Resetting in stochastic optimal control

Benjamin De Bruyne● and Francesco Mori●

LPTMS, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France

"When in a difficult situation, it is sometimes better to give up and start all over again." While this empirical truth has been regularly observed in a wide range of circumstances, quantifying the effectiveness of such a heuristic strategy remains an open challenge. In this paper, we combine the notions of optimal control and stochastic resetting to address this problem. The emerging analytical framework allows one not only to measure the performance of a given restarting policy, but also to obtain the optimal strategy for a wide class of dynamical systems. We apply our technique to a system with a final reward and show that the reward value must be larger than a critical threshold for resetting to be effective. Our approach, analogous to the celebrated Hamilton-Jacobi-Bellman paradigm, provides the basis for the investigation of realistic restarting strategies across disciplines. As an application, we show that the framework can be applied to an epidemic model to predict the optimal lockdown policy.

Finding the optimal strategy to operate a complex system is a longstanding problem and has attracted a lot of attention in recent decades. Since the seminal works of Pontryagin [1] and Bellman [2], optimal control theory has received renewed interest due to its applications in a wide range of contexts, such as artificial intelligence [3] and finance [4]. In a typical setting, optimal control considers a system whose state at time t can be represented by a d-dimensional vector $\mathbf{x}(t)$. For instance, the state $\mathbf{x}(t)$ could correspond to the degrees of freedom of an autonomous robot or the asset values in a financial portfolio. The system typically evolves in time following a deterministic law, e.g., the laws of motion for mechanical systems or the law of supply and demand for financial markets. Oftentimes, the mathematical modeling of these laws is prohibitively expensive and one introduces a stochastic contribution to account for the missing information on the environment. Given the laws of motion, optimal control aims to operate the system in the best possible way by using an external control, e.g., actuators for robots or market orders in finance.

One of the simplest ways to describe analytically the evolution in time of the system $\mathbf{x}(t)$ is a first-order differential equation of the form $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x},t)$. This law is often a simplified description of the system and a source of Gaussian white noise $\boldsymbol{\eta}(t)$ is introduced to capture the fluctuations around the deterministic model. In addition, the external control on the system is usually modeled as a drift $\mathbf{u}(\mathbf{x},t)$. Summing up these contributions, the full mathematical description of the system is given by

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x},t) + \sqrt{2D}\,\boldsymbol{\eta}(t) + \mathbf{u}(\mathbf{x},t), \qquad (1)$$

where $\sqrt{2D}$ is the strength of the noise. The external control $\mathbf{u}(\mathbf{x},t)$ can be tuned to achieve a given goal, e.g., performing a task for a robot or generating profits in finance. Of course, controlling the system will generate operating costs, such as electrical consumption or transaction fees. Optimally controlling the system requires balancing a trade-off between high rewards, measured over time by a function $R(\mathbf{x},t)$, and low operating costs, often taken to be proportional to $\mathbf{u}^2(\mathbf{x},t)$. To be precise, for a system located at position $\mathbf{x}$ at time t, the reward in a small time interval dt is $R(\mathbf{x},t)dt$ and the cost is $\mathbf{u}^2(\mathbf{x},t)dt/2$.

In principle, solving this optimization problem is intrinsically difficult due to the high dimensionality of the space of solutions. Remarkably, Bellman introduced a general way to solve this problem, known as dynamical programming, which consists in breaking down the optimization into simpler subproblems in a recursive manner such that the present action is taken to maximize the future outcome. In doing so, the key quantity to keep track of is the optimal payoff $J(\mathbf{x},t)$, defined as the expected payoff for an optimally controlled system located at $\mathbf{x}$ at time t. Using this framework, one can show that the optimal control is simply given by $\mathbf{u}^*(\mathbf{x},t) = \nabla_{\mathbf{x}}J(\mathbf{x},t)$, driving the system towards the direction in which the payoff increases the most. The optimal payoff $J(\mathbf{x},t)$ satisfies the celebrated Hamilton-Jacobi-Bellman (HJB) equation [5],

$$-\partial_t J = D\Delta_{\mathbf{x}}J + \mathbf{f}\cdot\nabla_{\mathbf{x}}J + \tfrac{1}{2}(\nabla_{\mathbf{x}}J)^2 + R, \qquad (2)$$

where $\Delta_{\mathbf{x}}$ and $\nabla_{\mathbf{x}}$ are, respectively, the Laplacian and the gradient operators. Here, for convenience, we dropped the $\mathbf{x}$ and t dependence in the functions in Eq. (2). The quadratic term $(\nabla_{\mathbf{x}}J)^2$ renders this equation nonlinear and difficult to solve for arbitrary reward functions. Nevertheless, there exist few analytically solvable cases. For instance, in the case of $d=1$, where $f(x,t)=0$ and $R(x,t)=-\alpha(x-x_f)^2\delta(t-t_f)/2$, the optimal control has the simple form $u^*(x,t) = -\alpha(x-x_f)/[1+\alpha(t_f-t)]$, which, in the limit $\alpha\to\infty$, is reminiscent of the effective force to generate bridge

Brownian motion [6]. This optimal control continuously drives the system to maximize the final reward by arriving close to the target x_f at time t_f. In more realistic systems, one has to rely on numerical methods to solve Eq. (2) [7].

Ideas from optimal control have also been proven successful in different areas of physics [8–10]. Moreover, stochastic optimal control has been applied to a variety of systems, such as supply-chain planning [11], swarms of drones [12], and fluctuating interfaces [13]. These systems all have in common that the optimal control can be orchestrated as a coordinated sequence of infinitesimal local changes. However, numerous systems do not fall in this class and require global changes to be optimally managed. A timely example of such circumstances is the Covid-19 crisis, during which the main control policies have been global measures such as regional and national lockdowns. Other instances arise in the context of search processes, both in the cases of computer algorithms [14] and time-sensitive rescue missions [15]. In the latter situations, a common and widely observed strategy is to stop and resume the operations from a predefined location. Such situations particularly arise in the context of search processes [16,17], chemical reactions [18], and catastrophes in population dynamics [19,20]. Unfortunately, the HJB framework is not well suited to study such resetting protocols. Indeed, resetting is known to be quite different from a local force and exhibits interesting features, including out-of-equilibrium dynamics [21–23], dynamical phase transitions [24–26], and nontrivial first-passage properties [16,17] (for a recent review, see [27]). This observation naturally called into question the existence of an analytical framework to devise the optimal resetting control policy.

In this paper, we combine the notions of stochastic resetting and optimal control into *resetting optimal control*, which provides a natural framework to operate systems through stochastic restarts. Our goal is not to provide an accurate description of a specific system, but rather to consider a minimal model to explore resetting in optimal control. To model resetting policies, we exchange the control force $\boldsymbol{u}(\boldsymbol{x}, t)$ for a resetting rate $r(\boldsymbol{x}, t)$. In a small time interval dt, the state of the system is reset to a predefined location $\boldsymbol{x}_{\text{res}}$ with probability $p = r(\boldsymbol{x}, t)dt$ and evolves freely otherwise. In sum, the dynamical system evolves according to

$$\boldsymbol{x}(t + dt) = \begin{cases} \boldsymbol{x}_{\text{res}}, & \text{prob.} = p \\ \boldsymbol{x}(t) + \boldsymbol{f}(\boldsymbol{x}, t)dt + \sqrt{2D}\boldsymbol{\eta}dt, & \text{prob.} = 1 - p, \end{cases}$$
(3)

where the subscript "res" in $\boldsymbol{x}_{\text{res}}$ stands for "resetting location." Similarly to the HJB framework, we aim to find the optimal resetting rate r, as a function of $\boldsymbol{x}$ and t, that balances the trade-off between high rewards, measured over time by the function $R(\boldsymbol{x}, t)$, and low operating costs, which depend on $r(\boldsymbol{x}, t)$. To mathematically pose the optimization problem, we naturally extend the HJB paradigm and define the following payoff functional:

$$\mathcal{F}_{\boldsymbol{x_0}, t}[r] = \left\langle \int_t^{t_f} d\tau [R(\boldsymbol{x}(\tau), \tau) - c(\boldsymbol{x}(\tau), \tau)r(\boldsymbol{x}(\tau), \tau)] \right\rangle_{\boldsymbol{x_0}},$$
(4)

where $c(\boldsymbol{x}, \tau)$ is the cost associated to resetting, t_f is the time horizon up to which the system is controlled, and the

symbol $\langle \cdot \rangle_{\boldsymbol{x_0}}$ indicates the average over all stochastic trajectories starting from $\boldsymbol{x_0}$ at time t and evolving according to Eq. (3). Note that the payoff $\mathcal{F}$ is a functional and depends on the full function r.

Remarkably, we find that the optimal resetting policy r^* that maximizes $\mathcal{F}$ is *bang-bang*, resulting in an impulsive control strategy [5]: the system is reset with probability one if its state is outside of a time-dependent domain $\Omega(t)$ and evolves freely otherwise:

$$r^*(\boldsymbol{x}, t)dt = \begin{cases} 0 & \text{if} \quad \boldsymbol{x} \in \Omega(t), \\ 1 & \text{if} \quad \boldsymbol{x} \notin \Omega(t). \end{cases}$$
(5)

The domain $\Omega(t)$ evolves according to

$$\Omega(t) = \{\boldsymbol{x} : J(\boldsymbol{x}, t) \geqslant J(\boldsymbol{x}_{\text{res}}, t) - c(\boldsymbol{x}, t)\},$$
(6)

where the optimal payoff function $J(\boldsymbol{x}, t) = \max_r \mathcal{F}_{\boldsymbol{x}, t}[r]$ is the solution of the differential equation

$$-\partial_t J = D\Delta_{\boldsymbol{x}} J + \boldsymbol{f} \cdot \nabla_{\boldsymbol{x}} J + R, \quad \boldsymbol{x} \in \Omega(t).$$
(7)

Equation (7) must be solved jointly with Eq. (6) starting from the final condition $J(\boldsymbol{x}, t_f) = 0$ and evolving backward in time with the Neumann boundary condition $\nabla_{\boldsymbol{x}} J(\boldsymbol{x}, t) \cdot \boldsymbol{n}(\boldsymbol{x}) = 0$, where $\boldsymbol{n}(\boldsymbol{x})$ is the normal unit vector to the boundary. Outside of the domain $\Omega(t)$, the solution is given by $J(\boldsymbol{x}, t) = J(\boldsymbol{x}_{\text{res}}, t) - c(\boldsymbol{x}, t)$. The definition of the domain $\Omega(t)$ in Eq. (6) has a clear interpretation: at any given time, the optimal policy is to restart the system if its expected payoff is less than the one at the resetting location minus the cost incurred for a restart. The emerging framework outlined in Eqs. (5)–(7) is the main result of this paper and provides a general method to optimally control stochastic systems through restarts. The derivation of this result is presented in Appendix A.

Going from Eq. (4) to Eq. (7), we have reduced a functional optimization problem to a partial differential equation, which is often easier to solve. Note, however, that the mathematical problem in Eqs. (6) and (7) is of a special kind as the evolution of the domain of definition $\Omega(t)$ is coupled to the solution $J(\boldsymbol{x}, t)$ of the differential equation. This kind of equations belongs to a class known as Stefan problems [28], which often arise in the field of heat transfer, where one studies the evolution of an interface between two phases, e.g., ice and water on a freezing lake. In this context, one must solve the heat equation for the temperature profile with an interface between the water and ice phases, which moves according to the temperature gradient. The interface is therefore to be considered as an additional unknown function, which must be jointly solved with the differential equation. To draw an analogy with our case, the optimal payoff function $J(\boldsymbol{x}, t)$ plays the role of the temperature and the boundary of the domain $\Omega(t)$ corresponds to the water-ice interface. Note, however, that the two Stefan problems have different boundary conditions. The domain $\Omega(t)$ can be obtained by solving the Stefan problem in Eqs. (6) and (7) numerically. This can be achieved, for instance, by using an Euler explicit scheme with a space-time discretization and updating the domain $\Omega(t)$ at each time step according to Eq. (6). The domain $\Omega(t)$ is illustrated in Fig. 1 for the case of a one-dimensional random walk with a final reward. Such a situation corresponds to rewarding the system by a finite amount for arriving at some target location at the final time, while penalizing it with a unit cost for each
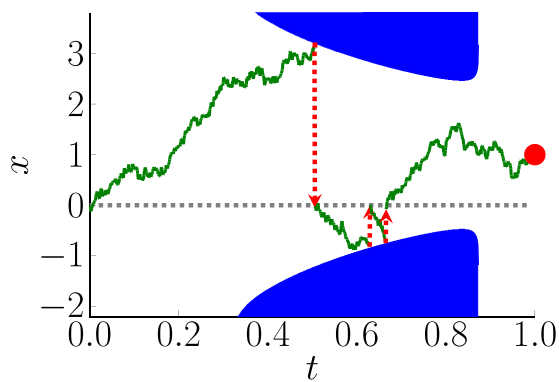
FIG. 1. Space-time illustration of the optimal resetting policy for a one-dimensional random walk (green line) to reach the target location $x_f = 1$ (filled circle) exactly at time $t_f = 1$. A reward α is received upon reaching the location x_f at the final time, while a unit cost is incurred upon resetting (red dashed arrows) to the origin $x_{\rm res} = 0$ (gray dashed line). The optimal strategy that maximizes the expected payoff is to reset upon touching the blue region and to evolve freely in the white region, which we denote as $\Omega(t)$ in the text. The domain $\Omega(t)$ is obtained by numerical integration of the Stefan problem in Eqs. (6) and (7), with $R(x, t) = \alpha\delta(t - 1)\delta(x - 1)$, $f(x, t) = 0$, $\alpha = 10$, and $D = 1$. The boundary of the domain $\Omega(t)$ guides the particle to the location x_f at time t_f, while avoiding resetting as much as possible. The shape of $\Omega(t)$ depends nontrivially on the reward value α. Further explanations for this shape are given in the text.

resetting event. This setting is further discussed in the next paragraph, where we illustrate our framework with various examples.

The Stefan problem in Eq. (7) is the analog of the HJB equation (2) for a resetting control. Despite the moving boundary, Eqs. (6) and (7) have a linear dependence in J. One might, therefore, wonder if exactly solvable models exist within this framework. Interestingly, we have found a time-independent formulation in the infinite time horizon limit $t_f \to \infty$ which allows for exact analytical solutions to be found. This is achieved by considering discounted rewards and costs of the form $R(x, t) = e^{-\beta t}\mathcal{R}(x)$ and $c(x, t) = e^{-\beta t}\mathcal{C}(x)$, where $\beta > 0$ is the discount rate. Accordingly, we also consider the drift to be time independent, $f(x, t) = f(x)$. Discounted payoffs are common in the control theory literature [29] and capture situations in which the effect of the payoff decays over a typical timescale $1/\beta$. Such effect is, for instance, observed in financial contexts, where β is related to interest rates and is used to compare future and present benefits. Using the ansatz $J(x, t) = e^{-\beta t}\mathcal{J}(x)$, we find that Eq. (7) becomes a time-independent ordinary differential equation of the form

$$\beta\mathcal{J} = D\Delta_x\mathcal{J} + f \cdot \nabla_x\mathcal{J} + \mathcal{R}, \quad x \in \Omega, \tag{8}$$

where the domain $\Omega = \{x : \mathcal{J}(x) \geqslant \mathcal{J}(x_{\rm res}) - \mathcal{C}(x)\}$ is also independent of time. This equation can be explicitly solved in the absence of external forces by choosing a quadratic reward $\mathcal{R}(x) = -\alpha x^2$ and a constant resetting cost $\mathcal{C}(x) = c$ to the origin $x_{\rm res} = 0$. Note that $\mathcal{R}(x) \leqslant 0$ and is maximized at $x = 0$, rewarding the system for being close to the origin.
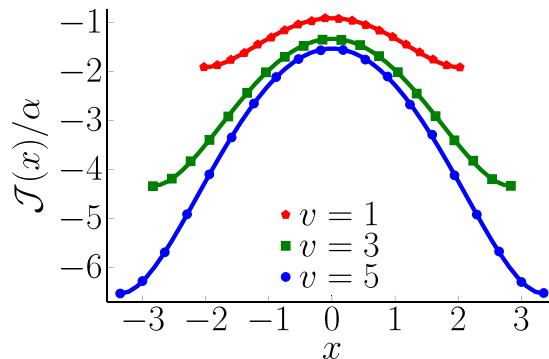

FIG. 2. The optimal expected payoff $\mathcal{J}(x)$ for a one-dimensional random walk starting at x for the discounted reward $\mathcal{R}(x) = -\alpha x^2$ and cost $\mathcal{C}(x) = c$ and for different values of the cost-reward ratio $v = c/\alpha$. The continuous lines correspond to the exact result in Eq. (9), while the symbols correspond to numerical simulations performed with $\beta = D = 1$. The optimal strategy is to reset for $|x| > u(v)$, where $u(v)$ is defined in the text. For $|x| < u(v)$, the system is let free to evolve and its payoff depends continuously on x.

Solving Eq. (8), we obtain, for $\beta = D = 1$, the exact expression for the optimal payoff,

$$\mathcal{J}(x) = \alpha[-2 - x^2 + 2u(v)\cosh(x)/\sinh(u(v))], \quad x \in \Omega, \tag{9}$$

where $u(v)$ is the boundary of the symmetric domain Ω, i.e.,

$$\Omega = \{x : |x| < u(v)\}. \tag{10}$$

The boundary $u(v)$ is the unique positive solution of the transcendental equation $v - u^2(v) + 2u(v)\tanh[u(v)/2] = 0$, where $v = c/\alpha$ is the cost-reward ratio. The optimal strategy thus corresponds to resetting the system if $|x| > u(v)$. When $v \ll 1$, the cost of resetting is much smaller than the reward, therefore the boundary is close to the origin $u(v) \sim \sqrt{2}(3v)^{1/4}$, allowing the state of the system to remain close to the optimal location $x = 0$. On the other hand, when $v \gg 1$, the cost of resetting is much larger than the running cost and the boundary is set far away from the origin, $u(v) \sim \sqrt{v} + 1$. Beyond the optimal resetting policy, our approach predicts the function $\mathcal{J}(x)$, measuring the expected payoff upon starting from x and following the optimal strategy. In Fig. 2, $\mathcal{J}(x)$ is shown for $|x| < u(v)$ and for various values of the cost-reward ratio v. As a function of x, $\mathcal{J}(x)$ has a symmetric bell shape centered around the origin, where the reward is maximal. As $|x|$ increases, $\mathcal{J}(x)$ decreases since the reward decreases and the resetting boundary comes closer. Note that in this special case, the optimal policy can also be recovered within the framework of first-passage resetting [30,31], as shown in Appendix B. This is only true for this particular example, where the control variable is a scalar.

Our previous example focused on the case of an infinite time horizon, corresponding to $t_f \to \infty$. We now investigate the effect of a finite time horizon. One of the simplest settings where such effect can be studied is the case of a one-dimensional random walk with a Dirac delta final reward $R(x, t) = \alpha \delta(x - x_f)\delta(t - t_f)$, with $\alpha > 0$, and a constant resetting cost $c(x, t) = c$. Such parameters correspond to rewarding the system by a finite amount α for arriving at the

target position x_f at time t_f, while penalizing it with a constant cost c for each resetting event. Note that the system needs to arrive at the location x_f exactly at time t_f to get the reward, while earlier visits do not provide any additional benefits (as for a *rendezvous*, where one needs to be at the right place at the right time). The δ reward function has to be understood as the continuum limit of a Gaussian function whose width tends to zero [32]. Before presenting the optimal policy for this problem, it is instructive to consider two limiting cases. For $\alpha \to 0$, one should never reset as the reward is not worth the resetting cost. On the other hand, for $\alpha \to \infty$, the cost of resetting becomes negligible and the optimal strategy is to reset if restarting would bring the system closer to x_f, independently of time. Interestingly, we observe that the crossover between these two regimes occurs as a sharp transition at the critical value $\alpha = \alpha_c$, where $\alpha_c = x_f c \sqrt{2\pi e} \approx 4.13273\, x_f c$, which we predict analytically (see Appendix C). For $\alpha < \alpha_c$, the optimal policy is to never reset, corresponding to $\Omega(t) = \mathbb{R}$ for all $0 \leqslant t \leqslant t_f$. The situation is more subtle for $\alpha > \alpha_c$, where resetting is only favorable in a specific time window before t_f. To describe this window, it is convenient to introduce the backward time $\tau = t_f - t$. We find that no boundary is present for $\tau < \tau^*$, where τ^* is the smallest positive solution of the transcendental equation $\alpha e^{-x_f^2/(4D\tau^*)} = c\sqrt{4\pi D\tau^*}$. At $\tau = \tau^*$, a boundary appears and one must resort to numerical integration techniques to find the solution for $\tau > \tau^*$ (see Fig. 1). We observe numerically that the boundary evolves with τ, i.e., backward in time, in a nonmonotonic way and eventually disappears. This optimal policy can be understood as follows. Close to t_f, where $\tau < \tau^*$, it is unlikely for the system to reach the target location x_f from the origin in the remaining time. Thus, it is not convenient to reset. On the other hand, for very early times, it is not yet necessary to reset since, after resetting, the system would typically evolve away from the target.

Our framework can be easily generalized to other resetting dynamics. For instance, in light of the recent experiments of stochastic resetting on colloidal particles [33–35], it is relevant to consider the case where the new state of the system $\boldsymbol{x}'$ after a resetting event is drawn at random from some probability distribution $P_R(\boldsymbol{x}'|\boldsymbol{x})$, which can eventually depend on the current $\boldsymbol{x}$ of the system. In practice, it is physically impossible to reset the particles to a fixed location since resetting is performed by using an optic trap. In this case, our main results in Eqs. (5) and (7) remain valid, while the definition of the domain $\Omega(t)$ is modified as

$$\Omega(t) = \left\{ \boldsymbol{x} : J(\boldsymbol{x}, t) \geqslant \int d\boldsymbol{x}' \, [J(\boldsymbol{x}', t) P_R(\boldsymbol{x}'|\boldsymbol{x})] - c(\boldsymbol{x}, t) \right\}. \tag{11}$$

Note that the case of a fixed resetting location $\boldsymbol{x}_{\mathrm{res}}$ corresponds to $P_R(\boldsymbol{x}'|\boldsymbol{x}) = \delta(\boldsymbol{x}' - \boldsymbol{x}_{\mathrm{res}})$. Similarly, one can extend the framework to describe situations in which a finite amount of time is required for each resetting event.

Finally, we illustrate the generality of our framework by applying it to the problem of finding the optimal lockdown policy to navigate a pandemic. We do not aim to provide an accurate description of a specific pandemic, but rather to illustrate the generality of our framework and to gain interesting insights. We consider a population of N individuals and we model the evolution of the epidemic with the susceptible-infectious-recovered (SIR) model, which is one of the simplest compartmental models in epidemiology [36]. Previous works have approached this problem by continuously controlling the infection rate [37]. However, in real situations, one cannot adapt the restrictions in a continuous way and our framework is well suited to describe discontinuous policies, such as lockdowns. In Appendix D, we show that our technique can be directly applied to this problem to find the optimal time to impose a lockdown.

In sum, we combined optimal control and stochastic resetting to address the effectiveness of restarting policies. The emerging framework, contained in Eqs. (6) and (7), provides a unifying paradigm to obtain the optimal resetting strategy for a wide class of dynamical systems. Our method can be generalized to discrete-time systems and quadratic costs associated with resetting (see Appendix E). Furthermore, it is a simple exercise to include a continuous control policy in addition to resetting to account for realistic systems. In addition, one would need to explore ways to solve the moving boundary problem in high dimensions, which might require approximation schemes. It would be interesting to investigate extensions to optimal stopping problems and to study cost functions that are first-passage functionals [38], for instance where the time horizon is a first-passage time. This would be particularly relevant in the context of search processes.

APPENDIX A: DERIVATION OF EQS. (6) AND (7)

In this Appendix, we derive our main result given in Eqs. (6) and (7) of the main text. This derivation is based on a dynamical programming argument (also known as a backward approach). We consider evolving the system from time t to $t + dt$. According to the equations of motion in Eq. (3) of the main text, the state of the system either (i) evolves from $\boldsymbol{x}$ to $\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x}, t)dt + \sqrt{2D}\,\boldsymbol{\eta}(t)dt$ with probability $1 - r(\boldsymbol{x}, t)dt$ or (ii) is reset to position $\boldsymbol{x}_{\mathrm{res}}$ with probability $r(\boldsymbol{x}, t)dt$. In the time interval dt, the payoff in Eq. (4) of the main text changes by the amount $R(\boldsymbol{x}, t)dt - c(\boldsymbol{x}, t)r(\boldsymbol{x}, t)dt$. For the subsequent evolution of the process from $t + dt$ to t_f, the new initial value is either $\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x}, t)dt + \sqrt{2D}\,\boldsymbol{\eta}(t)dt$ in the former case or $\boldsymbol{x}_{\mathrm{res}}$ in the latter case. Following this argument and using the notation $\mathcal{F}[r\,|\,\boldsymbol{x}, t] \equiv \mathcal{F}_{\boldsymbol{x},t}[r]$, we obtain

$$\begin{aligned} \mathcal{F}[r\,|\,\boldsymbol{x}, t] = {}& R(\boldsymbol{x}, t)dt - c(\boldsymbol{x}, t)r(\boldsymbol{x}, t)dt \\ & + \langle \mathcal{F}[r\,|\,\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x}, t)dt + \sqrt{2D}\,\boldsymbol{\eta}(t)dt, t + dt] \rangle \\ & \times [1 - r(\boldsymbol{x}, t)dt] \\ & + \mathcal{F}[r\,|\,\boldsymbol{x}_{\mathrm{res}}, t + dt]r(\boldsymbol{x}, t)dt, \end{aligned} \tag{A1}$$

where $\langle \cdot \rangle$ is an average over the noise realizations $\boldsymbol{\eta}(t)$. The goal is now to maximize both sides over the full function r. Note that the functional $\mathcal{F}$ on the right-hand side of Eq. (A1) does not depend on the value of the function $r(\boldsymbol{x}, t)$ at time

t, given that its initial time is $t + dt$. We can therefore first maximize over the function $r(\mathbf{x}, \tau)$ for $\tau \in [t + dt, t_f]$ and then over the value $r(\mathbf{x}, t)$ at time t, which gives

$$-\partial_t J(\mathbf{x}, t) = R(\mathbf{x}, t) + D\Delta_x J(\mathbf{x}, t) + \mathbf{f}(\mathbf{x}, t) \cdot \nabla_x J(\mathbf{x}, t)$$
$$+ \max_{r(\mathbf{x},t)} \{r(\mathbf{x}, t)[J(\mathbf{x}_{\text{res}}, t) - c(\mathbf{x}, t) - J(\mathbf{x}, t)]\},$$
$$(A2)$$

where we used the definition of $J(\mathbf{x}, t)$, expanded to first order in dt, and averaged over $\boldsymbol{\eta}(t)$. We are left with maximizing a linear function of the scalar variable $r(\mathbf{x}, t)$. Given that $r(\mathbf{x}, t)$ is a Poissonian resetting rate, it must be positive and less than $1/dt$ since $r(\mathbf{x}, t)dt$ is a probability. This immediately gives the expression for the bang-bang optimal control policy $r^*(\mathbf{x}, t)$ given in Eq. (5) in the main text along with the domain $\Omega(t)$. Plugging this back into Eq. (A2), we obtain the moving boundary problem introduced in the main text.

The Neumann boundary condition $\nabla_x J(\mathbf{x}, t) \cdot \mathbf{n}(\mathbf{x}) = 0$ can be obtained upon a careful analysis of Eq. (A1) close to the boundary. We start from Eq. (A1) and write

$$\mathcal{F}[r \,|\, \mathbf{x}, t] = R(\mathbf{x}, t)dt - c(\mathbf{x}, t)r(\mathbf{x}, t)dt$$
$$+ \langle \mathcal{F}[r \,|\, \mathbf{x} + \mathbf{f}(\mathbf{x}, t)dt + \sqrt{2D}\,\boldsymbol{\eta}(t)dt, t + dt] \rangle$$
$$\times [1 - r(\mathbf{x}, t)dt]$$
$$+ \mathcal{F}[r \,|\, \mathbf{x}_{\text{res}}, t + dt]r(\mathbf{x}, t)dt. \quad (A3)$$

We set $\mathbf{x} \in \partial\Omega(t)$ and we expand Eq. (A3) to the first nontrivial order in dt. The main difference with the case $\mathbf{x} \in \Omega(t)$, discussed above, is that the linear term in $\boldsymbol{\eta}(t)$ does not vanish as the gradient of the payoff functional with respect to $\mathbf{x}$ vanishes outside of the domain $\Omega(t)$. We obtain

$$0 = \sqrt{2D}\nabla_x \mathcal{F}[r \,|\, \mathbf{x}, t] \cdot \langle \boldsymbol{\eta}(t)dt \, \mathbb{1}[\mathbf{x} + \sqrt{2D}\,\boldsymbol{\eta}(t)dt \in \Omega(t)] \rangle,$$
$$(A4)$$

where $\mathbb{1}$ is the indicator function and the average over $\boldsymbol{\eta}(t)$ is of the order of $O(\sqrt{dt})$. By decomposing $\boldsymbol{\eta}(t)$ over the parallel and perpendicular components of the normal direction $\mathbf{n}(\mathbf{x})$ to the boundary at $\mathbf{x}$, Eq. (A4) immediately gives the boundary condition $\nabla_x \mathcal{F}[r|\mathbf{x}, t] \cdot \mathbf{n}(\mathbf{x}) = 0$, which is, by definition, also valid for J.

APPENDIX B: FIRST-PASSAGE RESETTING IN A FINITE INTERVAL

In this Appendix, we show that the optimal policy in the special case of discounted payoffs can be obtained within the framework of first-passage resetting. We consider a one-dimensional diffusive particle with diffusion coefficient D whose position x belongs to the finite interval $[-L, L]$ with $L > 0$. We assume that when the particle reaches any of the interval boundaries, it is instantaneously reset to the origin $x_{\text{res}} = 0$. We assume that at the initial time, the particle is at position x_0, with $-L < x_0 < L$. In analogy with the discounted payoff in the main text, we associate a discounted reward $R(x, t) = e^{-\beta t}\mathcal{R}(x)$ and discounted cost of resetting, $c(x, t) = e^{-\beta t}\mathcal{C}(x)$, to the trajectory of the particle. Considering an infinite time horizon, the average payoff is therefore
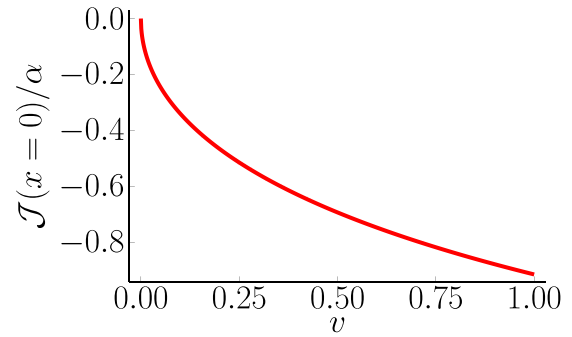

FIG. 3. The rescaled optimal expected payoff $\mathcal{J}(x)$ for a one-dimensional random walk starting at $x = 0$ for the discounted reward $\mathcal{R}(x) = -\alpha x^2$ and cost $\mathcal{C}(x) = c$ as a function of the cost-reward ratio $v = c/\alpha$.

given by

$$\mathcal{J}(x|L) = \left\langle \int_0^\infty d\tau\, e^{-\beta\tau}\mathcal{R}(x) - \sum_{i=1}^\infty e^{-\beta t_i}\mathcal{C}(x) \right\rangle_{x_0}, \quad (B1)$$

where t_i is the time of the ith resetting event and the average is computed over all trajectories $\{x(\tau)\}$ with $x(0) = x_0$. In particular, for $\mathcal{R}(x) = -\alpha x^2$ and $\mathcal{C}(x) = c$, we obtain

$$\mathcal{J}(x|L) = -\alpha \int_0^\infty d\tau\, e^{-\beta\tau}\langle x(\tau)^2 \rangle_{x_0} - c\sum_{i=1}^\infty \langle e^{-\beta t_i} \rangle_{x_0}. \quad (B2)$$

The two average quantities can be computed by standard renewal techniques and one obtains

$$\mathcal{J}(x|L) = -\alpha \left[\frac{2D}{\beta^2} + \frac{x_0^2}{\beta} - \frac{L^2}{\beta}\frac{\cosh(x_0\sqrt{\beta/D})}{\cosh(L\sqrt{\beta/D}) - 1} \right]$$
$$- c\left[\frac{\cosh(x_0\sqrt{\beta/D})}{\cosh(L\sqrt{\beta/D}) - 1} \right]. \quad (B3)$$

Upon minimizing over L and setting $\beta = D = 1$, we recover the expression in Eqs. (9) and (10) in the main text. The optimal payoff in Eq. (9) in the main text is shown as a function of the cost-reward ratio in Fig. 3. The optimal payoff decreases monotonically as the cost-reward ratio is increased.

APPENDIX C: DIRAC δ FINAL REWARD

In this Appendix, we consider the case of a Dirac δ final reward in $d = 1$, corresponding to $R(x, t) = \alpha\delta(x - x_f)\delta(t - t_f)$, where $\alpha > 0$ is the magnitude of the reward and x_f is the target location. We also assume that $f(x, t) = 0$, $c(x, t) = c > 0$ and we consider the optimal payoff in the time-reversed dynamics, defined as $I(x, \tau) = J(x, t_f - \tau)$. It is easy to show that $I(x, \tau)$ satisfies the diffusion equation,

$$\partial_\tau I(x, \tau) = D\partial_{xx}I(x, \tau), \quad (C1)$$

with initial condition $I(x, \tau = 0) = \alpha\delta(x - x_f)$. Assuming that no boundary appears for small τ (to be verified *a posteriori*), the optimal payoff function is the usual Gaussian profile,

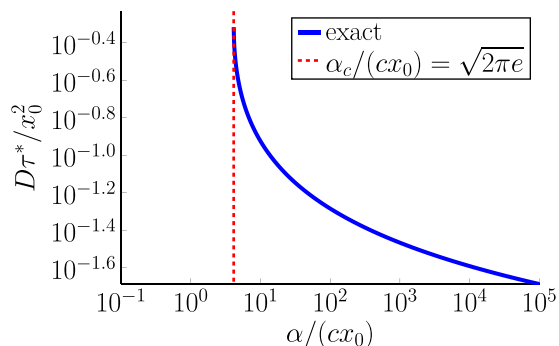$$I(x, \tau) = \alpha \frac{1}{\sqrt{4\pi D t}}e^{-(x-x_f)^2/(4\pi D\tau)}. \quad (C2)$$

FIG. 4. Scaled time $D\tau^*/x_0^2$ at which the barrier first appears as a function of $\alpha/(cx_0)$. The continuous blue line corresponds to the smallest positive solution of $\alpha = c\sqrt{4\pi D\tau}\, e^{x_f^2/(4D\tau)}$, while the dashed red line corresponds to the critical value $\alpha_c = \sqrt{2\pi e}cx_0$. For $\alpha < \alpha_c$, no barrier is present.

A boundary appears at time τ when the condition $I(x, \tau) < I(0, \tau) + c$ is verified for the first time for some value of x. Using Eq. (C2), this condition can be rewritten, for $|x - x_f| > x_f$, as

$$\alpha > \frac{c\sqrt{4\pi D\tau}}{e^{-x_f^2/(4D\tau)} - e^{-(x-x_f)^2/(4D\tau)}}. \tag{C3}$$

Minimizing the right-hand side of Eq. (C3) over x and τ, we obtain $\alpha > cx_f\sqrt{2e\pi}$. Thus, for $\alpha < \alpha_c = cx_f\sqrt{2e\pi}$, no boundary appears and the cost function is given by Eq. (C2) for any τ. On the other hand, for $\alpha > \alpha_c$, two boundaries appear at time τ^*, which is the smallest solution of $\alpha = c\sqrt{4\pi D\tau}\, e^{x_f^2/(4D\tau)}$. Thus, for $\tau < \tau^*$, the cost function is given by Eq. (C2), while it is hard to determine analytically for $\tau > \tau^*$. We obtain numerically the boundary for $\tau > \tau^*$ (see Fig. 1 of the main text). Note that at $\tau = \tau^*$, the condition in Eq. (C3) is only verified for $x \to \pm\infty$, meaning that the two boundaries start from infinity at the critical time. The critical time τ^* is shown in Fig. 4 as a function of α. The asymptotic behaviors of τ^* as a function of α are given by

$$\tau^* = \begin{cases} x_f^2/(2D) & \text{for } \alpha \to \alpha_c, \\ [x_f^2/(4D)]/\ln(\alpha) & \text{for } \alpha \to \infty. \end{cases} \tag{C4}$$
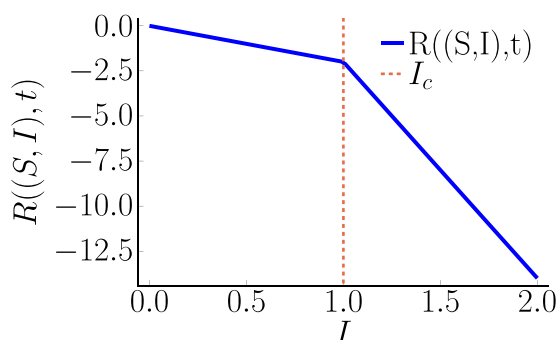

FIG. 5. Reward function $R((S, I), t)$ in (D3) describing the public-health cost in the context of the optimal lockdown policy problem. The reward is a piecewise linear function whose slope increases above the hospital alert level threshold I_c.

APPENDIX D: APPLICATION TO LOCKDOWN POLICIES

In this Appendix, we apply the framework developed in the main text to a problem of finding the optimal lockdown policy to navigate a pandemic. To describe the propagation of an epidemic within a population, we employ the SIR model. We consider a population of N individuals, divided into three groups (S, I, R), where S, I, and R are, respectively, the number of susceptible, infected, or recovered individuals. Note that by definition, $S + I + R = N$ and therefore a given state of the system is completely specified by the vector $\boldsymbol{x} = (S, I)$. We assume that new infections occur at a rate βSI, while infected individuals recover at a rate γI. In other words, the system evolves as a Markov jump process with rates

$$w[(S, I) \to (S - 1, I + 1)] = \beta SI \tag{D1}$$

and

$$w[(S, I) \to (S, I - 1)] = \gamma I, \tag{D2}$$

where $w[\boldsymbol{x} \to \boldsymbol{x}']$ indicates the rate of the transition from $\boldsymbol{x}$ to $\boldsymbol{x}'$.

During a real pandemic, national governments are usually presented with conflicting objectives. Indeed, the rapid spread of the disease carries heavy public-health costs, in particular when the number of infected individuals that require treatment exceeds the hospital alert level of a country. This spread can
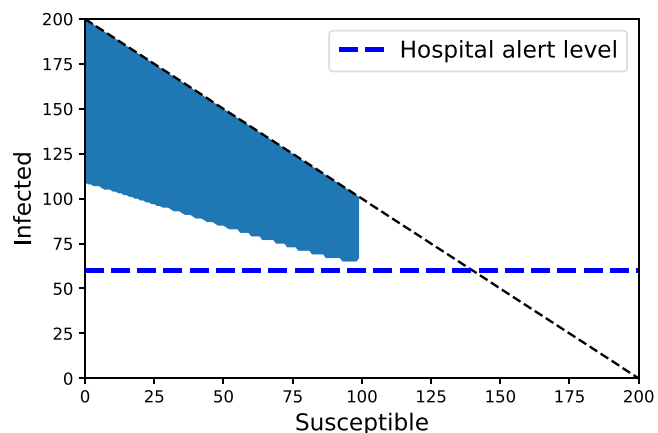

FIG. 6. Phase-space illustration of the optimal lockdown policy for a SIR epidemic model with infection rate β and recovery rate γ. Since the total number N of individuals is fixed, only configurations below the black dashed line, where $S + I \leqslant N$, are allowed. We consider a (negative) reward $R((S, I), t) = -aI - [b(I - I_a)]\, \theta(I - I_a)$, where I_a is the hospital alert level (horizontal dashed line) and $\theta(x)$ is the Heaviside step function. A fixed cost c is incurred upon resetting from the state (S, I) to the state $(S, \lfloor \alpha I \rfloor)$, which describes the effects of a lockdown. The optimal strategy that maximizes the expected payoff is to reset upon touching the blue region and to evolve freely in the white region, which we denote as $\Omega(t)$ in the text. The domain $\Omega(t)$ is obtained by numerical integration of the Stefan problem presented in (D4). The boundary of the domain $\Omega(t)$ guides the epidemic below the hospital alert level, while avoiding lockdowns as much as possible. The figure corresponds to the choice of parameters $t = 0$, $a = 2$, $b = 10$, $\alpha = 0.5$, $N = 200$, $I_a = 0.3 N$, $\beta = 5/N$, $t_f = 1$, and $\gamma = 1$. An optimally controlled epidemic trajectory is shown in Fig. 7.

be countered by imposing lockdowns, which, however, have an impact on the economy of a country.

To describe the public-health cost, we introduce the (negative) reward function (see Fig. 5),

$$R((S, I), t) = -aI - [b(I - I_c)] \, \theta(I - I_c), \qquad (D3)$$

where I_c represents the hospital alert level, $a > 0$ is the cost per infected person when the hospitals are not saturated, and $b > 0$ is the excess cost when the hospitals are saturated. Here, $\theta(z)$ is the Heaviside theta function. We describe lockdowns by allowing the possibility for governments to decrease the number of infected people by a constant fraction, without changing the number of susceptible people. In other words, the effect of a lockdown is to reset the system from state $\boldsymbol{x} = (S, I)$ to state $\boldsymbol{x}' = (S, \lfloor \alpha I \rfloor)$, where $0 < \alpha < 1$. We assume that each lockdown comes with a fixed cost $c > 0$, due to the negative socioeconomic impact. The goal in then to find the optimal resetting (or lockdown) policy $r^*(\boldsymbol{x}, t)$ which minimizes the payoff function defined in Eq. (4) of the main text, in order to control the system up to a given time horizon t_f.

It is easy to employ our framework in this case and to show that the optimal payoff function $J((S, I), t)$ evolves according to

$$
\begin{aligned}
-\partial_t J((S, I), t) = {}& R((S, I), t) + \beta SI[J((S-1, I+1), t) \\
& - J((S, I), t)] - \gamma I[J((I-1, S), t) \\
& - J((S, I), t)],
\end{aligned}
\qquad (D4)
$$

for $(S, I) \in \Omega(t)$, and

$$J((S, I), t) = J((S, \lfloor \alpha I \rfloor), t) - c, \qquad (D5)$$

for $(S, I) \notin \Omega(t)$. Here the domain $\Omega(t)$ is defined as

$$\Omega(t) = \{\boldsymbol{x} : J((S, I), t) \geqslant J((S, \lfloor \alpha I \rfloor), t) - c\}. \qquad (D6)$$
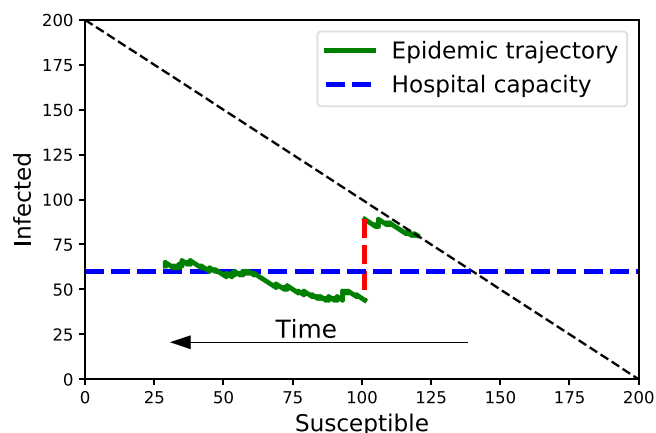

FIG. 7. Epidemic evolution in the SIR model controlled by an optimal lockdown policy according to the reward function given in (D3) and a fixed cost c for resetting. The red dashed line corresponds to the lockdown event. Since the total number N of individuals is fixed, only configurations below the black dashed line, where $S + I \leqslant N$, are allowed. The horizontal dashed line describes hospital alert level I_c. The figure corresponds to the choice of parameters $a = 2$, $b = 10$, $\alpha = 0.5$, $N = 200$, $I_c = 0.3 \, N$, $\beta = 5/N$, $\gamma = 1$, and $t_f = 1$, which are the same as those used in Fig. 6.

We employ our framework numerically and obtain the optimal policy in the $(S - I)$ plane, as shown in Fig. 6. As can be seen in this figure, the policy is to never reset if the number of susceptible individuals is larger than a critical value. Then, as the number of susceptible individuals is lowered, the optimal policy is to reset if the number of infected individuals is larger than a threshold which increases as the number of susceptible individuals is lowered. The optimal domain $\Omega(t)$ guides the epidemic below the hospital alert level, while avoiding lockdowns as much as possible. An optimally managed epidemic trajectory is shown in Fig. 7.

APPENDIX E: GENERALIZATION TO A QUADRATIC RESETTING COST

In this Appendix, we generalize our framework to the case in which a quadratic cost is associated to resetting with the following payoff functional:

$$\mathcal{F}[r|\boldsymbol{x}_0, t] = \left\langle \int_t^{t_f} d\tau \left[R(\boldsymbol{x}(\tau), \tau) - \frac{1}{2} c(\boldsymbol{x}(\tau), \tau) r(\boldsymbol{x}(\tau), \tau)^2 \right] \right\rangle_{\boldsymbol{x}_0}, \qquad (E1)$$

with $c(\boldsymbol{x}, \tau) \geqslant 0$. Following the procedure outlined in the main text, we obtain the moving boundary problem,

$$-\partial_t J(\boldsymbol{x}, t) = D\Delta_{\boldsymbol{x}} J(\boldsymbol{x}, t) + \boldsymbol{f}(\boldsymbol{x}, t) \cdot \nabla_{\boldsymbol{x}} J(\boldsymbol{x}, t) + R(\boldsymbol{x}, t), \quad \boldsymbol{x} \in \Omega(t)$$

$$-\partial_t J(\boldsymbol{x}, t) = D\Delta_{\boldsymbol{x}} J(\boldsymbol{x}, t) + \boldsymbol{f}(\boldsymbol{x}, t) \cdot \nabla_{\boldsymbol{x}} J(\boldsymbol{x}, t) + R(\boldsymbol{x}, t) + \frac{1}{2 \, c[\boldsymbol{x}(\tau), \tau]} [J(\boldsymbol{x}, t) - J(\boldsymbol{0}, t)]^2, \quad \boldsymbol{x} \notin \Omega(t), \qquad (E2)$$

where

$$\Omega(t) = \{\boldsymbol{x} : J(\boldsymbol{x}, t) \geqslant J(\boldsymbol{x}_{\mathrm{res}}, t)\}. \qquad (E3)$$

The differential equation (E2) must by solved by imposing the continuity of the solution and its derivative at the boundary of $\Omega(t)$. The optimal resetting rate $r^*(\boldsymbol{x}, t)$ is no longer bang-bang and is given by

$$r^*(\boldsymbol{x}, t) = \begin{cases} 0, & \boldsymbol{x} \in \Omega(t) \\ [J(\boldsymbol{x}_{\mathrm{res}}, t) - J(\boldsymbol{x}, t)]/c[\boldsymbol{x}(\tau), \tau], & \boldsymbol{x} \notin \Omega(t). \end{cases} \qquad (E4)$$

APPENDIX F: GENERALIZATION TO DISCRETE-TIME SYSTEMS

In this Appendix, we generalize our framework to the case of discrete-time systems whose state x_n evolves according to the Markov rule,

$$x_n = \begin{cases} x_{\text{res}} & \text{with probability } p_n(x) \\ x_{n-1} + f_n(x_{n-1}) + \eta_n & \text{with probability } 1 - p_n(x), \end{cases} \tag{F1}$$

where η_n are independently and identically distributed random variables drawn from a probability distribution $q(\eta)$ and $f_n(x)$ is an external force. The distribution $q(\eta)$ is arbitrary and includes, for instance, the case of fat-tailed distributions. The payoff functional in Eq. (4) of the main text straightforwardly generalizes to

$$\mathcal{F}(\{p_m, \ldots, p_n\}|x_m, m) = \left\langle \sum_{l=m}^{n} R(x_l, l) - c(x_l, l)p_l(x_l) \right\rangle_{x_m}, \tag{F2}$$

where the control is the sequence of resetting probabilities, $\{p_m, \ldots, p_n\}$, and the average on the right-hand side is taken over all system trajectories starting from x_m at step m. Following a similar approach as in the previous sections, we obtain that the optimal policy is given by

$$p_m(x) = \begin{cases} 0 & \text{if } x \in \Omega_m, \\ 1 & \text{if } x \notin \Omega_m, \end{cases} \tag{F3}$$

where $\Omega_m \subseteq \mathbb{R}^d$ is a time-dependent domain. The corresponding discrete moving boundary problem for the optimal payoff function, $J(x, m) = \max_{\{p_m, \ldots, p_n\}} \mathcal{F}[\{p_m, \ldots, p_n\}|x, m]$, is given by

$$J(x, m) = R(x, m) + \max \left\{ J(x_{\text{res}}, m + 1) + c(x, m), \int_{-\infty}^{\infty} d\eta \, q(\eta) J[x + f_m(x) + \eta, m + 1] \right\}, \tag{F4}$$

with

$$\Omega_m = \left\{ x : \int_{-\infty}^{\infty} d\eta \, p(\eta) J[x + f_m(x) + \eta, m + 1] \geqslant J(x_{\text{res}}, m + 1) + c(x, m) \right\}. \tag{F5}$$

[1] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko, *The Mathematical Theory of Optimal Processes* (Interscience, New York, 1962).

[2] R. Bellman and R. E. Kalaba, *Dynamic Programming and Modern Control Theory* (Academic Press, New York, 1965).

[3] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* (Prentice Hall, Englewood Cliffs, NJ, 2003).

[4] H. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications* (Springer Science & Business Media, Berlin, 2009).

[5] R. Stengel, *Optimal Control and Estimation* (Dover, New York, 1993).

[6] S. N. Majumdar and H. Orland, Effective Langevin equations for constrained stochastic processes, J. Stat. Mech. (2015) P06039.

[7] H. J. Kappen, Linear Theory for Control of Nonlinear Stochastic Systems, Phys. Rev. Lett. **95**, 200201 (2005).

[8] T. Schmiedl and U. Seifert, Optimal Finite-Time Processes In Stochastic Thermodynamics, Phys. Rev. Lett. **98**, 108301 (2007).

[9] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, Macroscopic fluctuation theory, Rev. Mod. Phys. **87**, 593 (2015).

[10] U. Boscain, M. Sigalotti, and D. Sugny, Introduction to the pontryagin maximum principle for quantum optimal control, Phys. Rev. X Quantum **2**, 030203 (2021).

[11] A. Dolgui, D. Ivanov, S. P. Sethi, and B. Sokolov, Scheduling in production, supply chain and Industry 4.0 systems by optimal control: fundamentals, state-of-the-art and applications, Intl. J. Product. Res. **57**, 411 (2019).

[12] V. Gómez, S. Thijssen, A. Symington, S. Hailes, and H. J. Kappen, Real-time Stochastic optimal control for multi-agent quadrotor systems, in *Twenty-Sixth International Conference on Automated Planning and Scheduling*, edited by A. Coles, A. Coles, S. Edelkamp, D. Magazzeni, and S. Sanner (AAAI Press, Palo Alto, CA, 2016).

[13] K. Khanin, S. Nechaev, G. Oshanin, A. Sobolevski, and O. Vasilyev, Ballistic deposition patterns beneath a growing KPZ interface, Phys. Rev. E **82**, 061107 (2010).

[14] A. Montanari and R. Zecchina, Optimizing Searches via Rare Events, Phys. Rev. Lett. **88**, 178701 (2002).

[15] M. F. Shlesinger, Search research, Nature (London) **443**, 281 (2006).

[16] M. R. Evans and S. N. Majumdar, Diffusion with Stochastic Resetting, Phys. Rev. Lett. **106**, 160601 (2011).

[17] M. R. Evans and S. N. Majumdar, Diffusion with optimal resetting, J. Phys. A: Math. Theor. **44**, 435001 (2011).

[18] S. Reuveni, M. Urbakh, and J. Klafter, Role of substrate unbinding in Michaelis Menten enzymatic reactions, Proc. Natl. Acad. Sci. USA **111**, 4391 (2014).

[19] A. Economou and D. Fakinos, A continuous-time Markov chain under the influence of a regulating point process and applica-

tions in stochastic models with catastrophes, Eur. J. Operat. Res. **149**, 625 (2003).

[20] P. Visco, R. J. Allen, S. N. Majumdar, M. R. Evans, Switching and growth for microbial populations in catastrophic responsive environments, Biophys. J. **98**, 1099 (2010).

[21] V. Méndez and D. Campos, Transport properties and first-arrival statistics of random motion with stochastic reset times, Phys. Rev. E **93**, 022106 (2016).

[22] S. Eule and J. J. Metzger, Nonequilibrium steady states of stochastic processes with intermittent resetting, New J. Phys. **18**, 033006 (2016).

[23] F. Mori, S. N. Majumdar, and G. Schehr, Distribution of the time of the maximum for stationary processes, Europhys. Lett. **135**, 30003 (2021).

[24] L. Kuśmierz, S. N. Majumdar, S. Sabhapandit, G. Schehr, First Order Transition for the Optimal Search Time of Lévy Flights with Resetting, Phys. Rev. Lett. **113**, 220602 (2014).

[25] S. N. Majumdar, S. Sabhapandit, and G. Schehr, Dynamical transition in the temporal relaxation of stochastic processes under resetting, Phys. Rev. E **91**, 052131 (2015).

[26] D. Campos and V. Méndez, Phase transitions in optimal search times: How random walkers should combine resetting and flight scales, Phys. Rev. E **92**, 062115 (2015).

[27] M. R. Evans, S. N. Majumdar, and G. Schehr, Stochastic Resetting and Applications, J. Phys. A: Math. Theor. **53**, 193001 (2020).

[28] J. Crank, *Free and Moving Boundary Problems* (Oxford University Press, New York, 1984).

[29] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions* (Springer Science & Business Media, New York, 2006).

[30] B. De Bruyne, J. Randon-Furling, and S. Redner, Optimization in First-Passage Resetting, Phys. Rev. Lett. **125**, 050602 (2020).

[31] B. De Bruyne, J Randon-Furling, and S. Redner, Optimization and growth in first-passage resetting, J. Stat. Mech. (2021) 013203.

[32] This is the same way as for the standard deviation of the white noise.

[33] O. Tal-Friedman, A. Pal, A. Sekhon, S. Reuveni, and Y. Roichman, Experimental realization of diffusion with stochastic resetting, J. Phys. Chem. Lett. **11**, 7350 (2020).

[34] B. Besga, A. Bovon, A. Petrosyan, S. N. Majumdar, and S. Ciliberto, Optimal mean first-passage time for a Brownian searcher subjected to resetting: Experimental and theoretical results, Phys. Rev. Res. **2**, 032029(R) (2020).

[35] B. Besga, F. Faisant, A. Petrosyan, S. Ciliberto, and S. N. Majumdar, Dynamical phase transition in the first-passage probability of a Brownian motion, Phys. Rev. E **104**, L012102 (2021).

[36] W. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. London A **115**, 700 (1927)

[37] J. Samuel and S. Sinha, Optimal control in pandemics, Phys. Rev. E **103**, L010301 (2021).

[38] S. N. Majumdar, Persistence in nonequilibrium systems, Curr. Sci. **89**, 2076 (2005).