this problem.

C.1 Optimization over deterministic discrete policies through non-deterministic discrete policies. 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

$$\bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}) = \sum_{c_1,\dots,c_q=0,1} \prod_{k=1}^{q} p_k(c_k) \int dt \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; c, \boldsymbol{\mu}, \boldsymbol{x}_0)), \tag{35}$$

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 (35). 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 Section 3.

Let us first differentiate Eq (35) with respect to the continuous variables $\boldsymbol{\mu}$. The result is

$$\frac{\partial \bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu})}{\partial \mu_j} = \sum_{c_1,\dots,c_q=0,1} \prod_{k=1}^{q} p_k(c_k) v_j(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c) = \langle v_j(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c) \rangle_c, \tag{36}$$

where

$$v_j(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c) = \int dt \frac{\partial \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; c, \boldsymbol{\mu}, \boldsymbol{x}_0))}{\partial \mu_j} \tag{37}$$

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 (37) can be computed using the techniques discussed in Section B.

Differentiating Eq (35) with respect to $\tilde{s}_k$ we find that

$$\frac{\partial \bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu})}{\partial \tilde{s}_k} = \sum_{a=0,1} \langle w_k(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c^{(k,a)}) \rangle_{c^{(k,a)}}, \tag{38}$$

with

$$w_k(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}|c^{(k,a)}) = \frac{\partial p_k(a)}{\partial \tilde{s}_k} \int dt \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; c^{(k,a)}, \boldsymbol{\mu}, \boldsymbol{x}_0)), \quad \text{for } a = 0, 1 \tag{39}$$

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

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

$$\begin{aligned} \langle v \rangle &= \nabla_{\boldsymbol{\mu}} \bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}) \\ \langle w \rangle &= \nabla_s \bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}). \end{aligned} \tag{40}$$

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 (27) 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

$$c_k^\star = \begin{cases} 0 & \text{for } p_k^\star(0) > 1/2 \\ \\ 1 & \text{otherwise} \end{cases} \tag{41}$$

has the same objective value. Indeed, for $k \in \{1, \ldots, n\}$, fix $\{s_j^\star : j \neq k\}$, then

$$\bar{A}(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star) = \sum_{a=0,1} p_k^\star(a) A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = a), \tag{42}$$

with

$$A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = a) \equiv$$
$$\sum_{c_1,\ldots c_{k-1}, c_{k+1},\ldots,c_q=0,1} \prod_{j \neq k}^q p_j^\star(c_j) \int dt \mathcal{L}(t, \boldsymbol{\mu}^\star, c, \boldsymbol{x}(t; c_1, \ldots, c_{k-1}, a, c_{k+1}, \ldots, c_q, \boldsymbol{\mu}^\star, \boldsymbol{x}_0)). \tag{43}$$

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

$$\bar{A}(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star) = A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = 0) = A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = 1), \tag{44}$$

or, for some $a \in \{0, 1\}$,

$$\bar{A}(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star) = A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = a) < A(\tilde{\boldsymbol{s}}^\star, \boldsymbol{\mu}^\star | a_k = 1 - a), \tag{45}$$

with $p_k^\star(a) = 1, p_k^\star(1 - a) = 0$. In either case, fixing c_k through the procedure (41) cannot increase the average value of the objective function. Iterating over $k = 1, \ldots, q$, we prove the claim.

C.1.1 Generalization to optimizations over adaptive policies. The method described above can be easily extended to tackle optimization problems over weekly discrete adaptive policies. Consider an adaptive policy where the government intervention $c_k \in \{0, 1\}$ on week k is decided on week $k - l$ by sampling from a Bernoulli distribution dependent on the value $\boldsymbol{\theta}^k$ of the cell state on week $k - l$ (remember from Appendix B that the cell state at time t represents the government's internal memory and is given by some of the entries of $\boldsymbol{x}(t)$). The functional form of this distribution is determined by a vector of parameters $\boldsymbol{v}^k$. We thus have that $p_k(c_k) = p_k(c_k | \boldsymbol{\theta}^k; \boldsymbol{v}^k)$.

Note that $\boldsymbol{x}^k = \boldsymbol{x}^k(c_1, \ldots, c_{k-l-1})$. Starting from the initial conditions $\boldsymbol{x}_0$, the average objective function can therefore be estimated by propagating $\boldsymbol{x}$ week by week, computing the contribution to (27) and sampling each c_k at time $7(k - l)$ as we go along. If we instead compute the contribution to the gradient of (27) with respect to the continuous variables $\boldsymbol{\mu}$, we will have a statistical estimate of the gradient of the average objective function (with respect to $\boldsymbol{\mu}$).

Estimating the gradient of the average objective function with respect to the variables $\boldsymbol{v}^k$ is done similarly, by averaging over the appropriate Markov chain. More specifically,

$$\nabla_{\boldsymbol{v}^k} \bar{A}(\tilde{\boldsymbol{s}}, \boldsymbol{\mu}) = \sum_{a=0,1} \left\langle w_k(\tilde{\boldsymbol{v}}, \boldsymbol{\mu} | c^{(k,a)}) \right\rangle_{c^{(k,a)}}, \tag{46}$$

with

$$w_k(\tilde{\boldsymbol{v}}, \boldsymbol{\mu} | c^{(k,a)}) = \nabla_{\boldsymbol{v}^k} p_k(a | \boldsymbol{\theta}^k; \boldsymbol{v}^k) \times \int dt \mathcal{L}(t, \boldsymbol{\mu}, \boldsymbol{x}(t; c^{(k,a)}, \boldsymbol{\mu}, \boldsymbol{x}_0)), \text{ for } a = 0, 1. \tag{47}$$

This time, $c^{(k,a)}$ is sampled sequentially while we solve the differential equation, as described

before to estimate the average of the objective function. The only difference is that, at time $t_0 + 7(k - l)$, instead of sampling c_k, we set it equal to a.

Now, let us assume that, for $y_1 \neq y_2$, the family of functions $p_k(a|\boldsymbol{\theta}^k;\boldsymbol{v}^k)$ available is rich enough to regard $p_k(a|\boldsymbol{\theta}^k = \boldsymbol{y}_1;\boldsymbol{v}^k)$, $p_k(a|\boldsymbol{\theta}^k = \boldsymbol{y}_2;\boldsymbol{v}^k)$ as independent. Then one can argue as for non-adaptive policies and conclude that the gradient method will converge to a deterministic policy.

C.2 Optimization over discrete policies with continuous lockdown times. Our second heuristic to devise discrete policies for disease control requires considering lock-down policies of the following form:

$$s(t) = \begin{cases} 0 & \text{if } (t - t_0) \leq \tau_1 \text{ or } \sum_{k=1}^{2i} \tau_k \leq (t - t_0) \leq \sum_{k=1}^{2i+1} \tau_k, \text{ for some } i, \\ 1 & \text{otherwise.} \end{cases} \tag{48}$$

Here t_0 is fixed and the variables $\{\tau_k\}_{k=1}^n$ are assumed to be non-negative and to add up to $t_f - t_0$; this policy can hence be parametrized by a vector $\boldsymbol{\mu} \in \mathbb{R}^n$, with $\boldsymbol{\tau} = (t_f - t_0)\text{softmax}(\boldsymbol{\mu})$, where softmax($\boldsymbol{\mu}$) denotes the vector $\boldsymbol{v}$ with components $v_i = \frac{\exp(\mu_i)}{\sum_j \exp(\mu_j)}$. 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 Fig 15.

At time t, the disease's basic reproduction number is given by (15), with $s(t)$ defined as above. To find out $y_j^i \equiv \frac{\partial x^i(t;\boldsymbol{\mu})}{\partial \mu_j}$, we invoke Eq (29). In computing the term

$$\frac{\partial G^i}{\partial \mu_j} = \sum_k \frac{\partial G^i}{\partial \tau_k} \frac{\partial \tau_k}{\partial \mu_j}, \tag{49}$$

we have the problem that, due to (48), G is not continuous or differentiable. 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 Fig 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

$$t^- := \left(t_0 + \sum_{k=1}^u \tau_k\right) \leq t \leq \left(t_0 + \sum_{k=1}^u \tau_k + \delta\right) =: t^+. \tag{50}$$

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

$$\frac{\partial}{\partial \tau_j} \hat{s}(t) = \frac{(-1)^u}{\delta}. \tag{51}$$

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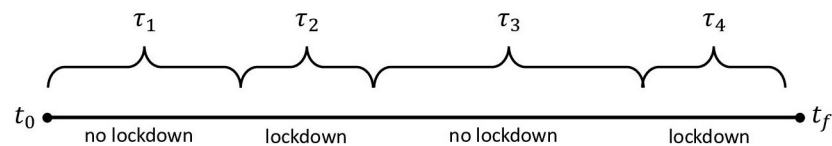

Fig 15. Lockdowns (discrete policies) parametrized with continuous time. In this example there are two lockdowns and two periods of freedom. The algorithm finds the optimal distribution and minimises the total time $\tau_2 + \tau_4$ in lockdown.
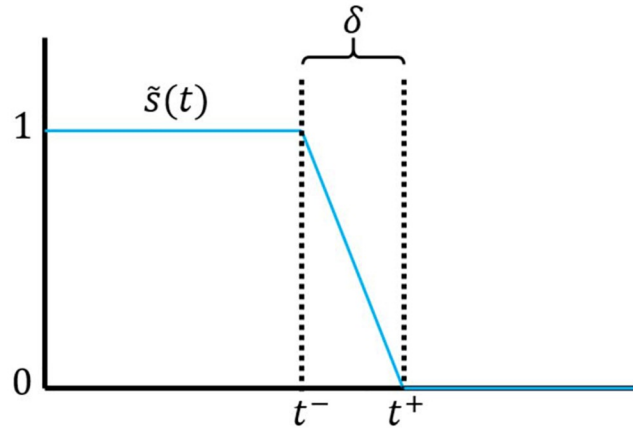
Fig 16. Modified continuous function $\tilde{s}(t)$.

right-hand side of (29) for $t \in [t^-, t^+]$ is therefore

$$
\begin{aligned}
\frac{\partial G^i}{\partial \mu_j} &= \frac{\partial G^i(t^-, \boldsymbol{x}(t^-))}{\partial \tilde{s}} \sum_{k=1}^{u} \frac{\partial \tilde{s}}{\partial \tau_k} \frac{\partial \tau_k}{\partial \mu_j} \\
&= (t_f - t_0) \frac{(-1)^u}{\delta} \frac{\partial G^i(t^-, \boldsymbol{x}(t^-))}{\partial \tilde{s}} \sum_{k=1}^{u} \frac{\partial \text{softmax}(\boldsymbol{\mu})_k}{\partial \mu_j} =: \frac{K_j^i(t^-, \boldsymbol{x}(t^-), \boldsymbol{\mu}, u)}{\delta}.
\end{aligned}
\tag{52}
$$

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

1. $y_j^i(t_0) = 0$.

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

$$
y_j^i(t) \to y_j^i(t) + K_j^i(t, \boldsymbol{x}, \boldsymbol{\mu}, u).
\tag{53}
$$

3. For all other values of t, y_j^i continuously evolves via the equation

$$
\frac{dy_j^i}{dt} = \sum_{l} \frac{\partial G^i(t, \boldsymbol{x}; \boldsymbol{\mu})}{\partial x^l} y_j^l; i = 1, ..., m, j = 1, ..., n.
\tag{54}
$$

D Continuous-space population models

Even though the focus of this article is that of compartmental models of the form (26), 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 [26], 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

$$
\frac{\partial}{\partial t} \boldsymbol{u} - \text{div}(\boldsymbol{v} \nabla \boldsymbol{u}) = A(\boldsymbol{u})\boldsymbol{u}.
\tag{55}
$$

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. v, 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

$$\nabla \boldsymbol{u}(t, X, Y) \cdot \boldsymbol{n}(X, Y) = 0, \ \text{ for } (X, Y) \in \partial\Omega, \tag{56}$$

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 [26] solve this equation numerically via the Finite Element Method (FEM) [34].

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 $v_j^i(t, X, Y; \boldsymbol{\mu}, u_0) \equiv \frac{\partial u^i(t, X, Y; \boldsymbol{\mu}, u_0)}{\partial \mu_j}$. Let $\boldsymbol{v}_j \in \mathbb{R}^3$ be the vector with components v_j^i and differentiate both (55) and (56) with respect to μ_i. This results in the equation

$$\frac{\partial}{\partial t} \boldsymbol{v}_j - \text{div}\left(\frac{\partial \boldsymbol{v}}{\partial \mu_j} \nabla \boldsymbol{u} + v \nabla \boldsymbol{v}_j\right) = \left(\frac{\partial A}{\partial \mu_j} + \frac{\partial A}{\partial u_i} v_j^i\right) \boldsymbol{u} + A(\boldsymbol{u}) \boldsymbol{v}_j$$

$$\nabla \boldsymbol{v}_j(t, X, Y) \cdot \boldsymbol{n} = 0, \ \text{ for } (X, Y) \in \partial\Omega. \tag{57}$$

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 [26] solved Eq (55), that is, via the FEM.

Acknowledgments

Author Contributions

Conceptualization: Miguel Navascués, Costantino Budroni, Yelena Guryanova.

Formal analysis: Miguel Navascués, Costantino Budroni, Yelena Guryanova.

Software: Miguel Navascués, Costantino Budroni, Yelena Guryanova.

References

1. Nations U. A UN framework for the immediate socio-economic response to COVID-19; 2020. Available from: https://unsdg.un.org/resources/un-framework-immediate-socio-economic-response-covid-19.

2. COVID-19 has made the health system's collapse complete in Yemen. Medecins sans Frontieres. 2020;.

3. People are dying at home amid collapsing health system in El Salvador. Medecins sans Frontieres. 2020;.

4. Pradhan TR. Government decides to lift the four-month-long coronavirus lockdown, but with conditions. The Kathmandu Post. 2020;.

5. Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton University Press; 2008. Available from: http://www.jstor.org/stable/j.ctvcm4gk0.

6. Hindes J, Bianco S, Schwartz I. Optimal periodic closure for minimizing risk in emerging disease outbreaks. PLoS ONE. 2021; 16(1 January). https://doi.org/10.1371/journal.pone.0244706 PMID: 33406106

7. Gaff H, Schaefer E. Optimal control applied to vaccination and treatment strategies for various epidemiological models. Mathematical Biosciences and Engineering. 2009; 6:469. https://doi.org/10.3934/mbe.2009.6.469 PMID: 19566121

8. de Pinho MdR, Kornienko I, Maurer H. Optimal Control of a SEIR Model with Mixed Constraints and L1 Cost. In: Moreira AP, Matos A, Veiga G, editors. CONTROLO'2014—Proceedings of the 11th Portuguese Conference on Automatic Control. Cham: Springer International Publishing; 2015. p. 135–145.

9. Obsu LL, Balcha SF. Optimal control strategies for the transmission risk of COVID-19. Journal of Biological Dynamics. 2020; 14(1):590–607. https://doi.org/10.1080/17513758.2020.1788182

10. Zamir M, Shah Z, Nadeem F, Memood A, Alrabaiah H, Kumam P. Non Pharmaceutical Interventions for Optimal Control of COVID-19. Computer Methods and Programs in Biomedicine. 2020; 196:105642. https://doi.org/10.1016/j.cmpb.2020.105642 PMID: 32688137

11. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. New England Journal of Medicine. 2020; 382(13):1199–1207. https://doi.org/10.1056/NEJMoa2001316 PMID: 31995857

12. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Eurosurveillance. 2020; 25:2000058. https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058

13. Volpicelli G. China has almost eliminated Covid-19. What can the world learn? Wired. 2020;.

14. Goodfellow IJ, Bengio Y, Courville A. Deep Learning. Cambridge, MA, USA: MIT Press; 2016. Available from: http://www.deeplearningbook.org.

15. Kissler S, Tedijanto C, Lipsitch M, Grad Y. Social distancing strategies for curbing the COVID-19 epidemic; 2020. Available from: https://dash.harvard.edu/handle/1/42638988.

16. Reich NG, Brooks LC, Fox SJ, Kandula S, McGowan CJ, Moore E, et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. Proceedings of the National Academy of Sciences. 2019; 116(8):3146–3154. https://doi.org/10.1073/pnas.1812594116 PMID: 30647115

17. Online repository with the code used for the simulations, optimizations, and plots presented;. Available from: https://github.com/costantinobudroni/opt-disease-control/.

18. Keeling MJ, Rohani P. Estimating spatial coupling in epidemiological systems: a mechanistic approach. Ecology Letters. 2002; 5(1):20–29. https://doi.org/10.1046/j.1461-0248.2002.00268.x

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

20. McLean AR, Blower SM. Imperfect vaccines and herd immunity to HIV. Proceedings of the Royal Society of London Series B: Biological Sciences. 1993; 253(1336):9–13. https://doi.org/10.1098/rspb.1993.0075 PMID: 8396781

21. Agur Z, Cojocaru L, Mazor G, Anderson RM, Danon YL. Pulse mass measles vaccination across age cohorts. Proceedings of the National Academy of Sciences. 1993; 90(24):11698–11702. https://doi.org/10.1073/pnas.90.24.11698 PMID: 8265611

22. Keeling MJ, Woolhouse MEJ, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, et al. Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. Science. 2001; 294(5543):813–817. https://doi.org/10.1126/science.1065973 PMID: 11679661

23. Boyd S, Xiao L, Mutapcic A. Subgradient methods. lecture notes of EE392o, Stanford University, Autumn Quarter. 2004;.

24. Duchi J. EE364b: Lecture Slides and Notes. https://webstanfordedu/class/ee364b/lectureshtml. 2018;.

25. Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press; 2000.

26. Keller JP, Gerardo-Giorda L, Veneziani A. Numerical simulation of a susceptible-exposed-infectious space-continuous model for the spread of rabies in raccoons across a realistic landscape. Journal of biological dynamics. 2013; 7 Suppl 1:31–46. https://doi.org/10.1080/17513758.2012.742578 PMID: 23157180

27. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: Bengio Y, LeCun Y, editors. 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings; 2015.Available from: http://arxiv.org/abs/1412.6980.

28. Cook SA. The Complexity of Theorem-Proving Procedures. In: Proceedings of the Third Annual ACM Symposium on Theory of Computing. STOC'71. New York, NY, USA: Association for Computing Machinery; 1971. p. 151–158. Available from: https://doi.org/10.1145/800157.805047.

29. et al F. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College. 2020;.

30. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020; 368(6493):860–868. https://doi.org/10.1126/science.abb5793 PMID: 32291278

31. Katella K. Comparing the COVID-19 Vaccines: How Are They Different?; 2021. Available from: https://www.yalemedicine.org/news/covid-19-vaccine-comparison.

32. Hochreiter S, Schmidhuber J. Long Short-term Memory. Neural computation. 1997; 9:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735 PMID: 9377276

33. Numerical Methods for Ordinary Differential Equations. John Wiley & Sons, Ltd; 2016. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119121534.fmatter.

34. Logan DL. A First Course in the Finite Element Method Using Algor. 2nd ed. USA: Brooks/Cole Publishing Co.; 2000.