Letter

Optimal control in pandemics

Joseph Samuel [1,2] and Supurna Sinha [1]
[1] *Raman Research Institute, Bangalore 560080, India*
[2] *International Center for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore 560089, India*

(Received 4 September 2020; accepted 4 January 2021; published 19 January 2021)

During a pandemic, there are conflicting demands that arise from public health and socioeconomic costs. Lockdowns are a common way of containing infections, but they adversely affect the economy. We study the question of how to minimize the socioeconomic damage of a lockdown while still containing infections. Our analysis is based on the SIR model, which we analyze using a clock set by the virus. This use of the "virus time" permits a clean mathematical formulation of our problem. We optimize the socioeconomic cost for a fixed health cost and arrive at a strategy for navigating the pandemic. This involves adjusting the level of lockdowns in a controlled manner so as to minimize the socioeconomic cost.

DOI: 10.1103/PhysRevE.103.L010301

Introduction. The COVID-19 virus presents a global threat to life and livelihoods and throws up challenges which societies across the world have to learn to deal with. Pandemics are not new, and there are mathematical models which have been developed over the years. The value of mathematical models is that they give us a simplified picture of the pandemic and let us explore the effects of different containment strategies without performing costly, and possibly fatal, social experiments. These models are the basis for a rational, science-based social response to a serious threat. While models do have their limitations, they are steadily improving with time, experience, and computational power.[1] It is imperative for us to understand the predictions of these models and compare them with data and experience.

In this Letter we consider one of the simplest models of disease spread, the SIR model [1]. Our focus here is to quantify the social cost of a pandemic within the framework of the SIR model. As a society we would like to use interventions in order to minimize the damage caused by the disease.

By far the most important interventions are medical: doctors, nurses, medical infrastructure, treatments (drugs), preventive measures (vaccines), testing, and contact tracing. When a pandemic breaks out, it takes time to develop some of these interventions. Safe vaccines and drugs take time to test and develop. If the infections get out of hand, contact tracing too becomes impractical. We are then left with nonmedical interventions, like lockdowns, which limit the spread of infection by changing the social behavior of the population. This is the focus of the present paper. Lockdowns limit the spread of disease by reducing social contact; however, they also prevent the economy from functioning normally and thus come with a socioeconomic (SE) cost. Like the health cost of a pandemic, the economic cost of a lockdown can be debilitating: lockdowns affect lives and livelihoods, cause physical and mental trauma, and even deaths.

Extreme strategies are

(i) to ignore the SE cost and impose strict lockdowns (to the grievous detriment of the economy) and

(ii) to ignore the health cost and keep the economy running normally (which results in a large human cost of suffering and death).

The SE cost and the health cost are like Scylla and Charybdis of Greek mythology. We would like to have a rational strategy of steering a course between these hazards, optimizing the extent and timing of lockdowns to minimize the total cost to society. In order to do this, we need to model these costs in mathematical terms. Before we do that we recall the SIR model for disease spread.

The SIR model divides the population into three compartments $\{S, I, R\}$, where $\{S, I, R\}$ are respectively the fractions of susceptible, infected, and removed populations. The removed population includes recoveries as well as deaths. The model assumes that the recovered population is immune to the disease, that there is no possibility of reinfection. The progress of the disease is described by a set of three ordinary differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta(t)IS\\ \frac{dI}{dt} &= \beta(t)IS - \gamma I \\ \frac{dR}{dt} &= \gamma I,\end{aligned} \tag{1}$$

where we allow for the possibility that β varies with time, as would happen when lockdowns are imposed and relaxed. Evidently,

$$S + I + R = 1, \tag{2}$$

[1] It is worth emphasizing that models for weather prediction which were unreliable a few decades ago have now come of age and give us reliable predictions of the path a cyclone will take.

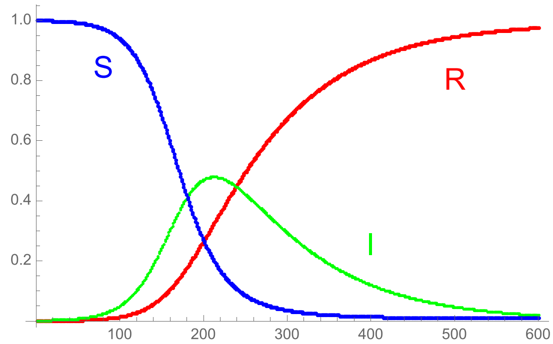

FIG. 1. The figure shows the susceptible (decreasing, blue), infected (nonmonotonic, green), and removed (increasing, red) fractions as a function of time in days. In this graph, for illustration we have taken $\beta = 0.5$ and $\gamma = 0.1$. This corresponds to a reproductive ratio of $r = 5$.

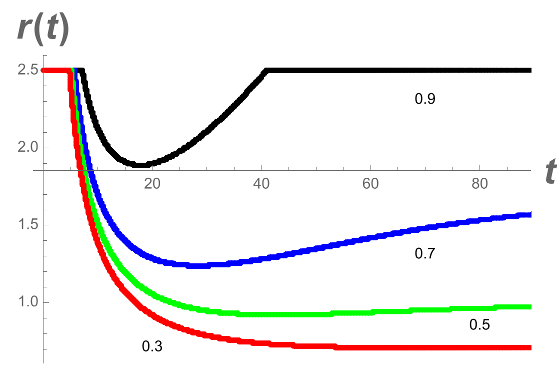

FIG. 2. Optimal lockdown profiles. Here we plot the reproduction ratio $r(t) = \beta(t)/\gamma$ as a function of human time. The different curves correspond to $\tau_f = 0.3, 0.5, 0.7, 0.9$ from bottom to top, corresponding to the colors red, green, blue, and black, respectively. $\tilde{\beta}$ and τ_0 (see main text) have been set to 0.25 and 0.0, respectively, and γ to 0.1.

and the equations of the SIR model (1) maintain this condition. The progress of the disease can be described by a point in a two-dimensional space, a plane (2) in the three-dimensional $\{S, I, R\}$ space.

The parameter β describes the rate at which the susceptible population becomes infected due to contact with the infected population. This parameter depends on how infectious the disease is, as well as the degree of contact between people. The effective infectivity $\beta(t)$ is a product $\beta(t) = u(t)\beta_0$ [2] of the biological infectivity β_0 (which depends on the disease) and $u(t)$, the degree of social contact between people. $\beta(t)$ can be controlled by reducing social contact $u(t)$, for example, by using lockdowns to ensure physical distancing and using masks. γ is the rate at which infected individuals either recover or die from the infection. Early detection and good medical care can increase the recovery rate. β and γ, which appear in the equations (1), are parameters of the model which are both positive. γ is assumed to be constant in time. The model is characterized essentially by one parameter, the reproduction ratio $r = \frac{\beta}{\gamma}$. The independent time variable t can be rescaled to set γ to 1.

The SIR model describes the evolution of the disease in a fixed population and is one of the simplest models capturing the essential features of disease spread. More detailed compartmental models have also been studied. Among these are the SEIR model and its variants [3,4], which have more compartments to allow for asymptomatic infections, etc. There is also a study [5] which questions the effectiveness of lockdowns in preventing fatalities. A suggestion for mitigating the SE cost of lockdowns has been made in Ref. [6]. The independent variable in the SIR model is the time t measured, say in days, and there are three dependent variables $\{S, I, R\}$ subject to a single constraint (2). Figure 1 shows the evolution of the SIR fractions as a function of time.

This paper is organized as follows. We first summarize our main results and then present a derivation of them, and finally, we end with some concluding remarks.

Main results. Here we summarize the main results of the study and describe the methods we use. Our objective is to minimize the damage caused by the pandemic on two fronts: the public health perspective and the economy. The demands of public health force us to impose lockdowns, which adversely affect the economy. The question of interest is when and how much to lock down so that the damage to the economy is minimized. The question is complicated by the fact that lockdown measures taken at a certain time can influence the infection rates at later times. To understand this influence requires the use of a model for the spread of infections. We work with the simplest SIR model. In order to gain a long-term perspective, we have to consider the entire duration of the pandemic and account for the integrated health and SE costs.

This is precisely the kind of problem which can be dealt with using the calculus of variations. To give a familiar example, the shape of a soap bubble is determined by the requirement that its surface area is a minimum, subject to the constraint that the volume of enclosed air is fixed. The shape which achieves this optimization is the sphere. In the case of the pandemic, the role of the "shape" is played by the profile of lockdown characterized by β as a function of time, which tells when and how much to lock down. The role of the "area" is played by the total integrated SE cost of the lockdown. The role of the fixed volume of the soap bubble is played by the total health cost, measured by the fraction of people affected over the duration of the pandemic. What emerges from this study is the optimal profile for $\beta(t)$, i.e., that which minimizes the SE cost for a fixed health cost. This is the analog of the spherical shape of a soap bubble.

A crucial ingredient in our study is the use of a new time variable. The time variable in the original SIR equations (1) is human time. Human time is counted in days or weeks and measured by the progress of stars in the sky. In contrast, virus time $\tau = R$ is measured by the progress of the virus through the population. The virus clock starts ticking at the beginning of the pandemic ($\tau = 0$, when $R = 0$), runs faster when there are more infections ($\frac{d\tau}{dt} = \gamma I$), and ceases to tick when the infections die out at the end of the pandemic.

The results of our study are presented in Fig. 2, which shows the optimal lockdown as a function of human time. These curves are the main results of this study. They represent the optimal way to modulate the lockdown so that the impact on the economy is minimal. Each of these curves represents a different fixed health cost in terms of the number of people affected by the virus.

Next we discuss the main results of this paper. We express the health and SE cost in mathematical form.

Socioeconomic cost. The parameter in the SIR model which represents the effect of lockdown is β. We suppose that when all measures which do not affect the economy (like wearing masks, washing hands) have been imposed, we have $\beta = \tilde{\beta}$. Further reduction in β can only come at an SE cost. The SE cost is a function that decreases with increasing β until $\tilde{\beta}$ and then drops down to zero. We make a simple choice of this function: for β values less than $\tilde{\beta}$, the SE cost of a lockdown is inversely proportional to β and directly proportional to the number of days it lasts. We emphasize that we are not interested in a detailed modeling of the economy. We only wish to describe the *damage* caused to the economy by the intensity and duration of the lockdown.

The total SE cost integrated over the duration of the pandemic is modeled as[2]

$$\mathcal{C}_E = \int_0^\infty \frac{dt}{\beta(t)}. \tag{3}$$

Note that the cost depends on the extent of the lockdown as well as the duration measured in human time t. Values of β above $\tilde{\beta}$ can be ignored as they come with no SE cost. If controlling the pandemic does not require β less than $\tilde{\beta}$, there is no conflict between the economic and public health objectives: the economy can function normally. Below we assume that we are always dealing with β values less than $\tilde{\beta}$, i.e., there *is* a conflict between the twin objectives.

Health cost. We model the health cost as $R_f = R(\infty)$, the total fraction of people affected by the disease during the entire course of the epidemic. Hospitalizations and deaths are some fixed fractions of R_f. Even some of those who do not need hospitalization suffer long term after effects from the ravages of COVID-19. We can therefore model the health cost mathematically as R_f, the final value of the removed fraction: $\mathcal{C}_H = R_f$. $\mathcal{C}_H$ is a dimensionless number. Our objective is to hold $\mathcal{C}_H$ fixed at the value $\mathcal{C}_{H0}$ and find the lockdown profile $\beta(t)$ which minimizes the SE cost. The fixed value $\mathcal{C}_{H0}$ of the health cost is a choice one has to make. Needless to say, there is a value judgment involved in making this choice. Choosing a small value for $\mathcal{C}_{H0}$ gives more weight to the health cost and a large value reverses the emphasis. Once this value judgment is made, we can use our ability to modulate β over time, varying the extent and timing of lockdowns to minimize the SE cost.

The independent variable t in the SIR equations (1) is the time measured in human time, for instance, days. This is relevant to the progress of the epidemic in human terms. In fact, the SE cost (3) grows with the duration of a lockdown, measured in human time. We find it advantageous to use a new time variable as set by the progress of the virus through the human population. Accordingly, we set $\tau = R$ as the fraction of people affected by the epidemic and regard this to be the "virus time." We will use τ and R interchangeably, preferring τ when we wish to emphasise its role as a "time" or independent variable. The virus time increases monotonically,

$$\frac{d\tau}{dt} = \frac{dR}{dt} = \gamma I \geqslant 0, \tag{4}$$

with human time, the rate of progress given by $\gamma I > 0$, which is proportional to the current infected fraction. As we will see, using the virus time instead of the human time gives us significant advantages in addressing our problem. First, it gives us an exact parametric solution of the SIR model. (This is equivalent to the parametric solutions given earlier by [7,8]). Second, we get a clean mathematical formulation of our problem of optimizing the total social cost. Letting the virus set the clock is one of the crucial ingredients of our approach.

Dividing the first of the equations [Eq. (1)] by the last, we find that (expressing β as a function of τ)

$$\frac{dS}{d\tau} = \frac{dS}{dR} = -\frac{\beta(\tau)}{\gamma} S, \tag{5}$$

which is readily integrated. Using the constraint (2) immediately gives us an exact solution of the SIR model in parametric form:

$$\begin{aligned} S(\tau) &= S_0 \exp\left[-\frac{\int_{\tau_0}^{\tau} \beta(\tau')d\tau'}{\gamma}\right] \\ I(\tau) &= 1 - S_0 \exp\left[-\frac{\int_{\tau_0}^{\tau} \beta(\tau')d\tau'}{\gamma}\right] - \tau \\ R(\tau) &= \tau, \end{aligned} \tag{6}$$

where the last equation is a tautology arising from the definition of τ. The virus time τ is measured from the beginning of the epidemic. τ_0 represents any fixed intermediate time. The relation between the virus time and human time is given by integrating the last of (1), where $I(\tau)$ is given by the second equation of (6):

$$t = \int_0^\tau \frac{d\tau'}{\gamma I(\tau')}. \tag{7}$$

Let us suppose that our lockdown response starts when the virus time is τ_0, when the susceptible fraction is S_0. τ_0 could be when the pandemic is initially detected or any subsequent time. Given a fixed value of the health cost R_f, our problem is to choose the function $\beta(\tau)$ so as to minimize the SE cost.

Let us introduce a new variable,

$$y(\tau) = \int_{\tau_0}^{\tau} \beta(\tau')d\tau', \tag{8}$$

so that $S(\tau) = S_0 \exp[-y(\tau)/\gamma]$ and $\beta(\tau) = \frac{dy}{d\tau} = \dot{y}$. Then $S_f = S(\tau_f) = S_0 \exp[-y(\tau_f)/\gamma]$ and at the end of the pandemic, when infections cease ($I = 0$), we have from $S_f + R_f = 1$,

$$S_0 \exp[-y(\tau_f)/\gamma] + \tau_f = 1, \tag{9}$$

fixing y_f in terms of τ_f,

$$y_f = y(\tau_f) = -\gamma \log(1 - \tau_f)/S_0. \tag{10}$$

[2]We could write this equation more correctly by multiplying the integrand by the Heaviside Theta function $\Theta(\tilde{\beta} - \beta)$. We have not done so, as this may not be familiar to some readers.

We now have a classic variational problem for $y(\tau)$, where $y(\tau_0) = 0$, $y(\tau_f) = y_f$ are held fixed and we have to minimize the SE cost,

$$\mathcal{C}_E = \int_{\tau_0}^{\tau_f} \frac{dt}{d\tau}\frac{d\tau}{\dot{y}}. \tag{11}$$

From the parametric solution to the SIR equations (1), we replace $\frac{dt}{d\tau}$ by $[\gamma I(\tau)]^{-1}$, leading to the variational problem of minimizing

$$\int_{\tau_0}^{\tau_f} \frac{d\tau}{I[y(\tau)]\dot{y}(\tau)}, \tag{12}$$

where we have dropped some constants. $I[y(\tau)]$ here is a functional of $y(\tau)$, which is found by solving the SIR equations (1). Its explicit form is given by the second of (6). We now have to minimize,

$$\int_{\tau_0}^{\tau_f} \frac{d\tau}{\dot{y}(\tau)(1 - S_0 \exp{[-y/\gamma]} - \tau)}. \tag{13}$$

We now vary $y(\tau)$ in Eq. (13) and as is usual in the calculus of variations, integrate by parts and discard the boundary terms, since y is held fixed at both boundaries.

We can read off the Lagrangian appearing in Eq. (13):

$$L(y, \dot{y}, \tau) = \frac{1}{\dot{y}(1 - S_0 \exp{[-y/\gamma]} - \tau)}, \tag{14}$$

and the resulting Euler-Lagrange equations can be rearranged to give

$$\ddot{y}(1 - S_0 \exp{[-y]}/\gamma - \tau) + \frac{\dot{y}^2 S_0 \exp{[-y/\gamma]}}{\gamma} - \frac{\dot{y}}{2} = 0. \tag{15}$$

This equation can be integrated by expressing it as

$$\frac{dK(y, \dot{y}, \tau)}{d\tau} = 0, \tag{16}$$

where the constant of the motion K has the form

$$K(y, \dot{y}, \tau) = \dot{y}(1 - S_0 \exp[-y/\gamma] - \tau) + y/2 = K_0. \tag{17}$$

These equations can be analytically solved by introducing an integrating factor $(y/2 - K)^{-3}$. The solution gives τ as a function of y expressed in terms of elementary functions including the exponential integral, which can be plotted to show y as a function of τ. From this it is easy to extract the quantity of interest, $\beta(\tau) = \dot{y}$, which determines the lockdown profile.

In making Fig. 2, we have numerically integrated (17) in the form

$$\frac{dy}{d\tau} = \frac{K_0 - y/2}{(1 - S_0 \exp[-y/\gamma] - \tau)} \tag{18}$$

and noted that our boundary conditions imply that the value of the constant is $K_0 = y_f/2$.

Figure 2 shows the optimal way of imposing lockdowns, plotting $r = \beta/\gamma$ as a function of human time. The optimal solution consists of an initial sharp lockdown followed by a gradual release of the lockdown. Intuitively, this is easy to understand: premature release of lockdown results in flareups of the disease, which then require further lockdowns which contribute to the SE cost.

Concluding remarks. First, a disclaimer: The authors of this paper are not epidemiologists. We are theoretical physicists who have addressed a socially relevant interdisciplinary problem, which can be addressed using the methods of our subject. As is common in theoretical physics, we work with the simplest model that captures the phenomena of interest. The conclusions we arrive at are not intended to be *directly* transferred to any real world context. Nevertheless, the ideas developed here can be developed further by introducing more realistic models. The main message we have to offer is that there is a competition between the twin social objectives of public health and the economy. It is then advantageous to use lockdown profiles derived from our formalism to minimize the total damage from a pandemic. For example, the SEIR model is a slightly more realistic model in which our analysis can be carried out. In this case, a purely analytic solution is not possible, but one can formulate the problem as we have done here and derive results for the optimal lockdown using numerical methods.

One of the main ideas introduced and used in this Letter is that of "virus time." At one level it is a convenient mathematical device. It leads to an exact parametric solution of the SIR model. This solution is considerably simpler than the existing ones [7]. At a conceptual level it is a more appropriate measure of the progress of the disease than human time measured in days. For example, virus time elapses differently in different countries. Some countries impose travel restrictions (e.g., travel within 5 km of one's house). In such cases virus time elapses differently in different locations.

Our graphs show the optimal lockdown profile in terms of the reproductive ratio β/γ, a quantity which is directly measurable (by testing a random sample of the population) and often used to describe the progress of a pandemic. One could consider more complicated functional dependence on β for the SE cost, for instance, one could consider the SE cost per day to be inversely proportional to β^2 and so on. We expect the main qualitative conclusions to remain unchanged by such a choice. It would take economists to realistically measure the cost of lockdowns. This is a task we do not undertake here.

Across the world there have been many disparate government responses to the COVID-19 pandemic. Sweden chose not to lock down at all; New Zealand chose to lock down hard and early. Many other nations adopted policies in between, some responding to flaring infections as they broke out, as firemen do with fires. After the pandemic is over, with the clarity of hindsight we will learn which of these strategies was most effective in preventing societal distress. Meanwhile we can gain insights by working with simple models to evaluate these different strategies. The work of this paper is a starting point in considering the SE as well as health costs of a pandemic. We import methods from variational calculus (which we illustrate by the example of soap bubbles) to arrive at an optimal strategy to navigate the pandemic, keeping both the SE and health costs in mind. A crucial ingredient in our problem is the virus time, a natural parameter which has been introduced to make the problem tractable. Our main result is that the best strategy to follow is one in which a sharp lockdown is imposed, followed by a gradual release. This is indeed to be expected, if one understands exponential growth of disease. However, our solution also prescribes when to lock down and how much. An important choice to be made in determining the strategy is the value of R_f, the health cost. This choice will depend on the resources of the nation. Nations which can afford a larger cost can opt for a lower value of R_f. Poorer nations will be forced to accept a higher health cost. However, given these limitations, the strategy we propose is optimal within the SIR model.

Acknowledgments. It is a pleasure to thank Abhishek Dhar, Suvrat Raju, Biman Nath, and Michael Berry for their comments on the manuscript. We thank the referees for their critical remarks, which have helped us improve the manuscript.

[1] W. Kermack and A. G. McKendrick, Proc. R. Soc. London, Ser. A **115**, 700 (1927).

[2] A. Dhar, arXiv:2004.05373.

[3] A. Das, A. Dhar, S. Goyal, A. Kundu, and S. Pandey, Chaos, Solitons & Fractals 110595 (2021), doi: 10.1016/j.chaos.2020.110595.

[4] B. F. Maier and D. Brockmann, Science **368**, 742 (2020).

[5] S. Raju, arXiv:2007.04219.

[6] M. Dror, N. Schulamann, R. Cohen, S. Haber, E. Yaniv, R. Sarid, and B. Barzel, arXiv:2004.01453.

[7] T. Harko, F. S. Lobo, and M. Mak, Appl. Math. Comput. **236**, 184 (2014).

[8] J. Miller, Bull. Math. Biol. **74**, 2125 (2012).