TECHNISCHE UNIVERSITÄT WIEN
Operations Research and Control Systems
SWM ORCOS

Optimal control and the Value of Information for a Stochastic Epidemiological SIS-Model

Peter Grandits, Raimund M. Kovacevic, Vladimir M. Veliov

Research Report 2016-14
December, 2016

Operations Research and Control Systems
Institute of Statistics and Mathematical Methods in Economics
Vienna University of Technology

Research Unit ORCOS
Wiedner Hauptstraße 8 / E105-4
1040 Vienna, Austria
E-mail: orcos@tuwien.ac.at

Optimal control and the Value of Information for a Stochastic Epidemiological SIS-Model*

Peter Grandits†, Raimund M. Kovacevic‡, and Vladimir M. Veliov §

Abstract

This paper presents a stochastic SIS-model of epidemic disease, where the recovery rate can be influenced by a decision maker. The problem of minimization of the expected aggregated economic losses due to infection and due to medication is considered. The resulting stochastic optimal control problem is investigated on two alternative assumptions about the information pattern. If a complete and exact measurement is always available, then the optimal control is sought in a state-feedback form for which the Hamilton-Jacobi-Bellman (H-J-B) equation is employed. If no state measurement is available at all, then the optimal control is sought in an open-loop form. Given at least an estimated initial probability density for the number of infected, the open loop problem can be reformulated as an optimal control problem for the associated Kolmogorov forward equation (describing the evolution of the probability density of the state). Optimality conditions are derived in both cases, which requires involvement of non-standard arguments due to the degeneracy of the involved H-J-B and Kolmogorov parabolic equations. The effect of the observations on the optimal performance is investigated theoretically and numerically.

Keywords: stochastic differential equations, optimal control, epidemiology, SIS models, optimal treatment, Hamilton-Jacobi-Belman equation, Fokker-Planck equation

Classification: 93E20, 60H10, 92D30

1 Introduction

SIS-models form a class of simple prototypical epidemiological models. They assume that individuals can be infected multiple times throughout their lives with no immunity after each infection. Typical examples mentioned in literature are rota-viruses, some sexually transmitted as well as bacterial infections, see e.g. [12, 3].

*This work was supported by the Austrian Science Foundation (FWF) under grant P24125-N13.

†Institute of Statistics and Mathematical Methods in Economics (Financial and Actuarial Mathematics), Vienna University of Technology, peter.grandits@tuwien.ac.at

‡Institute of Statistics and Mathematical Methods in Economics (Operations Research and Control Systems), Vienna University of Technology, raimund.kovacevic@tuwien.ac.at

§Institute of Statistics and Mathematical Methods in Economics (Operations Research and Control Systems), Vienna University of Technology, vladimir.veliov@tuwien.ac.at

In the following we denote the number of infected individuals[1] at time $t \geq 0$ by $x(t)$ and the number of susceptible individuals by $y(t)$. We will use capital letters, when referring to stochastic entities. These notations differ from the usual notations in epidemiology ($I(t)$ for infected and $S(t)$ for susceptible) but are more in line with the conventions in stochastic calculus. Throughout this paper we restrict the discussion to the simplest case of a stationary population. Normalizing the population size to one,

$$x(t) + y(t) = 1, \tag{1}$$

means that $x(t)$ and $y(t)$ represent the fractions of infected and susceptible individuals. In this setup it suffices to regard only the number of infected, replacing $y(t)$ by $1 - x(t)$ wherever it appears.

The basic SIS model involves two parameters: the disease transmission coefficient (or strength of infection) $\beta > 0$, leading to a force of infection (or incidence rate) of $\beta x(t)$, and the recovery rate $\gamma > 0$. The force of infection models the rate at which susceptible individuals become infected, while the recovery rate is the rate at which infected individuals become susceptible again. One can interpret the reciprocal $1/\gamma$ as the average duration of infection.

Within this setup, the deterministic SIS model is described by the ordinary differential equation

$$\begin{aligned} \frac{\mathrm{d}x(t)}{\mathrm{d}t} &= \beta x(t)\,(1 - x(t)) - \gamma x(t), && (2) \\ x(0) &= x_0 \in [0, 1], && (3) \end{aligned}$$

which has an explicit solution whose properties are well known.

While many different approaches for stochastic epidemiological models exist in literature, see e.g. [6] or Chapter 6 of [15], in the present paper we analyze a diffusion-type version of the above SIS model: the deterministic equation (2) is replaced by a stochastic (Ito) differential equation

$$\begin{aligned} \mathrm{d}X(t) &= [\beta\,(1 - X(t))\,X(t) - \gamma X(t)]\ \mathrm{d}t + \sigma\,[(1 - X(t))\,X(t)]\ \mathrm{d}\mathcal{B}(t), && (4) \\ X(0) &= x_0 \in (0, 1). \end{aligned}$$

Here, $\mathcal{B}(\cdot)$ represents a standard Brownian motion in $\mathbb{R}$ and the solution $X(\cdot)$ of (4) is a one dimensional Ito diffusion process (in contrast to $x(\cdot)$, which is a deterministic function).

Such a formulation can be obtained, if one e.g. assumes that the parameter β in (2) is not constant and deterministically known but random, when $\beta\,\mathrm{d}t$ is replaced with the specification

$$\mathrm{d}\hat{\beta}(t) = \beta\ \mathrm{d}t + \sigma\ \mathrm{d}\mathcal{B}(t).$$

In this way it is possible to model random fluctuations of the parameter around the value β. Here the volatility parameter $\sigma > 0$ models the uncertainty of the true parameter process $\hat{\beta}$. This simple stochastic SIS model has been analyzed e.g. in [11], see also [7, 4] for an extended versions and [18]

[1]Strictly speaking, "number of susceptible/infected individuals" means "size of the susceptible/infected subpopulation", which is a real number.

for an analysis of the related stationary distributions.

Keeling and Rohani [14] call the specification of noise used in (4) *external parameter noise* (in contrast to constant noise, scaled additive noise and heterogeneous parameter noise). While in other noise specifications the number of infected is directly subject to random fluctuations, here the parameters of an originally deterministic epidemiological model are disturbed by random noise due to external unpredictable forces. See also [1, 2, 5] for possible underlying mechanisms.

So far, the literature on stochastic epidemiological SDE-models mainly focuses on descriptive models, where the parameters β, σ and in particular γ are static and fixed. While optimal control theory has been applied many times to deterministic epidemiological models, optimization for stochastic epidemiological models is a (so-far) neglected topic. In the present paper we make a step from descriptive to optimization models in the stochastic context. To this end, we assume that the "natural" development of the disease can be influenced by a decision maker. Various specifications are possible, but here we focus on the case where the decision maker is able to control the recovery rate γ. Such a control can be implemented e.g. by increasing the treatment capacity or the efficiency of medication.

In order to clarify the issue we first consider the deterministic SIS model (2). As before, we assume that γ is the recovery rate without any action by the decision maker, while $\gamma + u(t)$ is the recovery rate at time t if a control $u(t) \geq 0$ is applied at t. The control is costly, and we assume that the cost is proportional to the number of treated individuals, $x(t)$, and is marginally increasing, namely we specify the control cost at t as $c_2 x(t) u(t)^2$. In addition, taking an economic point of view, we assume that any infected individual creates for the society cost c_1 per unit of time, due to lost working hours and standard medical care, not including the treatment u. Then it is reasonable to consider the following standard optimal control problem, referred further as to *Deterministic Problem* (DP):

$$\min_u \int_0^T e^{-rt} \left[F(x(t), u(t)) \right] \, \mathrm{d}t + c_3 x(T),$$

subject to

$$\begin{aligned} \frac{\mathrm{d}x(t)}{\mathrm{d}t} &= a(x(t), u(t)), \quad x(0) = x_0 \in [0,1], \\ & u(t) \geq 0, \end{aligned} \tag{5}$$

where - for brevity - allover the paper we use the notations

$$a(x,u) = \beta x \left(1 - x\right) - (\gamma + u) x, \qquad b(x) = \sigma^2 x^2 \left(1 - x\right)^2 \tag{6}$$

and

$$F(x,u) = c_1 x + c_2 x u^2. \tag{7}$$

(the notation b will come into the play a bit later). Here $[0,T]$ is a given planning horizon, $r \geq 0$ is a discount factor (which can be taken equal to zero), and $c_3 x(T)$ is the expected accumulated future cost of the individuals that remain infected at T (can be taken equal to zero if the disease is expected to nearly die out till time T). We mention that DP is formulated as an open-loop problem (the control function u is sought as a function of time only), but even if a feedback control is allowed, in absence of disturbances it would generate the same optimal trajectory $x(t)$ as the optimal open-loop control.

Now we consider stochastic counterparts of DP. Since the Ito equation (4) generates a stochastic process, rather than a single trajectory, the information pattern becomes of crucial importance. Note that throughout this paper we will write $\mathbf{E}[\cdot]$ for the expectation and - if $X(t)$ denotes a stochastic process - $\mathbf{E}_{t,x}[\cdot]$ or even $\mathbf{E}_t[\cdot]$ for the conditional expectation given $X(t) = x$. We consider the following two "extreme" cases.

1. *Full information case* (Section 2). Here we assume that at any time t the exact current number of infected individuals becomes known to the decision maker. Correspondingly, the optimal control (which is a stochastic process) can be sought in feedback form: $u = u(X(s), s)$. Then the following optimal control problem (called further *Stochastic Feedback Problem* – SFP) becomes relevant:

$$\min_u \mathbf{E} \int_0^T e^{-rs} \left[F(X(s), u(X(s), s)) \right] \mathrm{d}s + c_3 \mathbf{E}\left[X(T) \right]. \tag{8}$$

subject to

$$\begin{aligned} \mathrm{d}X(s) &= a(X(s), u(X(s), s))\,\mathrm{d}s + \sqrt{b(X(s))}\,\mathrm{d}\mathcal{B}(s), \quad X(0) = x_0 \in (0, 1], && (9)\\ & u(X(s)) \geq 0. && (10) \end{aligned}$$

Here the methods of stochastic control can be applied, in particular the Hamilton-Jacobi-Bellman (HJB) equation, which we employ in the analysis in Section 2. In particular, we show that the constraint $u(x,t) \geq 0$ is not necessary at the optimum and use the related necessary condition for the HJB equation to derive a reduced partial differential equation for the value function of the SFP. Strict arguments (including a regularity result and a verification theorem) are given for showing that under certain standing assumptions the unique solution of the reduced equation in fact is the value function of the SFP.

2. *No information case* (Section 3). Exact information about the current state of the disease is usually not available, especially in less developed countries. Therefore, we consider a second "extreme" scenario, where the probability density of $X(0)$ is given, but no information about the current state becomes available in the time horizon $[0, T]$. In this case the feedback control cannot be implemented, therefore the optimal control is sought as function only of time: $u = u(t)$. Then a *Stochastic Open-loop Problem* (SOP) is considered, which has a similar formulation as SFP, but with $u(X(t), t)$ replaced with $u(t)$, which is now a deterministic function of time. The HJB equation cannot be applied directly in this case. Still it is possible to use the Ito-Lemma in order to formulate a verification theorem that comes close to a HJB-formulation. However, we will see that this needs quite strong

restrictions on the control function. Therefore we take a different approach. Namely, we employ the Kolmogorov-forward (or Fokker-Planck) equation related to (9), which describes the evolution of the density of $X(t)$ over time. Then SOP can be reformulated as an optimal control problem for a second order parabolic partial differential equation (the Fokker-Planck equation). In Section 3 we present and strictly prove Pontryagin type optimality conditions, which provide a base for a numerical procedure.

These results are used for an extensive example in Section 4, where both, the SFP and the SOP are solved numerically. While the reduced partial differential equation for the value function analyzed in Section 3 is used for solving the SFP, we use the steepest descent method for PDE-constraint optimal control problems in order to solve the SOP. In addition, we also solve the deterministic problem in a standard way. In particular we use the numerical results for comparing optimal solutions under differing information patterns: Assuming that reality is represented by the dynamics of the SOP, we analyze the gain (in terms of the objective function of SOP) from using the SOP instead of using the deterministic model DP. This quantifies the value of using the stochastic approach SOP. Moreover we also analyze the situation where the same dynamics still represents the reality for a decision maker, but another agent is able to observe the states. The expected gain that the agent may achieve (using SFP) compared to the decision maker (SOP) gives us an estimate for the amount that the decision maker is willing to pay for the information related to the observed states.

2 The stochastic closed loop optimal control problem

In the following we consider a decision maker who is able to observe the realized states $X(t)$ at any time $t \in [0, T]$ and who therefore is also able to react to this observations by choosing a feedback (that is, a state and time dependent) control $u(x, t)$. In this section we discuss the basic specification and properties of the related optimization problem SFP (see (8)–(10)) in the HJB framework, including a regularity result and a verification theorem for the associated vale function.

2.1 Stochastic control for the SIS model with SDE constraints

With $X(t)$ observable at time t, the relevant σ-field is $\mathcal{F} = \{\mathcal{F}_t = \sigma(X(t))\}$, i.e. the σ-field generated by the process X. The control $u(\cdot, \cdot)$, is a function $(0, 1) \times [0, T] \to \mathbb{R}$ of observed states and time. The exact space of admissible functions will be stated later in assumption 2.

In order to apply the dynamic programming approach we embed SFB into a family of problems parameterized by the initial time t and initial state x:

$$V(x,t) = \min_u \mathbf{E}_{t,x}\left[\int_t^T e^{-rs}\left[F(X(s),u(X(s),s))\right]\,ds + c_3X(T)\right] \tag{11}$$

subject to

$$\mathrm{d}X(s) = a(X(s),u(X(s),s))\,\mathrm{d}s + \sqrt{b(X(s))}\,\mathrm{d}\mathcal{B}(s) \quad X(t)=x, \tag{12}$$

$$u(X(s),s) \geq 0. \tag{13}$$

Remark 1. It is important to keep in mind the fact that, provided $u(x,t) \geq 0$ as assumed in our problem specification, the process $X(t)$, described by (12), stays in $I=(0,1)$ forever, when started within this interval. The diffusion term $b(x) = \sigma^2x^2(1-x)^2$ equals zero at $x=0$ and $x=1$. Moreover, the derivative of $b(x)$ also is equal to zero at these points, which in consequence are prescribed boundaries: at $x=1$ the drift is negative, hence there is an entrance boundary, whereas at $x=0$ we have $a(0,t)=0$, which indicates a natural boundary (see e.g. [10] p 119). As a natural boundary, $x=0$ is absorbing but is never reached if the process starts in $I=(0,1)$. In similar manner the entrance boundary $x=1$ is never reached.

For a fixed function u, the backward operator $\mathcal{L}$ related to (12) is given by

$$\mathcal{L}V\,(x,t) = V_t(x,t) + a(x,u(x.t))V_x(x,t) + \frac{1}{2}b(x)V_{xx}(x,t),$$

and we denote

$$\mathcal{L}^uV\,(x,t) = V_t(x,t) + a(x,u)V_x(x,t) + \frac{1}{2}b(x)V_{xx}(x,t). \tag{14}$$

Moreover, the related Hamilton-Jacobi-Bellman (HJB) equation is given by

$$\inf_{u\geq 0}\left\{\mathcal{L}^uW(x,t) + e^{-rt}F(x,u)\right\} = 0 \tag{15}$$
$$W(x,T) = c_3x,$$

which can be written as

$$W_t(x,t) + a(x,0)W_x(x,t) + \frac{1}{2}b(x)W_{xx}(x,t) + e^{-rt}c_1x + \inf_{u\geq 0}\left\{e^{-rt}c_2xu^2 - uxW_x(x,t)\right\} = 0 \tag{16}$$
$$W(x,T) = c_3x.$$

In the following subsection we will show a regularity result and a verification theorem, in particular it is shown that under certain regularity conditions the HJB equation has a solution, which is the optimal value function of the SFB problem (11). Here, we start with some simple facts about the optimal value function V and the optimal control u, namely that V is monotonically increasing in its first argument and that the optimal u is nonnegative, even when the condition $u(X(s),s) \geq 0$ is omitted in SFB (11).

To this end, consider the optimization problem

$$\mathcal{V}(x,t) = \min_u \mathbf{E}_{t,x}\left[\int_t^T e^{-rs}\left[F(X(s),u(X(s),s))\right]ds + c_3 X(T)\right] \tag{17}$$

subject to

$$\mathrm{d}X(s) = a(X(s),u(X(s),s))\,\mathrm{d}s + \sqrt{b(X(s))}\,\mathrm{d}\mathcal{B}(s) \quad X(t)=x,, \tag{18}$$

which is in fact (11)-(13) without the restriction on u.

Lemma 1. *The value function $\mathcal{V}(x,t)$ of problem (17)-(18) is strictly increasing in x.*

Proof. Consider two states x_0, x_1 with $x_0 < x_1$. For $i \in \{0,1\}$, denote an optimal strategy for a process starting at x_i at time t by $u_i^*(x,t)$ and the corresponding state process by $X_i^*(t)$. Using some $\varepsilon > 0$ define the following (non-optimal) strategy for a controlled process $\tilde{X}_0(t)$, starting at state x_0 at time t:

$$\tilde{u}_0(x,t) \quad = \quad \begin{cases} u_1^*(x,t) - \varepsilon & \text{if } x < X_1^*(t) \\ u_1^*(x,t) & \text{if } x = X_1^*(t). \end{cases} \tag{19}$$

In addition, define the stopping time $\tau = \inf\left\{t : \tilde{X}_0(t) = X_1^*(t)\right\}$ with $\inf\emptyset = T$ and denote $\tilde{\mathcal{V}}(x_1,t) = E\left[\int_t^T e^{-rs}\left(c_1\tilde{X}_0(s) + c_2\tilde{X}_0(s)\tilde{u}_1(\tilde{X}_0(s),s)^2\right)\,\mathrm{d}s + c_3\tilde{X}_0(T)\right]$

Then we have $\tilde{X}_0(s) \leq X_1^*(s)$ almost surely at each time $t \in [0,T]$ and (due to continuity of the sample paths) $\tau > 0$ almost surely. For $t \in [0, \tau\wedge T)$ we have $\tilde{X}_0(s) < X_1^*(s)$. Furthermore $\tilde{u}_0(s,x) \leq u_1^*(s,x)$ is ensured by (19). This implies

$$\begin{aligned} \mathcal{V}(x_1,t) - \tilde{\mathcal{V}}(x_0,t) \quad = \quad & c_1\mathbf{E}\left[\int_t^{\tau\wedge T} e^{-rs}\left(X_1^*(s) - \tilde{X}_0(s)\right)\,\mathrm{d}s + \int_{\tau\wedge T}^T e^{-rs}\left(X_1^*(s) - \tilde{X}_0(s)\right)\,\mathrm{d}s\right] \\ & + c_2\mathbf{E}\left[\int_0^T e^{-rs}\left(X_1^*(s)u_1^*(X_1^*(s),s)^2 - \tilde{X}_0(s)\tilde{u}_0(\tilde{X}_0(s),s)^2\right)\,\mathrm{d}s\right] \\ & + c_3\left[X_1^*(T) - \tilde{X}_0(T)\right] > 0, \end{aligned}$$

because the first integral is positive ($\tau > 0$), whereas all other integrals as well as the last line are nonnegative.

Finally, because $\tilde{u}_0$ is not the optimal strategy for a controlled process starting at x_0, we have

$$\mathcal{V}(x_0,t) \leq \tilde{\mathcal{V}}(x_0,t) < \mathcal{V}(x_1,t).$$

□

It seems that the constraint $u(x,t) \geq 0$ may complicate the situation. However, based on Lemma

(1) it can be shown that this constraint is irrelevant for the optimal solution.

Lemma 2. *The optimal control u^* of (17)-(18) is nonnegative. It follows that it also solves (11)-(13) and therefore that $V(x,t)$ is strictly monotone increasing in x.*

Proof. Let $u^*(x,t)$ be an optimal control for (17)-(18) and denote the related state process by $X^*(t)$.

Suppose that $u^*(t) < 0$ with positive probability on a set S of points in time, where S has positive Lebesgue-measure and choose a point in time τ_1 such that $\mathbb{P}\left(u(x,\tau_1) < 0\right) > 0$ and $\mathbb{P}\left(u(x,t) < 0\right) > 0$ for t in a set $S_1 \subseteq S$ with $\lambda(S_1) > 0$. Define now the control law

$$u'(x,t) = \begin{cases} u^*(x,t) & \text{for } t < \tau_1 \\ |u^*(x,t)| & \text{for } t \geq \tau_1. \end{cases}$$

Write $X'(t)$ for the related process of states. Because $X^*(\tau_1) = X'(\tau_1)$ we have almost surely $d(X^*(\tau_1) - X'(\tau_1)) \geq 0$ and $d\left(X^*(\tau_1) - X'(\tau_1)\right) = -2 \cdot u^*(\tau_1)X^*(\tau_1) > 0$ with positive probability. Therefore it can be inferred that $X^*(t) \geq X'(t)$ almost surely and $X^*(t) > X'(t)$ with positive probability for t in some (maybe small) interval $[\tau_1, \tau_2)$. In addition, $u^*(x,t)^2 = u'(x,t)^2$ holds almost surely. These facts imply

$$\mathbf{E}_{\tau_1} \int_{\tau_1}^{\tau_2} e^{-rs} \left(c_1 X^*(s) + c_2 X^*(s) u^*(X^*(s), s)^2\right) \, ds > \mathbf{E}_{\tau_1} \int_{\tau_1}^{\tau_2} e^{-rs} \left(c_1 X'(s) + c_2 X'(s) u'(X'(s), s)^2\right) \, ds. \tag{20}$$

Moreover, at time τ_2 the inequality $X^*(\tau_2) > X'(\tau_2)$ implies

$$\mathcal{V}(X'(\tau_2), \tau_2) < \mathcal{V}(X^*(\tau_2), \tau) \tag{21}$$

by Lemma 1.

It follows that a strategy u'' with related process X'' that switches from u^* to u' at time τ_1 and switches back to u^*at time τ_2 is a better strategy than keeping the optimal strategy u^* over the whole planning horizon. This contradicts the assumption that u^* is optimal, if $\tau_1 < T$. Hence, $u^*(x,t) \geq 0$ for problem (17)-(18), the condition $u^*(x,t) \geq 0$ is redundant in (11)-(13), and $V(x,t) = \mathcal{V}(x,t)$ holds. By applying Lemma 1, the function $V(x,t)$ must then be strictly monotone increasing. □

The minimization in the HJB equation (15) leads to

$$u^*(t) \quad = \quad \frac{e^{rt}}{2c_2} W_x(x,t), \tag{22}$$

which is true only if the process starts in the open interval $\{0,1\}$ - when $x = 0$ is alsmost surely never reached. Applying this result, reduces the HJB equation to

$$W_t(c,t) + a(x,0)W_x(x,t) + \frac{1}{2}b(x)W_{xx}(x,t) + e^{-rt}c_1 x - \frac{e^{rt}}{4c_2} x \left[W_x(x,t)\right]^2 = 0.. \tag{23}$$

In fact we will show below (see Corollary 1) that there exists a smooth enough solution W of the equation (23) which fulfills $W_x(x,t) \geq 0$. This readily implies that W is also a solution of the HJB-equation (15) with optimal control (22).

2.2 A regularity result and the verification theorem

In order to show that the solution of the HJB equation (15), resp. (23), is indeed the value function of our problem, we will have to show a so called verification result. To prove this, one needs some regularity for the solution of the HJB equation. We provide this in the following theorem. Note that here and in the following $C^{2+\beta,1+\frac{\beta}{2}}$ denotes the space of functions, such that second derivative with respect to x is Hölder continuous with index *beta* and the first derivative with respect to t is Hölder continuous with index $\frac{\beta}{2}$.

Theorem 1. *Equation (23) has at least one solution $W(x,t)$, which is bounded by a positive constant M and which is an element of $C^{2+\beta,1+\beta/2}(Q_t)$, with $0<\beta<1$ and $Q_T := \{(x,t)|\epsilon \leq x \leq 1-\epsilon, 0 \leq t \leq T\}$, for arbitrary small positive ϵ. Moreover, $W(x,t)$ is uniformly continuous at $x=0$.*

Proof. We want to apply in this proof standard results for parabolic Cauchy problems, see [19], hence we will first transform our x-variable to a z-variable, varying on $]-\infty,\infty[$. So let $z=\phi(x)$, with $\phi \in C^\infty$ and $\phi :]0,1[\to]-\infty,\infty[$. We will need to have control of the transformation near the boundary, so we prescribe its value there explicitly, whereas we just impose ϕ to be strictly monotone increasing in the middle part, i.e.

$$\phi(x) := \begin{cases} \ln x, & x \in]0, \frac{1}{10}], \\ \text{strict. mon. increasing}, & x \in]\frac{1}{10}, \frac{9}{10}[, \\ -\ln(1-x), & x \in [\frac{9}{10}, 1[, \end{cases} \tag{24}$$

Hence, we also have

$$\begin{aligned} 0 < \text{const} \leq \phi'(x) \leq \text{Const} \\ |\phi''(x)| \leq \text{Const} \end{aligned} \tag{25}$$

for $x \in [\frac{1}{10}, \frac{9}{10}]$. Performing the transformation, we find

$$W_x = W_z z_x = \begin{cases} W_z e^{-z}, & x \in]0, \frac{1}{10}], \\ W_z \phi'(\phi^{(-1)}(z)), & x \in]\frac{1}{10}, \frac{9}{10}[, \\ W_z e^z, & x \in [\frac{9}{10}, 1[, \end{cases} \tag{26}$$

and

$$W_{xx} = \begin{cases} W_{zz} e^{-2z} - W_z e^{-2z}, & x \in]0, \frac{1}{10}], \\ W_{zz} \phi'(\phi^{(-1)}(z))^2 + W_z \phi''(\phi^{(-1)}(z)), & x \in]\frac{1}{10}, \frac{9}{10}[, \\ W_{zz} e^{2z} + W_z e^{2z}, & x \in [\frac{9}{10}, 1[. \end{cases} \tag{27}$$

It is a lengthy , but elementary calculation to plug this transformation into the equation (23). The result is

$$W_t + \kappa_1(z)W_z + \kappa_2(z)W_{zz} + \frac{e^{rt}}{4c_2}\kappa_3(z)W_z^2 + \kappa_4(z) = 0. \tag{28}$$

We do not provide the functions κ_i explicitly, but we give their properties, which we will need in the sequel, i.e.

$$\begin{aligned} \kappa_i(z) &\in C^\infty(-\infty,\infty), & i &= 1,...4 \\ |\kappa_i(z)| &\le const., & z &\in [-\ln 10, \ln 10], i = 1,...4 \\ \kappa_i(z) &\ge const. > 0, & z &\in [-\ln 10, \ln 10], i = 2,...4. \end{aligned} \tag{29}$$

Finally, the boundary condition transforms to

$$W(z,T) = c_3\phi^{(-1)}(z) = c_3 \begin{cases} e^z, & z \le -\ln 10, \\ \phi^{(-1)}(z), & z \in]-\ln 10, \ln 10[, \\ 1 - e^{-z}, & z \ge \ln 10. \end{cases} \tag{30}$$

We want to apply Theorem V.8.1 in [19], and in order to get useful information from this theorem, it is helpful to transform also the dependent variable via $\tilde{W}(z,t) := \frac{W(z,t)}{h(z)}$, with $h(z) \in C^\infty(\mathbf{R})$ and

$$h(z) := \begin{cases} e^z, & z \le -\ln 10, \\ \text{strict. mon. increasing}, & z \in]-\ln 10, \ln 10[, \\ 1, & z \ge \ln 10. \end{cases} \tag{31}$$

Moreover, we choose $h(z)$ s.t. it fulfills

$$\begin{aligned} |h'(z)| &\le \text{const.}, \forall z \in \mathbf{R}, \\ |h''(z)| &\le \text{const.}, \forall z \in \mathbf{R}, \\ h(z) &\in [1/10, 1], z \in [-\ln 10, \ln 10]. \end{aligned} \tag{32}$$

We also introduce the “backwards time” $\tau := T - t$. Performing all these transformations gives after some elementary calculations instead of (28)

$$\tilde W_\tau - \tilde W_{zz}\cdot\begin{cases}\sigma^2\left(1-e^z\right)^2\\ \kappa_2(z)\\ \sigma^2\left(1-e^{-z}\right)^2\end{cases} + \tilde W_z^2\frac{e^{r(T-\tau)}}{4c_2}\cdot\begin{cases}1\\ h(z)\kappa_3(z)\\ \left(1-e^{-z}\right)e^{2z}\end{cases} + \tilde W\tilde W_z\frac{e^{r(T-\tau)}}{4c_2}\cdot\begin{cases}2\\ h'(z)\kappa_3(z)\\ 0\end{cases}$$
$$+\tilde W_z\cdot\begin{cases}-\beta(1-e^z)+\gamma+\sigma^2(1-e^z)^2-2\sigma^2(1-e^z)^2\\ -\kappa_1(z)-2(h'(z)/h(z))\kappa_2(z)\\ -\beta(1-e^{-z})+\gamma(e^z-1)-\sigma^2(1-e^{-z})^2\end{cases} + \tilde W^2\frac{e^{r(T-\tau)}}{4c_2}\cdot\begin{cases}1\\ ((h'(z))^2/h(z))\kappa_3(z)\\ 0\end{cases}$$
$$+\tilde W\cdot\begin{cases}-\beta(1-e^z)+\gamma+\sigma^2(1-e^z)^2-\sigma^2(1-e^z)^2\\ -\kappa_1(z)(h'(z)/h(z))-\kappa_2(z)(h''(z)/h(z))\\ 0\end{cases} - e^{-(T-\tau)}c_1\cdot\begin{cases}1\\ \phi^{(-1)}(z)/h(z)\\ 1-e^{-z}\end{cases} = 0 \quad (33)$$

with the initial condition

$$\tilde W(z,0)=\frac{c_3\phi^{(-1)}(z)}{h(z)} = c_3\begin{cases}1, & z\le-\ln 10,\\ \phi^{(-1)}(z)/h(z), & -\ln 10<z<\ln 10,\\ 1-e^{-z}, & z\ge\ln 10.\end{cases} \quad (34)$$

In this equation and in some of the following ones we adopt the convention that the first line holds always for $z \le -\ln 10$, the third line for $z \ge \ln 10$ and the second one for the area inbetween.

As announced earlier, we now apply Theorem V.8.1 in [19]. In order to do this, we match first some notation and find using V.(6.4) there

$$\frac{\partial a_1}{\partial \tilde W_z}(z,\tau,\tilde W,\tilde W_z)=\begin{cases}\sigma^2(1-e^z)^2,\\ \kappa_2(z),\\ \sigma^2(1-e^{-z})^2.\end{cases} \quad (35)$$

Moreover, $A(z,\tau,\tilde W,\tilde W_z)$ denotes in our case all the terms of the l.h.s. of equation (33), but the $\tilde W_\tau$ and the $\tilde W_{zz}$-term.

We now start to check the assumptions of Theorem V.8.1. Concerning assumption a), it is clear that our initial condition, defined by (34), is bounded and an element of $C^{2+\beta}(\Omega)$, for all bounded domains $\Omega\subset\mathbf{R}$.

We proceed with condition b) : Equation V.(8.6) there is satisfied by our (35) and the third line of (29). Concerning the condition on A, we find that in our case

$$A(z,\tau,\tilde W,0)\tilde W \ge -b_1|\tilde W|^3-b_2$$

holds for some positive constants b_1, b_2, because all the coefficients of $\tilde W^2, \tilde W$ and "1" in (33) are uniformly bounded. This means that the second condition of b) with $\Phi(\tau)=b_1\tau^2$ is fulfilled.

Coming to condition c), we first find

$$a_1(z,\tau,\tilde{W},\tilde{W}_z) = \tilde{W}_z \begin{cases} \sigma^2(1-e^z)^2, \\ \kappa_2(z), \\ \sigma^2(1-e^{-z})^2, \end{cases} \tag{36}$$

and, using V.(6.4) of [19], one gets

$$a(z,\tau,\tilde{W},\tilde{W}_z) = A(z,\tau,\tilde{W},\tilde{W}_z) + \frac{\partial a_1}{\partial \tilde{W}}\tilde{W}_z + \frac{\partial a_1}{\partial z} = A(z,\tau,\tilde{W},\tilde{W}_z) + \tilde{W}_z \begin{cases} -2\sigma^2(1-e^z)e^z, \\ \kappa_2'(z), \\ 2\sigma^2(1-e^{-z})^2 e^{-z}. \end{cases} \tag{37}$$

We have to check the conditions b) and c) of Theorem V.6.1 in [19] now. One easily sees that a_1 and a are continuous with respect to its variables and that a_1 is also differentiable.

Let now $Q_T := [-N_1, N_2] \times [0,T]$, for some (large) positive constants N_1, N_2. One first finds that the constants of inequality V.(6.9/1) are independent of Q_T in our case.

Concerning the inequality V.(6.9/2), we realize for later use the

Fact 1: The constants in inequality V.(6.9/2) depend on N_2 but not on N_1.

The reason for this are the e^{2z}, resp. the $(e^z - 1)$ terms in the coefficient of $\tilde{W}_z^2$, resp. $\tilde{W}_z$ of (33).

Finally, the Hölder continuity conditions of Theorem V.6.1c are easily seen to be satisfied. All in all, we can now apply Theorem V.8.1, giving us the following.

Assertion 1: There exists at least one solution $\tilde{W}(z,\tau)$ of the Cauchy problem (33)+(34) in $[0,T]\times\mathbf{R}$, which does not exceed a positive constant M in modulus and belongs to $C^{2+\beta,1+\beta/2}(Q_T)$ with $Q_T := [-N_1, N_2] \times [0,T]$ for some (large) positive constants N_1, N_2. Moreover, because of Fact 1 above, it is also an element of $C^{2+\beta,1+\beta/2}(]-\infty, N_2] \times [0,T])$.

Now, transforming back to the original variables, Assertion 1 provides almost Theorem 1, we just have to show continuity at $x = 0$. But this is easy: By Assertion 1, we have $|\tilde{W}(z,\tau)| \leq const.$ for all $(z,t) \in [-\infty, N_2] \times [0,T]$, giving $|W(z,\tau)| \leq \text{const.}h(z)$, or $|W(z,\tau)| \leq \text{const.}e^z$, for $z \leq -\ln 10$. Hence, finally $|W(x,\tau)| \leq \text{const.}x$, for $x \leq 1/10$. □

In a next step we will have to show that a solution $W(x,t)$ of (23), provided by Theorem 1, fulfills $W_x(x,t) \geq 0$ (we know this already for the value function, see Lemma 2). Once we know this, one gets that the optimizer u^* appearing in the HJB equation (15) is also nonnegative and given by the formula $u^* = \frac{e^{rt}}{2c_2}W_x$. This allows us in the proof of the verification theorem to restrict to nonnegative controls.

The method of our proof will be the following. We first consider an approximating optimization problem, analogous to the original one, but where we prescribe an artificial boundary condition at $x = 1 - \epsilon$. For this problem we do not have to show that the upper boundary is an entrance boundary. So a standard verification theorem will show that the solution of the HJB equation $W^{(\epsilon)}$ is the valuefunction $V^{(\epsilon)}$ of our approximating problem. If we choose the artificial boundary condition properly, a simple modification of Lemma 2 will show $V_x^{(\epsilon)} > 0$, hence $W_x^{(\epsilon)} > 0$. Going with epsilon to zero, will yield the desired result.

So let us start with the definition of the approximating problem

$$V^{(\epsilon)}(x,t) = \min_{u \in \mathcal{A}^\epsilon} \mathbf{E}_{t,x} \left[\int_t^{\tau^{(\epsilon)}} e^{-rs} \left[h(X(s), u(X(s),s)) \right] ds \right.$$
$$\left. + c_3 X(\tau^{(\epsilon)}) \mathbf{1}_{\{X(\tau^{(\epsilon)}) < 1-\epsilon\}} + \Psi^{(\epsilon)}(X(\tau^{(\epsilon)})) \mathbf{1}_{\{X(\tau^{(\epsilon)}) = 1-\epsilon\}} \right], \tag{38}$$

where $\tau^{(\epsilon)}$ is the stopping time $\tau^{(\epsilon)} := \inf\{s \geq t | X(s) = 0\} \wedge \inf\{s \geq t | X(s) = 1-\epsilon\} = \inf\{s \geq t | X(s) = 1-\epsilon\}$, with the convention $\inf \emptyset = T$ and accounting for the fact that $x = 0$ is a natural boundary (see Remark 1). Moreover, $\mathcal{A}^\epsilon := \{u(x,t) | u \text{ is Lipschitz w.r.t. } t \in [0,T] \text{ and } x \in [0, 1-\epsilon]\}$ are the admissible strategies. The state process is as in Lemma 1. For the artificial boundary condition we assume

$$\Psi^{(\epsilon)}(t) \in C^\infty([0,T]), \qquad (\Psi^{(\epsilon)})'(t) \leq 0. \tag{39}$$

We now consider the boundary value problem

$$\begin{aligned} W_t^{(\epsilon)} + [\beta x(1-x) - \gamma x] W_x^{(\epsilon)} + \sigma^2 x^2 (1-x)^2 W_{xx}^{(\epsilon)} + e^{-rt} c_1 x - \frac{e^{rt}}{4c_2} x (W_x^{(\epsilon)})^2 = 0, \\ W^{(\epsilon)}(x,T) = c_3 x, \\ W^{(\epsilon)}(0,t) = 0, \\ W^{(\epsilon)}(1-\epsilon, t) = \Psi^{(\epsilon)}(t). \end{aligned} \tag{40}$$

If we impose a compatibility condition on our boundary data at the point $(1-\epsilon, T)$, i.e., if we assume that the first line of (40) holds in this point, we get

$$(\Psi^{(\epsilon)})'(T) + O(\epsilon) - \gamma c_3 + e^{-rT} c_1 - \frac{e^{rT}}{4c_2} c_3^2 = 0$$

As $(\Psi^{(\epsilon)})'(T)$ should be nonpositive by (39), we impose for the following the

Assumption 1. *The constant c_3 in our model fulfills $c_3 < \tilde{c}_3$, where $\tilde{c}_3$ is the positive solution of the equation*

$$-\gamma c_3 + e^{-rT} c_1 - \frac{e^{rT}}{4c_2} c_3^2 = 0.$$

This assumption guarantees that one can find a function $\Psi^{(\epsilon)}$ fulfilling (39), as well as the compatibility condition, for ϵ small enough. As the compatibility condition is fulfilled now, one finds, using the beginning of chapter V.8 of [19], the following lemma.

Lemma 3. *The boundary value problem (40) has a solution in $C^{2+\beta, 1+\beta/2}([0, 1-\epsilon] \times [0,T])$, which converges - as $\epsilon \to 0$, after possibly extracting a subsequence - together with its derivatives on $]0,1[$ to $W(x,t)$, a solution provided by Theorem 1.*

We proceed with the announced verification result for the approximating problem.

Proposition 1. *A $C^{2+\beta,1+\beta/2}([0,1-\epsilon]\times[0,T])$ solution $W^{(\epsilon)}$ of the boundary value problem (BVP)*

$$\begin{aligned}\inf_{u\in\mathbf{R}}\left\{\mathcal{L}^u W^{(\epsilon)} + F(x,u)\right\} &= 0,\\ W^{(\epsilon)}(x,T) &= c_3 x,\\ W^{(\epsilon)}(0,t) &= 0,\\ W^{(\epsilon)}(1-\epsilon,t) &= \Psi^{(\epsilon)}(t).\end{aligned}$$

is the value function of problem (38).(Hence, there is only one such solution.)

Proof. Note that the BVP above is just another way of writing the system (40) ($\mathcal{L}^u$ stands for the generator of our state process).

We first remark that our assumption $u\in\mathcal{A}^\epsilon$ guarantees a unique strong solution of the underlying SDE. Applying Ito's formula to $W^{(\epsilon)}$, which is regular enough, gives

$$W^{(\epsilon)}(X(\tau^{(\epsilon)}),\tau^{(\epsilon)}) = W^{(\epsilon)}(x,t) + \int_t^{\tau^{(\epsilon)}} \mathcal{L}^u W^{(\epsilon)}(X(s),s)\ \mathrm{d}s + \int_t^{\tau^{(\epsilon)}} W_x^{(\epsilon)}(X(s),s)\sqrt{b(X(s))}\ \mathrm{d}\mathcal{B}(s).$$

Since σ and $W_x^{(\epsilon)}$ are bounded, the last integral is a true martingale and vanishes in the expectation. This gives

$$\begin{aligned}\mathbf{E}_{t,x}\left[\int_t^{\tau^{(\epsilon)}} F(X(s),u(s))\ \mathrm{d}s + W^{(\epsilon)}(X(\tau^{(\epsilon)}),\tau^{(\epsilon)})\right] &= W^{(\epsilon)}(x,t) + \mathbf{E}_{t,x}\left[\int_t^{\tau^{(\epsilon)}} F(X(s),u(s))\right.\\ &+ \left.\mathcal{L}^u W^{(\epsilon)}(X(s),s)\ \mathrm{d}s\right].\end{aligned}$$

By the HJB equation in the formulation of the proposition, the expectational value in the r.h.s. of the last equation is greater or equal to 0 for a general admissible strategy u, whereas it is equal to zero for the optimizer u^*. This concludes our proof. □

Checking the proof of Lemma (2), we note that it holds as well for our approximating optimization problem, because of the second assumption in (39). This gives $W_x^{(\epsilon)}(x,t) = V_x^{(\epsilon)}(x,t) > 0$ and, together with Lemma 3, we end up with

Corollary 1. *A solution $W(x,t)$ provided by Theorem 1 fulfills $W_x(x,t)\geq 0$ for $(x,t)\in[0,1[\times[0,T]$.*

We are now in a position to restrict the strategies, which are allowed, to the nonnegative ones.

Assumption 2. *The set of admissible strategies $u(\cdot,\cdot)\in\mathcal{A}$ for our basic optimization problem is given by*

$$\mathcal{A} := \{u(x,t)|u \text{ is Lipschitz w.r.t. } t\in[0,T] \text{ and locally Lipschitz w.r.t. } x\in[0,1[, u\geq 0\} \tag{41}$$

Finally we come to the main result of this section

Theorem 2. *A solution W given by Theorem 1 is the value function of problem (11)-(13) for $0 \le x < 1$ under Assumptions 1 and 2. (Hence, there is only one such solution.)*

Proof. Let $\tau^{(\epsilon)}$ be as it is defined after equation (38) and let $u \in \mathcal{A}$, then we get, as in the proof of Proposition 1,

$$\mathbf{E}_{t,x}\left[\int_t^{\tau^{(\epsilon)}} F(X(s),u(s))\,ds + W(X(\tau^{(\epsilon)}),\tau^{(\epsilon)})\right] \ge W(x,t).$$

Since $x = 1$ is entrance (see the discussion in Remark 1), we can define the stopping time τ as $\tau := \inf\{s \ge t | X(s) = 0\}$, with the convention $\inf \emptyset = T$. There is a natural boundary at $x = 0$ (see Remark 1), hence we get a.s. $\tau^{(\epsilon)} \to \tau = T$, for $\epsilon \to 0$. Since the state process $X(t)$ is continuous and $W(x,t)$ is continuous on $[0,1[\times[0,T]$, we conclude $W(X(\tau^{(\epsilon)}),\tau^{(\epsilon)}) \to W(X(\tau),\tau)$ a.s. By bounded convergence (W is bounded) and monotone convergence (ν is positive), we end up with

$$\mathbf{E}_{t,x}\left[\int_t^{T} F(X(s),u(s))\,ds + W(X(T),T,)\right] \ge W(x,t).$$

If we take instead of an arbitrary u the optimizer u^*, we get instead of the inequality an equality above, which concludes the proof. □

Remark 2. Let us finally remark that the existence of a strong solution of the underlying SDE (12) for our admissible strategies u_t follows by Exercise IX.2.10 of the monograph [23].

3 The stochastic open-loop optimal control problem

In this section we investigate the stochastic optimal control problem SOP without state observation, in which only (deterministic) open-loop controls, $u = u(t)$ are implementable. The only state-information available is the probability distribution of $X(0)$, represented by its density function $p_0(\cdot) : [0,1] \to \mathbb{R}_0^+$, so that $\mathbb{P}\left(a \le X(0) \le b\right) = \int_a^b p_0(x)\,\mathrm{d}x$ for any interval $0 \le a \le b \le 1$. This initial distribution can be based on statistical estimation or on subjective probabilistic views.

The stochastic optimal open-loop control problem, SOP, reads as

$$\mathcal{U}(p_0(\cdot)) = \min_u \mathbf{E}\left[\int_0^T e^{-rs}\left[F(X(s),u(s))\right]\mathrm{d}s + c_3 X(T)\right] \tag{42}$$

subject to

$$\mathrm{d}X(s) = a(X(s),u(s))\,\mathrm{d}s + \sqrt{b(X(s))}\,\mathrm{d}\mathcal{B}(s),$$

$X(0)$ is a random variable with probability density $p_0(x)$.

We embed SOP into a family of optimization problems

$$U(t) = \min_u \mathbf{E} \left[\int_t^T e^{-rs} \left[F(X(s), u(s)) \right] \, \mathrm{d}s + c_3 X(T) \right] \tag{43}$$

subject to

$$\mathrm{d}X(s) = a(X(s), u(s)) \, \mathrm{d}s + \sqrt{b(X(s))} \, \mathrm{d}\mathcal{B}(s), \tag{44}$$

$$X(t) \text{ is a random variable with (unconditional) probability density } p(x,t), \tag{45}$$

prametrized by time $t \in [0, T]$. $U(t)$ is the optimal rest value of SOP on $[t, T]$. With $p(x, 0) = p_0(x)$ we have $\mathcal{U}(p_0(\cdot)) = U(0)$. Moreover, if u^* is the optimal control function of SOP, we define the conditional value function as the expectation

$$U(x,t) = \mathbf{E}_{t,x} \left[\int_t^T e^{-rs} \left[F(X(s), u^*(s)) \right] \, \mathrm{d}s + c_3 X(T) \right] \tag{46}$$

3.1 Analyzing the open-loop problem with stochastic calculus

In this section we use stochastic calculus for deriving a sufficient condition for optimal solutions of the open-loop problem. As we will see, this condition is in some sense very similar to the HJB-condition but accounts for the fact that the states cannot be observed.

For technical reasons, in this subsection we restrict the set of admissible controls to the functions $u : [0, T] \to [0, \infty)$ which have bounded derivatives up to order 3.

Nonnegativity again is a natural assumption for the control in our context. Moreover it ensures that there is an entrance boundary at $x = 1$ (which can be ensured also by the weaker assumption $u(t) > -\gamma$). This fact is used in the proof of the following Theorem. Note however, that the arguments in the proof of Lemma 2 - showing the nonnegativity is automatically fulfilled in the optimum - can be applied also to the open loop problem.

Theorem 3. *For any admissible control $u(\cdot)$ (in the sense of this subsection) the PDE*

$$\mathcal{L}^{u(s)} Q\,(x,s) + e^{-rs} F(x, u(s)) = 0, \tag{47}$$
$$Q(x,T) = c_3 x \text{ for } x \in (0,1),$$

with $\mathcal{L}^u$ defined in (14), has a unique $C^2([0,1] \times [0,T])$-solution Q. Moreover, if an admissible control $\hat{u}$ and the related solution $\hat{Q}$ of (47) satisfy

$$\hat{u}(s) = \operatorname*{argmin}_u \mathbf{E} \left[\mathcal{L}^u \hat{Q}(x,s) + e^{-rs} F(x,u) \right] \tag{48}$$

for all $s \in [t, T]$, then $\hat{u}$ is an optimal control of SOP (43) and

$$\hat{Q}(x,s) = U(x,s). \tag{49}$$

This also leads to

$$\mathbf{E}\hat{Q}(X(s),s) = U(s). \tag{50}$$

Proof. Equation (47) is a linear second order equation with non-negative characteristic form ($b(x)\cdot\xi^2$ is obviously nonnegative for any $x \in [0,1]$ and $\xi \in \mathbb{R}$). Applying the Fichera-classification (see e.g. [8]) to the piecewise smooth boundary of our equation, we get $\Sigma_0 = \{0\} \times (0,T)$, $\Sigma_1 = (0,1) \times \{0\} \cup \{1\} \times (0,T)$, $\Sigma_2 = (0,1) \times \{T\}$ and $\Sigma_3 = \emptyset$. This shows that (47) is a well formulated first boundary-value problem (i.e. the boundary condition is formulated on $\Sigma_2 \cup \Sigma_3$). Observe now that due to the assumed properties of u the coefficients of (47) as well as the function F have bounded derivatives up to order 3 on the whole domain $[0,1]\times[0,T]$. Moreover the coefficients can be extended outside the domain with the same smoothness and keeping nonnegativity of the characteristic form (e.g., set $b(x) = \sigma^2 x^2(1-2x)$ for small negative x). The function $x \mapsto c_3 x$ is linear. Furthermore the intersection of any two of the sets $\Sigma_3, \Sigma_2, \Sigma_0 \cup \Sigma_1$ is nonempty. Using the result in [20] p 775, we get that there exist a solution[2] with bounded derivatives up to order 3. Altogether, this implies that the solution is in $W^{3,p}((0,1)\times(0,T))$ for any p>0.

The rectangle $[0,1]\times[0,T]$ has a Lipschitz boundary and we can apply Sobolev's embedding results (see e.g. [13], Theorem 1.14) to infer that Q can be represented (almost everywhere) by a $C^2([0,1]\times[0,T])$-function. This gives enough regularity for applying Ito's formula to the solution Q. Moreover, we have a natural boundary at $x=0$ and an entry boundary at $x=1$ (recall that $u(t)\geq 0$ by assumption) - both points are never reached if the process starts in the interior. Based on the dynamics (44) and using some starting point $X(t)=x_0$, we therefore get

$$Q(X(T),T) = Q(x_0,t) + \int_t^T \mathcal{L}^u Q(X(s),s)\ \mathrm{d}s + \int_t^T Q_x(X(s),s)\sqrt{b(X(s))}\ \mathrm{d}\mathcal{B}(s).$$

Since b and Q_x are bounded on $(0,1)$, the last integral is a true martingale and vanishes in expectation. This gives

$$\begin{aligned}\mathbf{E}_{t,x}\left[\int_t^T e^{-rs}F(X(s),u(s))\ \mathrm{d}s + Q(X(T),T)\right] = Q(x_0,t) + \mathbf{E}_{t,x}\left[\int_t^T e^{-rs}F(X(s),u(s))\right. \\ \left. + \mathcal{L}^u Q(X(s),s)\ \mathrm{d}s\right].\end{aligned} \tag{51}$$

If (47) holds, the expectation on the right-hand side vanishes, i.e.

$$\mathbf{E}_{t,x}\left[\int_t^T e^{-rs}F(X(s),u(s))\ \mathrm{d}s + c_3 X(T)\right] = Q(x_0,t). \tag{52}$$

Moreover, the start value x_0 was chosen arbitrarily so far. Applying now the assumption that $X(t)$

[2]Strictly speaking, the result is valid for equations that involve a further term $c(x)Q(x,t)$ in the left hand side, such that $c(x)$ is negative and bounded by a constant that is small enough. For parabolic PDEs however, this always can be achieved by a simple transformation of Q.

is distributed with density function $p(x,t)$ and taking expectation at both sides of 51, we get

$$\mathbf{E}\left[\int_t^T e^{-rs}F(X(s),u(s))\,\mathrm{d}s + Q(X(T),T)\right] = \mathbf{E}\left[Q(X(t),t)\right] + \int_t^T \mathbf{E}\left[e^{-rs}F(X(s),u(s)) + \mathcal{L}^{u(t)}Q(X(s),s)\right]\,\mathrm{d}s. \tag{53}$$

If now $u^*(t)$ is the minimizer (48), it minimizes the right-hand and hence also the left hand side of (53). Again the integral at the right hand side vanishes by (47) and in consequence

$$\mathbf{E}Q(X(t),t) = U(t). \tag{54}$$

Plugging u^* into (52), we also get

$$Q(x,t) = U(x,t). \tag{55}$$

□

We mention that restricting the set of admissible control to 3 times differentiable functions, as in this subsection, may be too strong. Moreover, while it is clear that condition (48) is equivalent to the optimality condition

$$u(t) = \frac{e^{rt}}{2c_2}\frac{\mathbf{E}[X(t)Q_x(X(t),t)]}{\mathbf{E}X(t)}, \tag{56}$$

it is not clear at this point, how the conditions given in Theorem 1 can be used for calculations, in particular how the expectations in (56) could be calculated. For this reason, in the next subsection, we employ a more general approach, which uses arguments from PDE-constraint optimization.

3.2 The optimal control problem for the related Fokker-Planck equation

Denote $D := [0,1]\times[0,T]$. Because $X(t)$ is a (Markovian) diffusion process, the objective function in (42) can be rewritten in terms of the transition density function $p : D \to [0,\infty)$. This leads to the equivalent formulation

$$\mathcal{U} = \min_u \iint_D e^{-rs}F(x,u(t))\,p(x,t)\;\mathrm{d}x\,\mathrm{d}s + \int_0^1 c_3\,x\,p(x,T)\,\mathrm{d}x \tag{57}$$

subject to

$$p_t(x,t) = -[a(x,u(t))p(x,t)]_x + \frac{1}{2}[b(x)p(x,t)]_{xx}, \tag{58}$$

$$p(x,0) = p_0(x), \qquad p(1,t) = 0, \tag{59}$$

where the subscripts t and x mean differentiation. This is the Kolmogorov-forward (or Fokker-Planck) equation of the diffusion process (see e.g. [9] p. 132).

We remind that according to the specification of the functions a and b in (6), we have

$$a(0,u) = 0, \quad a(1,u) = -(\gamma + u) < 0, \quad b(0) = b'(0) = b(1) = b'(1) = 0. \tag{60}$$

Then the boundary condition (59) is equivalent to the zero-flux condition at $x = 1$:

$$0 = a(1,u(t))\,p(1,t) - \frac{1}{2}\frac{\partial}{\partial x}[b(x)p(x,t)]_{x=1} = a(1,u(t))\,p(1,t).$$

This condition ensures conservation of the probability mass at the “entrance boundary" $x = 1$. A similar boundary condition is not needed at the "natural boundary" (in the Feller terminology, see e.g. [10], p. 119) $x = 0$, since the flux there is automatically equal to zero due to (60).

Introducing the notations

$$A(x) = \frac{1}{2}b(x), \quad B(x,u) = a(x,u) - \frac{1}{2}b_x(x), \quad C(x,u) = a_x(x,u) - \frac{1}{2}b_{xx}(x), \quad f(x,t) = 0, \tag{61}$$

we can rewrite (58) (for a given measurable control $u(t) \geq 0$ inserted in B and C) in the symmetric form

$$p_t - (Ap_x)_x + Bp_x + Cp = f. \tag{62}$$

Thus the problem (57) can be written in the following more general form:

$$\text{minimize } \Big\{J(u) := \iint_D h(x,t,u(t))\,p(x,t)\,\mathrm{d}x\,\mathrm{d}t + \int_0^1 g(x)\,p(x,T)\,\mathrm{d}x\Big\}, \tag{63}$$

where in our particular case $h(x,t,u) = e^{-rt}F(x,u)$ and $g(x) = c_3 x$.

In fact, in the analysis below we do not use the particular forms of A, B, C, and f. We consider the more general problem (63) subject to (62) and (59). Admissible controls are all measurable and bounded functions $u : [0,T] \to [0,\bar{u}]$, where $\bar{u} \geq 0$ is an upper bound for the control.

For this more general problem we assume that the functions A, B, C, h, and f (where A depends only on x, B and C depend on (x,u), f depends on (x,t), h depends on (x,t,u)) and the derivatives A_x, A_{xx}, B_x, B_u, C_u, and h_u are Lipschitz continuous; there are constants $\hat{c}_1$ and $\hat{c}_2$ such that the following inequalities are fulfilled for every $x \in [0,1]$ and $u \in [0,\bar{u}]$:

$$\begin{aligned} &A(0) = A(1) = A'(0) = A'(1) = 0, \quad B(1,u) < 0, \\ &A(x) \geq \hat{c}_1 x^2, \quad |B(x,u)| + |B_u(x,u)| + |B_{uu}(x,u)| \leq \hat{c}_2 x. \end{aligned} \tag{64}$$

Clearly, the last conditions are fulfilled for our original SIS problem due to (61) and (6). Some of them may be weakened, but we do not aim for generality in this paper.

Below we briefly introduce some preliminary material, which is not quite standard and seems not to be available in the literature in the form that is needed in the optimal control context. The reason is that the function A in the elliptic term of (62) degenerates (even together with its first derivative) on the boundary $x \in \{0,1\}$, which requires special care.

Denote

$$V := \{v : (0,1] \to \mathbb{R} : \ v \text{ is locally absolutely continuous, } v \in L_2(0,1), \ x\, v_x \in L_2(0,1)\}.$$

Here v_x is the derivative of v, which exists almost everywhere. With the scalar product

$$\langle v, q\rangle := \int_0^1 [v(x)q(x) + x^2 v_x(x) q_x(x)]\, \mathrm{d}x + v(1)q(1), \qquad v,\, q \in V,$$

V is a Hilbert space. Denote by V^* its dual space, and by $\langle v^*, v\rangle_{V^*,V}$ – the pairing of V^* and V.

For a Banach space Y we denote by $L_2(0,T;Y)$ the space of all functions $y : (0,T) \to Y$ such that

$$\|y\|_{L_2(0,T;Y)} := \left(\int_0^T \|y(t)\|^2 \,\mathrm{d}t\right)^{1/2} < \infty.$$

It is said that $y \in L_2(0,T;Y)$ has a weak derivative z, written later as $\dot y = z$, if

$$\int_0^T \dot\varphi(t) y(t)\,\mathrm{d}t = -\int_0^T \varphi(t) z(t)\,\mathrm{d}t \qquad \forall\, \varphi \in C_c^\infty(0,T),$$

where the integrals are in the sense of Bochner. In particular, we will use the spaces $L_2(0,T;V)$, $L_2(0,T;V^*)$ and

$$W := \{p \in L_2(0,T;V) : \ \dot p \in L_2(0,T;V^*)\}.$$

It is known (see e.g. [26, Theorem 3.10]) that every element of W actually belongs to $C([0,T];V)$ (defined similarly as $L_2(0,T;V)$ but using the uniform norm), and for every two elements $p,\, \lambda \in W$ the formula for integration by parts is valid:

$$\int_0^T \langle \dot p(t), \lambda(t)\rangle_{V^*,V}\,\mathrm{d}t = \langle p(T), \lambda(T)\rangle_{L_2(0,1)} - \langle p(0), \lambda(0)\rangle_{L_2(0,1)} - \int_0^T \langle \dot\lambda(t), p(t)\rangle_{V^*,V}\,\mathrm{d}t. \tag{65}$$

In the space $V \times V$ we define the following bi-linear functional depending on $u \geq 0$:

$$\mathbf{B}(q,v;u) := \int_0^1 [A(x) q_x v_x + B(x,u) q_x v + C(x,u)\, q\, v]\,\mathrm{d}x - B(1,u)\, q(1) v(1).$$

From the definition of V and the properties of A and B it is easy to verify that $\mathbf{B}$ is well defined. Indeed, using (64) we have

$$\|A q_x v_x\|_{L_1} \leq \frac{1}{2}\|A_{xx}\|_{L_\infty}\, \|x q_x\|_{L_2}\, \|x v_x\|_{L_2} \leq \frac{1}{2}\|A_{xx}\|_{L_\infty}\, \|q\|_V\, \|v\|_V,$$

$$\|B(x,u) q_x v\|_{L_1} \leq \|B/x\|_{L_\infty}\, \|x q_x\|_{L_2}\, \|v\|_{L_2} \leq \hat c_2 \|q\|_V\, \|v\|_V.$$

Definition. A function $p : D \to \mathbb{R}$ is a *weak solution* of equation (62) with side conditions (59) (for a fixed admissible control $u(\cdot)$) if $p \in W$,

$$\langle \dot{p}(t), v\rangle_{V^*,V} + \mathbf{B}(p(t), v; u(t)) = \langle f(t), v\rangle_{L_2(0,1)} \quad \text{for every } v \in V \text{ and a.e. } t \in (0,T),$$

and $p(\cdot, 0) = p_0(\cdot)$.

We mention that, thanks to the inequality $B(1, u) \neq 0$, the last term of $\mathbf{B}$ ensures satisfaction of the boundary condition $p(1, t) = 0$.

Now, let us fix a reference admissible control u, and introduce the so-called adjoint equation:

$$\dot{\lambda} + (A\lambda_x)_x + (B\lambda)_x - C\lambda = -h, \tag{66}$$

with the end-condition

$$\lambda(x, T) = g(x). \tag{67}$$

Notice that by changing the time variable $t \longrightarrow T - t$ we obtain the following equation for $\mu(x,t) = \lambda(x, T-t)$:

$$\dot{\mu} - (A(x)\mu_x)_x - (B(x, u(T-t))\mu)_x + C(x, u(T-t))\mu = h(x, T-t, u(T-t)).$$

This equation has a similar form as (62) with the difference that the transport term is now $-B\mu$, and since $-B(1, u(t)) > 0$ and $A(1) = 0$, there is no need of a boundary condition at $x = 1$ (“natural boundary" in the Feller terminology, see e.g. [10], p. 119).

Correspondingly, the definition of a solution is a slight modification of the previous one. Namely, a solution of (66), (67) is any element $\lambda \in W$ such that

$$\langle \dot{\lambda}(t), v\rangle_{V^*,V} + \mathbf{D}(\lambda(t), v; u(t)) = -\langle h(\cdot, t, u(t)), v\rangle_{L_2(0,1)} \quad \forall\, v \in V \text{ and a.e. } t \in (0,T),$$

and $\lambda(T) = g$, where

$$\mathbf{D}(q, v; u) := \int_0^1 [-A(x)q_x v_x + (B(x,u)q)_x v - C(x,u)qv]\, \mathrm{d}x, \quad q,\, v \in V,$$

Proposition 2. *Given any measurable selection $u(\cdot)$ of the interval $[0, \bar{u}]$, both systems (62), (59) and (66), (67) have unique solutions, p and λ, respectively, in the space W. Moreover, there exists a constant c (depending on the functions A, B, and C, and the numbers $\bar{u}$ and T, but not on f, p_0, and the particular choice of the measurable selection u of $[0, \bar{u}]$), such that*

$$\max_{t\in[0,1]} \|p(\cdot, t)\|_{L_2(0,1)} + \|p\|_{L_2(0,T;V)} \le c \left(\|p_0\|_{L_2(0,1)} + \|f\|_{L_2(D)} \right).$$

The proof of this proposition adopts known approaches combining [21][Chapter 1] and [26, Chapter 3]. It is long but does not involve new ideas, therefore we do not present it here. Issues of regularity of the solution of problem (62), (59) will not be discussed (and used) in order to keep the presentation as focused as possible.

We introduce the notation

$$H(x,t,p,q,u,\mu) := h(x,t,u)p - (B(x,u)q + C(x,u)p)\,\mu.$$

The following is the main result of this section.

Theorem 4. *Let $\hat u$ be an admissible control and let $\hat p$ and λ be the corresponding solutions of the primal equation (62), (59) and the adjoint equation (66), (67), respectively.*

(i) Then for every number $M \ge 0$ there exists a constant $c(M)$ such that the inequality

$$\left| J(u) - J(\hat u) - \int_0^T J'(\hat u)(t)\,(u(t) - \hat u(t))\,\mathrm{d}t \right| \le c(M)\|u - \hat u\|^2_{L_2(0,T)}$$

is fulfilled for every measurable selection $u(\cdot)$ of $[0, \bar u + M]$. Here $J'(\hat u) \in L_\infty(0,T)$ is defined as

$$J'(\hat u)(t) = \int_0^1 H_u(x,t,\hat p(x,t),\hat p_x(x,t),\hat u(t),\lambda(x,t))\,\mathrm{d}x.$$

(ii) If $\hat u$ is an optimal control, then the following maximum condition holds:

$$\int_0^1 H(x,t,\hat p(x,t),\hat p_x(x,t),\hat u(t),\lambda(x,t))\,\mathrm{d}x = \min_{u\in[0,\bar u]} \int_0^1 H(x,t,\hat p(x,t),\hat p_x(x,t),u,\lambda(x,t))\,\mathrm{d}x.$$

We mention that property (i) directly implies Fréchet differentiability of the mapping $L_\infty(0,T) \ni u \mapsto J(u)$ with the L_∞-representative $J'(u)(\cdot)$ of the derivative. However, property (i) is stronger (as resembling some properties of an L_2-derivative), and this is essential for application of the gradient projection method for solving the optimal control problem at hand.

Proof. Let M be arbitrarily fixed and u be an arbitrary measurable selection of $[0, \bar u + M]$. Denote $\varepsilon := \|u - \hat u\|_{L_2(0,T)}$. Let p be the corresponding solution of equations (62), (59). Further, we abbreviate

$$\begin{aligned} \hat B(x,t) := B(x,\hat u(t)), \quad \hat C(x,t) := C(x,\hat u(t)), \quad \hat h(x,t) := h(x,t,\hat u(t)), \\ \Delta p = p - \hat p, \quad \Delta B(x,t) := B(x,u(t)) - B(x,\hat u(t)), \\ \Delta C(x,t) := C(x,u(t)) - C(x,\hat u(t)), \quad \Delta h(x,t) := h(x,t,u(t)) - h(x,t,\hat u(t)). \end{aligned}$$

Then the following equations are satisfied:

$$\begin{aligned} &\Delta p_t - (A\,\Delta p_x)_x + B\,\Delta p_x + C\,\Delta p = -\Delta B(x,t)\,\hat p_x(x,t) - \Delta C(x,t)\,\hat p(x,t), \qquad (68)\\ &\Delta p(x,0) = 0, \quad \Delta p(1,x) = 0. \end{aligned}$$

We shall apply to the above equation Proposition 2 with $p_0 = 0$ and $f = -\Delta B(x,t)\,\hat{p}_x(x,t) - \Delta C(x,t)\,\hat{p}(x,t)$. For this we need the following estimations that use the assumptions about A and B, the definition of the space V and the fact that $p \in C([0,T];V)$. First,

$$\iint_D |\Delta C(x,t)\hat{p}(x,t)|^2 \,\mathrm{d}x\,\mathrm{d}t \leq \int_0^T \hat{c}_4 |u(t) - \hat{u}(t)|^2 \int_0^1 |\hat{p}(x,t)|^2 \,\mathrm{d}x\,\mathrm{d}t$$
$$= \hat{c}_4 \int_0^T |u(t) - \hat{u}(t)|^2 \, \|\hat{p}(\cdot,t)\|^2_{L_2(0,1)} \,\mathrm{d}t \leq \hat{c}_4 \|\hat{p}\|^2_{C([0,T];V)}\, \varepsilon^2,$$

where $\hat{c}_4$, and also $\hat{c}_5$, etc., below, are constants (independent of ε). Second, for some measurable function $B^u(x,t) \in B_u(x,[u(t),\hat{u}(t)])$ we have

$$\iint_D |\Delta B(x,t)\hat{p}_x(x,t)|^2 \,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_0^1 |B^u(x,t)(u(t) - \hat{u}(t))|^2 \, |\hat{p}_x(x,t)|^2 \,\mathrm{d}x\,\mathrm{d}t$$
$$\leq \int_0^T \int_0^1 |B^u(x,t)/x|^2 \; |u(t) - \hat{u}(t)|^2 \, \|\hat{p}(\cdot,t)\|^2_V \,\mathrm{d}x\,\mathrm{d}t \;\leq\; c_3^2 \|\hat{p}\|^2_{C([0,T];V)}\, \varepsilon^2.$$

The obtained estimation of $\|f\|_{L_2(D)}$, together with Proposition 2, implies existence of a constant $\hat{c}_5$ which is independent of ε and such that

$$\|\Delta p\|_{L_2(0,T;V)} \leq \hat{c}_5 \varepsilon. \tag{69}$$

Now, we rewrite equation (68) in the more detailed form

$$\Delta p_t - (A\,\Delta p_x)_x + \hat{B}\,\Delta p_x + \hat{C}\,\Delta p = -\Delta B\,\hat{p}_x - \Delta C\,\hat{p} + \gamma,$$

where $\gamma = -\Delta B\,\Delta p_x - \Delta C\,\Delta p$. We notice that

$$\hat{c}_5 \varepsilon \geq \|\Delta p\|_{L_2(0,T;V)} \geq \|\Delta p\|_{L_2(D)} + \|x \Delta p_x\|_{L_2(D)}.$$

Thus

$$\|\gamma\|_{L_2(D)} \leq \|\Delta B\,\Delta p_x\|_{L_2(D)} + \|\Delta C\,\Delta p\|_{L_2(D)} \leq \|(\Delta B^u/x)(u - \hat{u})\,(x\Delta p_x)\|_{L_2(D)} + \hat{c}_6 \varepsilon^2 \leq \hat{c}_7 \varepsilon^2.$$

Let λ be the solution of the adjoint equation (66), (67) corresponding to $\hat{u}$ and $\hat{p}$. Since $\lambda(t) \in V$ for a.e. t, we have

$$\langle \dot{\Delta p}(t), \lambda(t) \rangle_{V,V^*} + \mathbf{B}(\Delta p(t), \lambda(t); \hat{u}(t)) = -\langle \Delta B(t)\,\hat{p}_x(t) + \Delta C(t)\,\hat{p}(t), \lambda(t) \rangle_{L_2(0,1)} + \langle \gamma(t), \lambda(t) \rangle_{L_2(0,1)}.$$

Using the integration by part formula (65) and the relations $\Delta p(x,0) = 0$, $\lambda(x,T) = g(x)$ we obtain that

$$- \int_0^T \langle \dot{\lambda}(t), \Delta p(t) \rangle_{V^*,V} \, \mathrm{d}t + \langle g, \Delta p(T) \rangle_{L_2(0,1)} + \int_0^T \mathbf{B}(\Delta p(t), \lambda(t); \hat{u}(t)) \, \mathrm{d}t$$
$$= - \int_0^T \langle (\Delta B(t) \hat{p}_x(t) + \Delta C(t) \hat{p}(t)), \, \lambda(t) \rangle_{L_2(0,1)} \, \mathrm{d}t + \zeta_1, \tag{70}$$

where $|\zeta_1| \le \hat{c}_8 \varepsilon^2$ and $\hat{c}_8$ is independent of ε and the particular choice of the selection u of $[0, \bar{u} + M]$.

The term $\int_0^T \mathbf{B}(\Delta p(t), \lambda(t); \hat{u}(t)) \, \mathrm{d}t$ contains the integral $\iint_D \hat{B} \lambda \Delta p_x \, \mathrm{d}x \, \mathrm{d}t$, in which we want to shift the differentiation from Δp to λ. Applying integration by parts needs a special care, due to the (a priori) possible unboundedness of Δp and λ at $x = 0$. For this reason we first show that for every $t \in [0, T]$ there exists a sequence $x_k \in (0, 1]$, $x_k \to 0$, such that $x_k \Delta p(x_k, t) \lambda(x_k, t)$ converges to zero. Indeed, if for some t such a sequence does not exist then there exist $\alpha > 0$ and $\beta > 0$ such that $x |\Delta p(x, t) \lambda(x, t)| \ge \alpha$ for any $x \in (0, \beta]$. Since $\Delta p(\cdot, t), \, \lambda(\cdot, t) \in L_2(0, 1)$, the function $|\Delta p(\cdot, t) \, \lambda(\cdot, t)|$ is integrable, which contradicts the inequality $|\Delta p(\cdot, t) \, \lambda(\cdot, t)| \ge \alpha / x$ on $(0, \beta]$.

Then, due to the absolute integrability of $\hat{B} \lambda \Delta p_x$, $\hat{B} \lambda_x \Delta p$ (already shown) and $\hat{B}_x \lambda \Delta p$ with respect to x, we have (skipping the fixed arguments t and $\hat{u}(t)$ in the notations)

$$\int_0^1 \hat{B}(x) \lambda(x) \Delta p_x(x) \, \mathrm{d}x = \lim_k \int_{x_k}^1 \hat{B}(x) \lambda(x) \Delta p_x(x) \, \mathrm{d}x$$
$$= \lim_k \left[\hat{B}(1) \lambda(1) \Delta p(1) - \hat{B}(x_k) \lambda(x_k) \Delta p(x_k) - \int_{x_k}^1 [B_x(x) \lambda(x) \Delta p(x) + [B(x) \lambda_x(x) \Delta p(x)] \, \mathrm{d}x \right]$$
$$= \hat{B}(1) \lambda(1) \Delta p(1) - \int_0^1 [B_x(x) \lambda(x) \Delta p(x) + [B(x) \lambda_x(x) \Delta p(x)] \, \mathrm{d}x,$$

where we also use the inequality $|B(x, u)| \le \hat{c}_2 x$ in (64). Substituting the obtained expression in (70) we obtain the equality

$$- \int_0^T \langle \dot{\lambda}(t), \Delta p(t) \rangle_{V^*,V} \, \mathrm{d}t + \int_0^1 \lambda^T(x) \Delta p(x, T) \, \mathrm{d}x + \int_0^T \mathbf{D}(\lambda(t), \Delta p(t); t, \hat{u}(t)) \, \mathrm{d}t$$
$$= - \iint_D (\Delta B(x, t) \hat{p}_x(x, t) + \Delta C(x, t) \hat{p}(x, t)) \, \lambda(x, t) \, \mathrm{d}x \, \mathrm{d}t + \zeta_1.$$

Having in mind the definition of solution of the adjoint equation and the fact that $\Delta p(t) \in V$, we obtain that

$$\int_0^1 g(x) \Delta p(x, T) \, \mathrm{d}x = - \iint_D \hat{h}(x, t) \Delta p(x, t) \, \mathrm{d}x \, \mathrm{d}t \tag{71}$$
$$- \iint_D (\Delta B(x, t) \hat{p}_x(x, t) + \Delta C(x, t) \hat{p}(x, t)) \, \lambda(x, t) \, \mathrm{d}x \, \mathrm{d}t + \zeta_1.$$

Now consider the difference

$$J(u) - J(\hat{u}) = \iint_D [h(x,t,u(t))\, p(x,t) - h(x,t,\hat{u}(t))\, \hat{p}(x,t)]\, \mathrm{d}x\, \mathrm{d}t + \int_0^1 g(x)\, \Delta p(x,T)\, \mathrm{d}x,$$

which, using the properties of h and (69), can be represented (skipping the arguments (x,t)) as

$$\begin{aligned} J(u) - J(\hat{u}) = & \iint_D [\hat{h}\, \Delta p + \Delta h\, \hat{p} + \Delta h\, \Delta p]\, \mathrm{d}x\, \mathrm{d}t + \int_0^1 g(x)\, \Delta p(x,T)\, \mathrm{d}x \\ = & \iint_D [\hat{h}\, \Delta p + \Delta h\, \hat{p}]\, \mathrm{d}x\, \mathrm{d}t + \int_0^1 g(x)\, \Delta p(x,T)\, \mathrm{d}x + \zeta_2, \end{aligned}$$

where $|\zeta_2| \le \hat{c}_9 \|u - \hat{u}\|^2_{L_2(0,T)} = \hat{c}_9\, \varepsilon^2$. Using (71), we obtain that

$$\begin{aligned} J(u) - J(\hat{u}) &= \iint_D [\hat{h}\, \Delta p + \Delta h\, \hat{p}]\, \mathrm{d}x\, \mathrm{d}t - \iint_D \hat{h}\, \Delta p\, \mathrm{d}x\, \mathrm{d}t \\ &\quad - \iint_D (\Delta B(x,t) \hat{p}_x(x,t) + \Delta C(x,t) \hat{p}(x,t))\, \lambda(x,t)\, \mathrm{d}x\, \mathrm{d}t + \zeta_1 + \zeta_2 \\ &= \iint_D [H(x,t,\hat{p}(x,t), \hat{p}_x(x,t), u(t), \lambda(x,t)) \\ &\qquad - H(x,t,\hat{p}(x,t), \hat{p}_x(x,t), \hat{u}(t), \lambda(x,t))]\, \mathrm{d}x\, \mathrm{d}t + \zeta_1 + \zeta_2. \end{aligned} \tag{72}$$

Using the abbreviation $\hat{H}(x,t,u) := H(x,t,\hat{p}(x,t), \hat{p}_x(x,t), u, \lambda(x,t))$ we represent in a standard way

$$\begin{aligned} & \hat{H}(x,t,u(t)) - \hat{H}(x,a,\hat{u}(t)) \\ = & \hat{H}_u(x,t,\hat{u}(t))(u(t) - \hat{u}(t)) + \int_0^1 [\hat{H}_u(x,t,\hat{u}(t) + s(u(t) - \hat{u}(t))) - \hat{H}_u(x,t,\hat{u}(t))]\, \mathrm{d}s\, (u(t) - \hat{u}(t)). \end{aligned}$$

Combining this with (72) and using the particular form of the function H and again the properties (64) in a similar way as above, we obtain the representation in part (i) of the theorem and the boundedness of the function $J'(\hat{u})(t)$.

Now let $\hat{u}$ be an optimal control. Then applying (72) to controls of the form

$$u(t) = \begin{cases} v & \text{if } t \in [\tau, \tau + \varepsilon'], \\ \hat{u}(t) & \text{if } t \notin [\tau, \tau + \varepsilon']\}, \end{cases}$$

where $v \in [0, \bar{u}]$, $\tau \in (0,T)$, and $\varepsilon' > 0$ is so small that $\tau + \varepsilon' \le T$ and $4\bar{u}^2 \varepsilon' \le \varepsilon$ (so that $\int_\tau^{\tau+\varepsilon'} |v - \hat{u}(t)|\, \mathrm{d}t \le \varepsilon$), one can prove the second claim of the theorem in a standard way. □

The second part of Theorem 4 gives a Pontryagin type necessary optimality condition for the problem (63), (62), (59). The first part provides a formula for calculation of the derivative of the objective functional with respect to the control function in the space $L_2(0,T)$, in directions that are bounded. This derivative can be used for computing an approximate solution by using the gradient projection method, as we do it in our numerical experiments in Section 4 (although the justification needs a lot more work that we do not present in this paper).

We mention that with the particular specifications for our SIS model the adjoint equation (66) (by substituting the definitions (61)) takes the form

$$-\lambda_t = a(x,u)\lambda_x + \frac{1}{2}b(x)\lambda_{xx} + e^{-rt}F(x,t,u). \tag{73}$$

If we apply Theorem 4 to the original problem (57)-(59), respectively (42), we can see a very close relation to Theorem 1.

Corollary 2. *Let the assumptions of Theorem 4 be fulfilled with optimal* $\hat{p}, \hat{u}, \hat{\lambda}$. *Moreover let* $\hat{X}(s)$ *denote the process defined by*

$$\mathrm{d}\hat{X}(t) = a(\hat{X}(t), \hat{u}(t))\,\mathrm{d}t + \sqrt{b(\hat{X}(t))}\,\mathrm{d}\mathcal{B}(t). \tag{74}$$

Then the equations

$$\begin{aligned} \mathcal{L}^{\hat{u}(s)}\hat{\lambda}\,(x,s) + e^{-rs}F(x,\hat{u}(s)) = 0, \\ \hat{\lambda}(x,T) = c_3 x \text{ for } x \in (0,1), \end{aligned} \tag{75}$$

are fulfilled and

$$\hat{u}(s) = \underset{u}{\operatorname{argmin}}\, \mathbf{E}\left[\mathcal{L}^u\hat{\lambda}(\hat{X}(s),s) + e^{-rs}F(\hat{X}(s),u)\right] \tag{76}$$

holds for all $s \in [t,T]$. *(We remind that the notation* $\mathcal{L}^u$ *is introduced in (14)).*

Proof. Clearly, the adjoint equation (73) for $\hat{u}, \hat{\lambda}$ can be reformulated as (74). Observe now that (using the definitions (61) and applying integration by parts) we have (without denoting the dependencies on (x,s))

$$\begin{aligned} &\int_0^1 H(x,s,\hat{p}(x,s),\hat{p}_x(x,s),\hat{u}(s),\hat{\lambda}(x,s))\,\mathrm{d}t \\ &\quad = \int_0^1 \left[h(x,s,\hat{u}(s)) + \left(a(x,\hat{u}(s)) - \frac{1}{2}b_x(x)\right)\hat{\lambda}_x(x,s)\right]\hat{p}(x,s)\,\mathrm{d}s \\ &\quad = \mathbf{E}\left[h(\hat{X}(s),s,\hat{u}(s)) + \left(a(\hat{X},s),\hat{u}(s)) - \frac{1}{2}b_x(\hat{X}(s))\right)\hat{\lambda}_x(\hat{X}(s),s)\right]. \end{aligned} \tag{77}$$

Because $b(\cdot)$ and $\hat{\lambda}_t$ do not depend on u, minimizing (77) is equivalent to minimizing (76). ☐

This shows that the results in theorems 4 and 1 are fully consistent with each other: in fact, (75) and (76) are the main conditions in Theorem 3.

4 The role of information in the stochastic SIS optimal control problem: a numerical study

In this section we investigate what improvement in the decision performance is brought by the availability of exact measurements of the current state. The numerical analysis is based on the results in sections 2 and 3, where state-feedback (for the SFP problem) or open-loop controls (SOP) are involved, respectively. In addition, we compare these performance assessments with the ones obtained by using the deterministic model (DP), assuming that the stochastic model better represents the reality.

In order to ensure comparability we take a viewpoint immediately before $X(0) = x_0$ is realized and assume that $X(0)$ is distributed according to the density $p_0(x)$ as assumed in Section 3. The expectation

$$\mathcal{W}(p_0(\cdot)) = \mathbf{E}\left[V(X(0), 0)\right] = \int_0^1 V(x)p_0(x)\,\mathrm{d}x,$$

where V was defined for SFP in (11), then gives the minimal expected cost for a decision maker, who is able to observe all values $X(t)$ and to calculate state-dependent strategies. This expectation can be compared with the optimal value $\mathcal{U}(p_0(\cdot))$, i.e. the minimal expected cost for a decision maker who is not able to observe any values of the process $X(t)$, see (42).

The *value of observation*

$$VO = \mathcal{U}(p_0(\cdot)) - \mathcal{W}(p_0(\cdot))$$

expresses the value of the additional information, achieved by observing the state values over time. Clearly, VO is positive, because restricting $u(x,t)$ to u(t) puts an additional constraint on the original optimization problem. In our case the difference can be interpreted as the (maximal) amount of money, a decision maker without the possibility to observe the process $X(t)$ is willing to pay in order to acquire full information about $X(t)$ at any time t. In the context of disease modeling, VO helps to assess the meaningfulness of improved reporting schemes: a small value speaks against improved efforts, whereas large values suggest that investment in information will pay off later. The idea of VO is closely related to the notion of *value of information*, where $\mathcal{W}(p_0(\cdot))$ is calculated as the expectation of a clairvoyant's optimal value. The value of information, introduced in [22], has a long history in stochastic optimization, see e.g. also [25].

When considering optimal control in a stochastic context, one may ask the question, whether the substantial extra effort for using SDE and/or PDE constraints is worthwile. In our case, the deterministic SIS model (2) could be easily reduced to the deterministic control problem (2), where the starting value is chosen as $x_0 = \mathbf{E}\left[X(0)\right] = \int_0^1 xp_0(x)\,dx$. In this way one replaces distributions and stochastic processes by (instantaneous) expectations and gets a simpler, ODE constraint problem, which can be treated by classical methods. The question then is, how much (in terms of objective value) we loose if the system is in fact stochastic (according to either (11) or (57)), but decisions are based on the deterministic model (2). We therefore calculate the optimal control $u_d(t)$ for (2) and plug it into the Fokker-Planck equation of the stochastic system, which leads to a transition density $p_d(x,t)$ for a decision maker who faces the stochastic system but nevertheless uses the deterministic model:

$$\begin{aligned}\frac{\partial p_d(x,t)}{\partial t} &= -\frac{\partial}{\partial x}a(x,u_d(t))p_d(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x)p_d(x,t)\right],\\ p_d(1,t) &= 0,\\ p_d(x,0) &= p_0(x).\end{aligned}$$

The cost for this strategy is given by

$$\mathcal{R}(p_0(\cdot)) = \int_t^T \int_0^1 e^{-rs}F(X(s),u(s))p_d(x,s)\,\mathrm{d}x\,\mathrm{d}s + \int_0^1 c_3\,x\,p_d(x,T)\,\mathrm{d}x.$$

It is then possible to compare $\mathcal{R}$,$\mathcal{U}$ and $\mathcal{W}$. Clearly we have $\mathcal{R}(p_0(\cdot)) \geq \mathcal{U}(p_0(\cdot)) \geq \mathcal{V}(p_0(\cdot))$. In particular the difference

$$VSOL = \mathcal{R}(p_0(\cdot)) - \mathcal{U}(p_0(\cdot))$$

expresses the *value of the stochastic open loop solution* (57), whereas

$$VSCS = \mathcal{R}(p_0(\cdot)) - \mathcal{W}(p_0(\cdot))$$

is the value of the *stochastic closed loop solution* (11).

In the following example the value function $V(x,t) = W(x,t)$ of the stochastic control problem (11) was calculated by solving the HJB equation (15) numerically. The related optimal control $u(x,t)$ is then calculated from (22).

For finding the optimal solution of the Fokker-Planck system (57) we used the Fréchet derivative J' given in Theorem 4 and applied the gradient projection method described e.g. in [26], Section 5.9.1 (see also [16] for convergence results). All partial differential equations (the HJB equation as well as all PDEs used within the gradient projection descent algorithm) were solved by the numerical method of lines (see e.g. [24]). Finally, we solved the deterministic problem (2) by the steepest descent method for ODE constraint control problems, see e.g. [17] p. 334.

All the algorithms were implemented in Wolfram Mathematica$^{\mathrm{TM}}$.

In the numerical example we choose $\gamma = 1$ and $\beta = 4.5$ and time is measured in weeks. This setup may be interpreted as a disease where the mean length of infection is one week, whereas each infected person infects 4.5 persons in the mean. In the deterministic model without control (2) such a setup describes a disease which does not die out (the basic reproduction numer $R = 4.5$) and has an endemic state (stationary point)

$$\bar{x} = 1 - \frac{\gamma}{\beta} = 0.\dot{7}.$$

For the volatility parameter σ we consider two cases: low volatility with $\sigma = 0.5$ and high volatility with $\sigma = 2.7$. In all cases we plan over a horizon of $T = 8$ weeks and the cost parameters

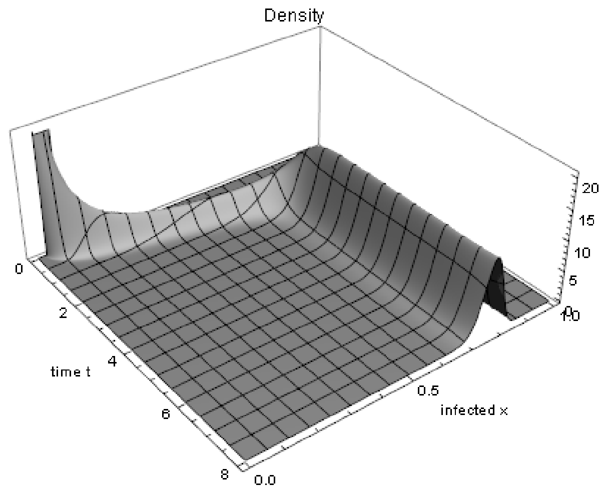

Figure 1: Low volatility case $\sigma = 0.5$: Transition density $p(x,t)$ for the optimal solution of the Fokker-Planck constraint problem (57).

are chosen as $c_1 = 15.0$, $c_2 = 80.0$, $c_3 = 0.0$. The discount rate is set to $r = 0.03$ p.a. In this way we get a setup, where the disease does not die out quickly without any measures. Moreover, the cost c_2 of control measures is high enough that it is not possible to extinct the disease using the control. Finally we assume that the random initial value $X(0)$ follows a uniform distribution with density

$$p_0(x) \quad = \quad \begin{cases} 20 & \text{if } x \in [0.03,\, 0.08]\,, \\ 0 & \text{else.} \end{cases}$$

This expresses the view that the initial fraction of infected individuals lies between 0.03 and 0.08, but each fraction in this range is equally likely.

Figure 1 shows the transition density function $p(x,t)$ for the optimal solution of the Fokker-Planck constraint problem (57) when $\sigma = 0.5$ (low volatility case). The optimal value is (within the accuracy of three digits) equal to the optimal value of the deterministic solution, i.e.

$$\mathcal{U}(p_0(\cdot)) \approx \mathcal{R}(p_0(\cdot)) \approx 73.816.$$

This shows that the value of the open loop solution $VSOL$ is vanishingly small in this case, which can be interpreted as a sort of justification for using deterministic control in case of low volatility. Visually, the transition density in Figure 1 shows time dependent densities (cuts parallel to the x-axis) with small variation.

We consider now the case with larger volatility parameter $\sigma = 2.7$. Figure (2) shows a comparison between the deterministic optimal control and the stochastic optimal open loop control. While both strategies use more control at the beginning (which is due to the fact that we chose $c_3 = 0$ in this example), the stochastic control uses higher values over the whole planning period.

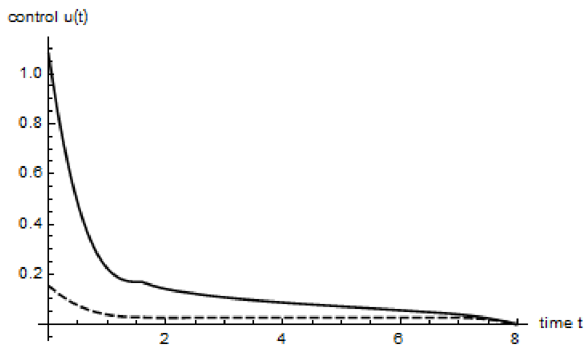

Figure 2: High volatility case $\sigma = 2.7$: Comparison between the optimal control for the deterministic problem (dashed line) and the optimal open loop control for the stochastic model (solid line).

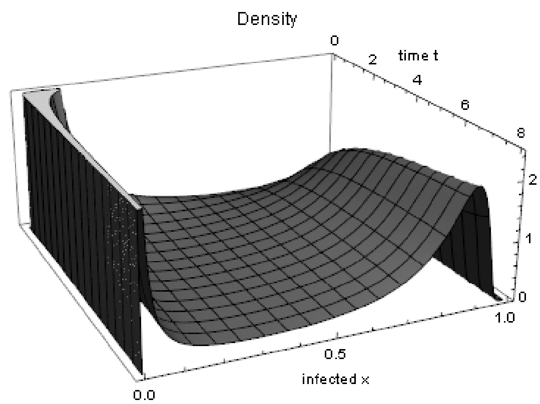

Figure 3: High volatility case $\sigma = 2.7$: Transition density $p(x,t)$ for the optimal solution of the Fokker-Planck constraint problem (57). The third axis (density $p(x,t)$) is cut at $p = 2.5$, so the small values contain substantially more probability mass then shown in the picture.

The related transition density when applying the optimal open loop control is shown in Figure 3. Here the control is able to bring down the number of infected near to zero and hold them there with high probability. Still there is a substantial probability that the disease grasps a high fraction of the population.

The stochastic open loop solution leads to a substantial reduction in expected costs: we have $\mathcal{U}(p_0(\cdot)) \approx 36.167$ for the optimal value of SOP and $\mathcal{R}(p_0(\cdot)) \approx 38.147$ for the expected cost of the deterministic solution, hence the value of the stochastic open loop problem is given by $VSOL \approx 1.98$. Note that already using the deterministic solution leads to a slight reduction of costs compared to the case without control ($u(x,t) = 0$), where the expected cost is equal to 39.669.

Finally we consider the full information case SFP, i.e. the solution of problem (11). The optimal transition density and the optimal control function are shown in Figure 4.

The transition density shows that the probabilities of staying near zero can be substantially improved, when being able to observe the states. The likelihoods of high fractions of infected then can be reduced a lot, also compared with the optimal open loop solution. The optimal control

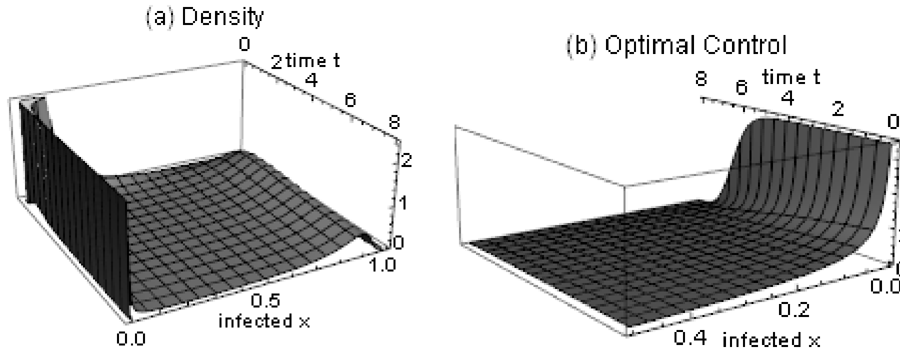

Figure 4: High volatility case: (a) transition density and (b) optimal feedback control $u(x,t)$ for the stochastic control problem (11). The third axis (density $p(x,t)$) is cut at $p = 2.5$, so the small values contain substantially more probability mass then shown in the picture. Note also that the x-axis is cut at $x = 0.5$ as the surface becomes very flat for larger values of x.

function again shows some decrease over time. More important are the differences in treatment intensities when compared over possible states. It shows that when it is possible to observe the states, one would put high effort per case if the fraction of infected is small, whereas the effort per case has to be reduced a lot if the number of cases becomes larger. Being able to observe the values of the process $X(t)$ is very valuable: the optimal value of the control problem is $\mathcal{W}(p_0(\cdot)) = 24.507$ and the value of observation therefore is $VO \approx 11.66$.

For both stochastic optimization model SOP and SFP and also for the deterministic approach, with $\sigma = 2.7$ the respective transition densities shows a similar twin-peaked visual pattern. There is a region near zero and a region with higher fraction of infected where the system tends to be concentrated. Because the pictures would look quite similar to Figure 3 at first glance, we do not show the transition densities for all the models. Instead we summarize the differences between the cases in Table 1, which shows the values of the related cumulative distribution functions of $X(T)$ at the end time $T = 8$ for selected fractions of infected. Not only SFP dominates SOP, SOP dominates the deterministic and the deterministic solution dominates the model without control in terms of the expected costs. In addition the same order of dominance can be observed at all considered fractions of infected: the probability that the fraction of infected is smaller or equal to the value given in the first column is larger for SFP than for SOP, larger for SOP than for the deterministic solution and larger for the deterministic solution than without any control.

Conclusions

We analyzed both the cases introduced in the introduction, the full information case (SFP) and the no information case (SOP). In particular we gave (under some technical assumptions)

- a sufficient condition (the HJB equation) together with a verification theorem for the optimal solution of the SFP.

infected x	no control	deterministic	no observation (SOP)	full observation (SFP)
	39.669	38.147	36.167	24.507
0.001	0.24535	0.25891	0.31358	0.85869
0.010	0.31168	0.32767	0.38703	0.86568
0.100	0.38831	0.40631	0.46629	0.87559
0.200	0.41923	0.43767	0.49630	0.88084
0.500	0.49929	0.51762	0.57010	0.89649

Table 1: Ecpected costs and cumulative distribution function for the analyzed approaches (no control with $u(x,t) = 0$, deterministic DP, no observation SOP and full observation SFP).
The second line shows the expected costs of the respective (optimal) strategies. The rest of the table shows the values of the cumulative distribution function $P_m(X(T) \leq x)$ at time $T = 8$ for the respective models at the fraction x of infected individuals specified in the first column.

- a sufficient condition with verification theorem for the SOP under quite strong assumptions about the control function, using stochastic calculus as the basis.
- a necessary and sufficient condition for the SOP under weaker assumptions, applying the idea of optimization under PDE-constraints.

It turns out that the two approaches for the SOP lead to identical optimality conditions (derived in different spaces), as can be expected. In a numerical part we compared the solutions (the control functions) of SFP and SOP with each other and with the solution of the deterministic problem. Plugging these solutions into the same stochastic dynamics led to the insight that using stochastic optimization has an advantage over using just (instantaneous) expectation for defining the dynamics. Moreover it also leads to decreasing expected costs if one is able to observe the number of infected individuals over time - so information has a numberable value for the decision maker.

Clearly several open questions have to be left for further research: The SIS model is one of the simplest epidemiological models, so it is natural to apply the basic ideas of stochastic SFP and SOP to more elaborated models (e.g. more compartments or heterogeneity age). In the optimization model so far we restricted to controlling the recovery rate (which only influences the drift), while it might also be important to influence the infection process (e.g. vaccination). This in turn may lead to a control that also acts on the diffusion part of the dynamics. Moreover, one may also consider different forms of the cost function.

References

[1] P. R. M. Abad, F. X. and A. Bosch. Survival of enteric viruses on environmental fomites. *Applied and Environmental Microbiology*, 60(3704-3710), 1994. 3

[2] E. Albina. Epidemiology of porcine reproductive and respiratory syndrome (PRRS): An overview. *Veterinary Microbiology*, 55:309–316, 1997. 3

[3] F. Brauer, L. J. S. Allen, P. Van den Driessche, and J. Wu. *Mathematical Epidemiology*. Lecture Notes in Math. 1945, Math. Biosci. Subser. Springer, 2008. 1

[4] C. Chen and Y. Kang. Dynamics of stochastic SIS epidemic model with saturated incidence. *Abstract and Applied Analysis*, 2014, 2014. 2

[5] N. Dexter. Stochastic models of foot and mouth disease in feral pigs in the australian semi-arid rangelands. *Journal of Applied Ecology*, 40:293–306, 2003. 3

[6] O. Diekmann, H. Heesterbeck, and T. Britton. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton Series in Theoretical and Computational Biology. Princeton University Press, Princeton, 2013. 2

[7] J. Dureau, K. Kalogeropoulos, and M. Baquelin. Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics*, 14(3):541–555, 2013. 2

[8] G. Fichera. On a unified theory of boundary value problems for elliptic-parabolic equations of second order. *University of Wisconsin Press, Madison/Wisconsin*, pages 97–120, 1960. 17

[9] W. H. Fleming and R. W. Rishel. *Deterministic and Stochastic Optimal Control*. Springer-Verlag, New York, Heidelberg, Berlin, 1975. 18

[10] C. Gardiner. *Stochastic Methods - a Handbook for the Natural and Social Sciences*. Springer, Berlin, Heidelberg, New York, 2009. 6, 19, 21

[11] A. Gray, D. Greenhalgh, L. Hu, X. Mao, and J. Pan. A stochastic differential equation SIS epidemic model. *SIAM Journal of Applied Mathematics*, 71(3):870–902, 2011. 2

[12] H. W. Hethcote and J. A. Yorke. *Gonorrhea Transmission Dynamics and Control*. Lecture Notes in Biomath. 56. Springer, 1984. 1

[13] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, Heidelberg, New York, 2009. 17

[14] M. J. Keeling and P. Rohani. *Modeling Infectuous Disease in Humans and Animals*. Princeton University Press, Princeton and Oxford, 2008. 3

[15] M. J. Keeling and J. V. Ross. On methods for studying stochastic disease dynamics. *Journal of the Royal Society Interface*, 5:171–181, 2008. 2

[16] C. T. Kelley. *Iterative Methods for Optimization*. SIAM, Philadelphia, 1999. 28

[17] D. E. Kirk. *Optimal Control Theory - An Introduction*. Dover, Mineola, New York, 2004. 28

[18] R. M. Kovacevic. Stochastic SIS epidemic models - the PDE-approach. *Research Report 2015-09, ORCOS, TU Wien*, 2015. 2

[19] V. A. Ladyženskaja O. A., Solonnikov and N. N. Uralceva. *Linear and quasilinear equations of parabolic type*. American Mathematical Society, Providence, Rhode Island, 1968. 9, 10, 11, 12, 13

[20] O. A. Oleinik. On linear second order equations with non-negative characteristic form. *Actes, Congres intern Math.*, 2:771–777, 1970. 17

[21] O. A. Oleinik, Oleĭnik and E. V. Radkevich. *Second order equations with nonnegative characteristic form*. Plenum Press, New York, London, 1973. 21

[22] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory*. Harvard University Press, 1961. 27

[23] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*. Springer, Berlin, 3rd edition, 1998. 15

[24] W. E. Schiesser. *The Numerical Method of Lines*. Academic Press inc., San Diego, New York, Boston, 1991. 28

[25] W. A. Spurr and C. Bonini. *Statistical Analysis of Business Decisions*. Richard D. Irwin Publishing House, 1973. 27

[26] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*. Graduate Studies in Mathematics, Volume 112. American Mathematical Society, Providence, Rhode Island, 2010. 20, 21, 28