OPTIMAL ADAPTIVE CONTROL WITH SEPARABLE DRIFT UNCERTAINTY

SAMUEL N. COHEN*, CHRISTOPH KNOCHENHAUER†, AND ALEXANDER MERKEL‡

Abstract. We consider a problem of stochastic optimal control with separable drift uncertainty in strong formulation on a finite time horizon. The drift of the state Y^u is multiplicatively influenced by an unknown random variable λ, while admissible controls u are required to be adapted to the observation filtration. Choosing a control actively influences the state and information acquisition simultaneously and comes with a learning effect. The problem, initially non-Markovian, is embedded into a higher-dimensional Markovian, full information control problem with control-dependent filtration and noise. To that problem, we apply the stochastic Perron method to characterize the value function as the unique viscosity solution of the HJB equation, explicitly construct ε-optimal controls, and show that the values in the strong and weak formulation agree. Numerical illustrations show a significant difference between the adaptive control and the certainty equivalence control, highlighting a substantial learning effect.

Key words. adaptive control, drift uncertainty, exploration-exploitation trade-off

AMS subject classifications. 93E20, 93E11, 93C40, 49L25

1. Introduction.

Active learning in stochastic control is a topic of considerable interest, particularly in situations where unobservable components can affect the state evolution. In Feldbaum's seminal work [21], the concept of the *dual effect* was introduced (see [3] for a treatment in the field of stochastic control). The dual effect is the interplay between the control's effect on the state and its influence on the estimation of unobservable components through the controlled state. Consequently, the dual effect plays a key role in problems with a learning effect in stochastic control, as it describes a case of what is called, in "modern" language, the much-studied trade-off between exploration and exploitation. Here, exploration is in terms of knowledge / uncertainty about the unobservable component and exploitation refers to cost optimization.

In this paper, we investigate the problem of Bayesian adaptive optimal stochastic control in continuous time on a finite time horizon with separable drift uncertainty introduced via a hidden, static random variable. We take a Bayesian view of the estimation problem, that is we assume the prior is known and subsequently update our beliefs. In this context, ε-optimal controls are constructed using stability of viscosity solutions, and strong and weak formulations are shown to agree in value.

1.1. Problem description.

In the following, the controlled state Y^u satisfies

$$\mathrm{d}Y_s^u = \lambda^\intercal b(s, Y_s^u, u_s)\mathrm{d}s + \sigma(s, Y_s^u, u_s)\mathrm{d}W_s, \qquad Y_0^u = 0$$

where u is the control and λ is an unobservable random variable with prior distribution μ. For each control, the state generates a filtration $\mathcal{Y}^u = \sigma(Y^u)$ called the "observation filtration" which is explicitly control-dependent in the strong formulation. As a result of the nonlinear drift, the *separation principle* first formulated in [47] generally does not hold. The separation principle roughly says that, under certain

*University of Oxford, Mathematical Institute, Oxford, United Kingdom, OX2 6GG (cohens@maths.ox.ac.uk).

†Technische Universität München, School of Computation, Information and Technology, Parkring 11–13, 85748 Garching bei München, Germany (knochenhauer@tum.de).

‡Technische Universität Berlin, Institut für Mathematik, Straße des 17. Juni 136, 10623 Berlin, Germany (merkel@math.tu-berlin.de).

conditions (typically linearity of the state dynamics), control and estimation can be decoupled, and the problem of simultaneous control and estimation separates into two problems; this does not apply here. The goal is to minimize the cost functional

$$\mathcal{J}(u) = \mathbb{E}\Big[\int_0^T k(t, Y_t^u, u_t, \lambda)\mathrm{d}t + g(Y_T^u, \lambda)\Big]$$

over a class of controls u which are adapted to their own generated filtration $\mathcal{Y}^u$, that is, they only rely on the information on λ and W generated from observing Y^u.

In this formulation, the control is of closed-loop type and directly influences the controller's knowledge of the hidden parameter λ, resulting in a learning effect. More precisely, a trade-off between exploration and exploitation arises, as the control must balance improving the estimation and minimizing the cost functional.

Mathematically, the dependence of the observation filtration $\mathcal{Y}^u$ on the control u is a result of the strong formulation of the control problem. To make the problem approachable using techniques of stochastic optimal control, especially dynamic programming, we rewrite the state dynamics in the filtration $\mathcal{Y}^u$. It turns out that, by introducing two control-dependent auxiliary states, we can fully describe the conditional distribution of λ given $\mathcal{Y}^u$ in a Markovian way. As such, the original problem is embedded into a finite-dimensional Markovian problem, and we can apply the techniques of dynamic programming.

To gain further insight, we also study an alternative weak formulation of the problem. A priori, given the effects of controlling the information flow and the dependence of the filtration on the control in the strong formulation, it is not clear if the optimal cost in the weak and strong formulation agree. Nevertheless, we link the two formulations using the stochastic Perron method [6], which yields a characterization of the strong and weak value functions as the unique viscosity solution of the same HJB equation, that is, the two value functions agree. The stochastic Perron method allows for the derivation of the viscosity characterization of the value functions without explicitly proving the dynamic programming principle (DPP) in continuous time. Instead, one uses suitable notions of sub- and supersolutions carrying the necessary intertemporal structure required for the proof.

Choosing a family of auxiliary control problems with the control set restricted to piecewise constant controls as the class of supersolutions, we establish the DPP for these controls and then build an approximation scheme in the sense of [4] converging from above to a viscosity subsolution of the HJB equation which dominates the value function in the strong formulation. Regarding the approximation from below, we consider stochastic subsolutions (in the weak formulation) and show that their pointwise supremum is a viscosity supersolution dominated by the value function in the weak formulation. Consequently, a comparison principle for the HJB equation implies that the limit of the control problems with piecewise constant controls and *the value functions of the problems in strong and weak formulation agree*, which is to say that additional randomization does not decrease the value of the control problem. Moreover, by establishing existence of optimizers for the problems with piecewise constant controls, we are able to *construct ε-optimal controls* which can be efficiently computed.

Problems with unknown dynamics and cost have highly relevant applications in, for example, the problem of optimal execution in mathematical finance. In this context, Y^u represents the asset price under price impact and $b(t, y, u) = u$ is a simple model for unobservable permanent price impact λ (see the monographs [13, 24, 45] for an introduction to problems of this type). This is a challenging issue in problems

of optimal execution, as price impact factors are generally unobservable and have to be estimated from the affected price. Another field of application is in motion control of robots; we solve a stylized example numerically in section 6 to demonstrate and compare our results to a naïve control obtained by replacing λ by its expectation $\mathbb{E}[\lambda]$ and a certainty equivalent (CE) control (c.f. [23]) which uses the current conditional mean as the best estimate, but neglects the control of future available information. This can be considered as a very specific continuous time continuous space version of a POMDP ([34]), which have a wide range of applications.

The HJB equation is solved numerically using the deep Galerkin method [43] combined with policy iteration; convergence of this method in the linear case was established recently in [29].

1.2. Related literature. Previous studies involving unobservable components, for example in other problems of adaptive control or partially observable control (e.g. [9, 10, 22]), usually pose the problem in a weak formulation such that the filtration does not depend on the control. There, the probability space and filtration are fixed, weak solutions to the state equation are considered, and the control is introduced via a change of measure to the cost functional. Another approach is to require that for a sufficiently large class of controls, the corresponding observation filtration agrees with the uncontrolled observation filtration and then work with the latter one. In the stochastic optimal control literature, direct dependence of the observation filtration on the control is usually avoided or resolved via the separation principle, which has a variety of technical formulations: Wonham [47] defines this as the existence of an optimal feedback control, while [9, 3] define the principle as the ability to replace the hidden variable with its conditional expectation. For a more in depth discussion, we refer the interested reader to [23].

For the sake of clarity, we will interpret the separation principle as the statement that 'the optimal control problem can be posed in such a way that the relevant filtration is fixed independently of the control'. With this formulation, we are not aware of any work studying such problems in the strong formulation when the separation principle does not hold. In many works [22, 11, 20, 2], the control does not directly influence the observation process (possibly due to additional randomization), allowing weak formulations of their control problems via change of measure, under which the separation principle (in our formulation) holds. In these contexts a different 'separation principle' is often given, where an extended state variable (based on the conditional distribution) is constructed and used as a basis for analysis. We will see in this paper that this approach is also valid for a problem where the control directly alters the filtration.

In [9, 30] on an infinite horizon and in [31] on a finite horizon, a special case of our control problem is considered in a weak formulation for quadratic cost on the state variable in a class of what the authors refer to as "wide-sense" admissible controls. They show the intriguing result that there does not exist an optimal control in the class of "strict-sense" admissible controls in the weak formulation but construct optimal controls in the class of "wide-sense" admissible controls. This, together with the fact that the optimal wide-sense admissible control is Markovian, suggests also that, in the strong formulation, in general there does not exist an optimal admissible control, thus justifying our search for ε-optimal controls.

A survey on the stochastic adaptive control problem from a control theory perspective is given in [37] and a recent broader view on adaptive control is, e.g., the monograph [1]. Similarly, model predictive control deals with control of unknown

systems, usually in discrete time and only sometimes with noise (see [27] for an overview). There, the goal is generally not to find optimal controls but implementable "good controls" having "good" asymptotic stability and robustness properties and to prove suboptimality guarantees. They also treat non-Bayesian approaches. Often, for example, an ergodic criterion is prescribed, and we mention [17, 18] as examples in this direction.

The problem of adaptive control has also attracted attention over the last decade due to the interest in reinforcement learning and, more generally, machine learning and its applications in control theory. We mention two recent examples: [44] who examine the linear convex episodic reinforcement learning problem in continuous time and [40] who derive suboptimality bounds for a certainty equivalent controller in the linear-quadratic problem in discrete time. The reinforcement learning community mostly considers asymptotic regret optimality and other types of asymptotic optimality, whereas we consider optimality on a finite horizon. As a consequence, in our formulation control effort matters at every point in time, whereas for asymptotic optimality, control effort on every finite time interval is irrelevant.

Finally, regarding our construction of ε-optimal controls, we point out that a way to construct ε-optimal controls in the class of piecewise constant controls was already suggested in the seminal monograph [36]. In [35] the author approximates the value function using value functions over the class of piecewise constant controls and proves the DPP in that class. We use regularity results appearing in [36] to establish the DPP for the case of unbounded cost functions over the class of piecewise constant controls. As a consequence of measurable selection, we obtain optimal controls for the approximating problems, which are Markovian "on a time grid". Convergence, and thus ε-optimality, for the original problem is not shown via direct analysis of the cost functional as in [36] and [35], but using instead the theory of viscosity solutions and the main stability result of [4], allowing us to additionally characterize the value function as the unique continuous viscosity solution of the HJB equation. The connection to the HJB equation was not made in [35], but explicit convergence rates in terms of the step size were obtained, something that is not included in our approach.

The rest of our work is structured as follows. In section 2 we formulate the optimal control problem in strong formulation. In section 3 we rewrite the state's dynamics in its own filtration and introduce two additional auxiliary states. A dynamic Markovian optimal control problem in strong formulation is introduced into which the original problem is embedded. In section 4 we furthermore introduce the Markovian control problem in weak formulation. In section 5 we establish the main results of this paper by showing that the weak and strong value functions coincide with the unique continuous viscosity solution of the HJB equation and construct ε-optimal controls. Finally, in section 6 we present an application of our results to a toy problem of optimal control in robotics, highlighting a substantial learning effect.

Notation. Throughout, we fix a complete probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ and denote by $T > 0$ a finite time horizon. The symbols $\mathrm{D}_x, \mathrm{D}_{xy}$ denote the gradient and Hessian with respect to the (multivariate) components x, y, whereas ∂_z denotes the partial derivative with respect to the (scalar) component z. Finally, $\mathcal{S}_d$ denotes the set of symmetric $d \times d$ matrices for any $d \in \mathbb{N}$.

2. The control problem. We pose the control problem beginning with an $\mathbb{R}^m$-valued random variable $\lambda = (\lambda_1, \dots, \lambda_m)^\intercal$ with distribution μ under $\mathbb{P}$. We furthermore suppose that $(\Omega, \mathfrak{F}, \mathbb{P})$ supports a one-dimensional Brownian motion $W = (W_t)_{t\in[0,T]}$ independent of λ, and we denote by $\mathcal{F}^{\lambda,W}$ the filtration generated by λ

and W, augmented by all $\mathbb{P}$-nullsets.

Assumption 2.1. λ is bounded, that is $|\lambda| \leq K$ for some $K \geq 0$.

This assumption guarantees that the estimator function introduced in the next section is Lipschitz continuous. Next, we consider controls taking values in a compact metric space $\mathcal{U}$. This is a standard assumption which allows us to construct controls using measurable selection. With this, the set of *pre-admissible controls* is

$$\mathcal{A}^{pre} := \big\{u : \Omega \times [0,T] \to \mathcal{U} : u \text{ is } \mathcal{F}^{\lambda,W}\text{-progressively measurable}\big\}.$$

For any $u \in \mathcal{A}^{pre}$, the controller observes a controlled one-dimensional state process $Y^u = (Y^u_t)_{t\in[0,T]}$, defined as the unique strong solution of

$$\mathrm{d}Y^u_t = \lambda^\intercal b(t, Y^u_t, u_t)\mathrm{d}t + \sigma(t, Y^u_t, u_t)\mathrm{d}W_t, \qquad Y^u_0 = 0, \tag{2.1}$$

where $b : [0,T] \times \mathbb{R} \times \mathcal{U} \to \mathbb{R}^m$ and $\sigma : [0,T] \times \mathbb{R} \times \mathcal{U} \to (0,\infty)$, and the initial state $Y_0 = 0$ is chosen for simplicity.

Assumption 2.2. The functions b, σ are jointly continuous in all arguments. Furthermore, there exist constants $L, M > 0$ such that, for all $t \in [0,T], u \in \mathcal{U}$,

$$\begin{aligned} |b(t,y_1,u) - b(t,y_2,u)| + |\sigma(t,y_1,u) - \sigma(t,y_2,u)| &\leq L|y_1 - y_2|, \qquad \forall y_1, y_2 \in \mathbb{R}, \\ |b(t,y,u)| + |\sigma(t,y,u)| + |\sigma(t,y,u)^{-1}| &\leq M, \qquad \forall y \in \mathbb{R}. \end{aligned}$$

Lipschitz-continuity and boundedness ensure existence of a strong solution of (2.1) and boundedness of b, σ^{-1} ensure that a certain Girsanov transform used to derive an estimator for λ is valid; boundedness of σ is for convenience. Uniformity of these estimates is required for regularity of the approximating problems in subsection 5.2.

By a standard existence result such as [36, Theorem 2.5.7], for each pre-admissible control $u \in \mathcal{A}^{pre}$, there exists a pathwise unique $(\mathcal{F}^{\lambda,W}, \mathbb{P})$-strong solution of (2.1) which generates a filtration, called the *observation filtration*, which we highlight to be control-dependent in the strong formulation.

Definition 2.3. *For $u \in \mathcal{A}^{pre}$, the observation filtration $\mathcal{Y}^u = (\mathcal{Y}^u_t)_{t\in[0,T]}$ is defined as the completed filtration generated by Y^u, that is $\mathcal{Y}^u_t := \sigma(Y^u_s, s \in [0,t]) \vee \mathcal{N}$ for all $t \in [0,T]$, where $\mathcal{N}$ denotes the system of $\mathbb{P}$-nullsets.*

Remark 2.4.
1) By definition, for every $u \in \mathcal{A}^{pre}$, we have $\mathcal{Y}^u \subset \mathcal{F}^{\lambda,W}$.
2) In the state dynamics (2.1), the unobservable λ does not appear in the diffusion coefficient. If it did, the problem would be fundamentally different.

We now restrict the controls to those which are adapted to their corresponding filtration $\mathcal{Y}^u$, that is, they only rely on the information on λ and W obtained from observing the controlled state.

Definition 2.5. *The set of admissible controls is defined as*

$$\mathcal{A} := \big\{u \in \mathcal{A}^{pre} : u \text{ is } \mathcal{Y}^u\text{-progressively measurable}\big\}.$$

Admissible controls are therefore *closed-loop controls* in the sense that they can utilize their effect on the observations; see [3] for an elaborate discussion.

Remark 2.6. Care must be taken in defining the set of admissible controls, as can be seen from the example of the Tsirel'son SDE (see, e.g., [41, p.362]). There, a bounded, nonanticipating but path-dependent drift (potentially a feedback map) is

constructed, which introduces additional independent randomness in the generated filtration, such that "$\mathcal{Y}^u \not\subseteq \mathcal{F}^{\lambda,W}$". Our choice of admissible controls, specifically $\mathcal{A} \subset \mathcal{A}^{pre}$, excludes these controls from being admissible.

In general, a definition of the set of admissible controls adapted to the filtration of a solution of a controlled SDE is circular. Indeed, in order for the observation filtration to exist, for each control there needs to exist a solution of the state equation (2.1), but to ensure such an existence, one needs to specify the set of admissible controls. To the best of our knowledge, there are three approaches:

- Fix an observation filtration $\mathcal{F}$ and work with weak solutions to the state equation and controls u which are $\mathcal{F}$-progressive (as in, e.g., [9, 22]).
- Choose as controls nonanticipative feedback maps $\hat{u} : [0,T] \times \mathcal{C}([0,T]) \to \mathcal{U}$, regular enough to define the state process. This is an approach often followed by the reinforcement learning community.
- Use a "reference" filtration, in our case $\mathcal{F}^{\lambda,W}$, to define a superset of controls, in our case $\mathcal{A}^{pre}$, and guarantee existence of the controlled state. Then restrict to those controls that are adapted to a control-dependent subfiltration, in our case $\mathcal{Y}^u$.

Note that the set of controls $\mathcal{A}$ is not a nice set to work with. For example, it is not closed under addition, even when the sum takes values in $\mathcal{U}$; i.e. for $u_1, u_2 \in \mathcal{A}$ it is not clear whether $u_1 + u_2$ is $\mathcal{Y}^{u_1+u_2}$-progressive and thus admissible.

For a control $u \in \mathcal{A}$, we define the *cost functional* as

$$\mathcal{J}(u) := \mathbb{E}\Big[\int_0^T k(t, Y_t^u, u_t, \lambda)\mathrm{d}t + g(Y_T^u, \lambda)\Big] \qquad \text{subject to (2.1)}, \tag{2.2}$$

where $k : [0,T] \times \mathbb{R} \times \mathcal{U} \times \mathbb{R} \to [0,\infty)$ and $g : \mathbb{R} \times \mathbb{R} \to [0,\infty)$. The goal is to minimize the cost functional $\mathcal{J}(u)$ over all $u \in \mathcal{A}$.

Assumption 2.7. k and g are jointly continuous. In addition, $k(t,y,u,\ell)$ is continuous in y uniformly over u for each $(t,\ell) \in [0,T] \times \mathbb{R}$ and $g(y,\ell)$ is uniformly continuous in y for each ℓ. Furthermore, we assume that there exist $C, p > 0$ with

$$|k(t,y,u,\ell)| + |g(y,\ell)| \le C(1 + |y|^p) \qquad \forall (t,y,u,\ell) \in [0,T] \times \mathbb{R} \times \mathcal{U} \times \mathbb{R}.$$

Definition 2.8. *We say that a control $u^* \in \mathcal{A}$ is optimal if*

$$\inf_{u \in \mathcal{A}} \mathcal{J}(u) = \mathcal{J}(u^*) \tag{2.3}$$

with $\mathcal{J}$ as in (2.2). Moreover, for every $\varepsilon > 0$, a control $u^{*,\varepsilon} \in \mathcal{A}$ is ε-optimal provided that $\mathcal{J}(u^{*,\varepsilon}) \le \inf_{u \in \mathcal{A}} \mathcal{J}(u) + \varepsilon$.

Remark 2.9. In the problem formulation considered here, Y^u is generally not a $\mathcal{Y}^u$-Markov process. In the next section, we derive a finite-dimensional control problem under full information where for each control $u \in \mathcal{A}$ the coefficients of Y^u are adapted to $\mathcal{Y}^u$, and which can be solved via dynamic programming in the sense of ε-optimal controls. We thus also obtain ε-optimal controls for the original problem (2.3).

Remark 2.10.

1) Our results extend to the case of a multidimensional state Y^u under suitably adjusted assumptions. Since this only adds to the notation, we stick to the one-dimensional case.
2) A motivating example for a multidimensional hidden parameter λ can be constructed by considering $\lambda = (1, \bar{\lambda}, \bar{\lambda}^2, \dots, \bar{\lambda}^N)$ for some real-valued $\bar{\lambda}$ and

fixed $N \in \mathbb{N}$. This, together with suitable coefficients $b_i(\,\cdot\,)$, $i = 0, \dots, N$, could be used for a polynomial approximation of a general drift function $b(\,\cdot\,, \bar{\lambda}) \approx \lambda^\intercal b(\,\cdot\,)$. The validity of such an approximation is left open for future research, but may provide a way to study the case of non-separable drift uncertainty.

3. Transformation to a full information problem. We now embed the problem into one in which the coefficients and cost are adapted to the observation filtration. Making use of techniques of Bayesian inference, we find a finite-dimensional parametrization of the conditional distribution of λ by two $\mathcal{Y}^u$-adapted information states. This suffices to transform the problem into a control problem under full information.

To begin with, we introduce a new probability measure $\mathbb{Q}^u$ as follows. Let $u \in \mathcal{A}$ and define two density processes $\Lambda^u = (\Lambda^u_t)_{t\in[0,T]}$ and $Z^u := (Z^u_t)_{t\in[0,T]}$ via

$$\begin{aligned}\Lambda^u_t := \frac{1}{Z^u_t} &:= \mathcal{E}\Big(-\int_0^\cdot \frac{\lambda^\intercal b}{\sigma}(s, Y^u_s, u_s)\mathrm{d}W_s\Big)_t \\ &= \exp\Big(-\int_0^t \frac{\lambda^\intercal b}{\sigma}(s, Y^u_s, u_s)\mathrm{d}W_s - \frac{1}{2}\int_0^t \frac{\lambda^\intercal bb^\intercal\lambda}{\sigma^2}(s, Y^u_s, u_s)\mathrm{d}s\Big) \\ &= \exp\Big(-\int_0^t \frac{\lambda^\intercal b}{\sigma^2}(s, Y^u_s, u_s)\mathrm{d}Y^u_s + \frac{1}{2}\int_0^t \frac{\lambda^\intercal bb^\intercal\lambda}{\sigma^2}(s, Y^u_s, u_s)\mathrm{d}s\Big).\end{aligned}$$

Since λ is bounded by Assumption 2.1 and b, σ^{-1} are bounded by Assumption 2.2, Λ^u is an $(\mathcal{F}^{\lambda,W}, \mathbb{P})$-martingale and defines a $\mathbb{P}$-equivalent probability measure $\mathbb{Q}^u$ on $(\Omega, \mathcal{F}^{\lambda,W}_T)$ with

$$\mathbb{Q}^u(A) := \mathbb{E}[\Lambda^u_T \mathbb{1}_A] \qquad \forall A \in \mathcal{F}^{\lambda,W}_T.$$

By Girsanov's theorem [32, Theorem 3.5.1], we find that Y^u is a standard $(\mathcal{F}^{\lambda,W}, \mathbb{Q}^u)$-Brownian motion independent of λ and, as $\mathcal{Y}^u \subset \mathbb{F}^{\lambda,W}$, it is also a $(\mathcal{Y}^u, \mathbb{Q}^u)$-Brownian motion. Furthermore, λ retains distribution μ under $\mathbb{Q}^u$; see [33, Lemma 2.2]. With this, we define another density process $\hat{Z}^u = (\hat{Z}^u_t)_{t\in[0,T]}$, which is also a $(\mathcal{Y}^u, \mathbb{Q}^u)$-martingale, by

$$\begin{aligned}\hat{Z}^u_t &:= \int_{\mathbb{R}} \exp\Big(\int_0^t \frac{\ell^\intercal b}{\sigma^2}(s, Y^u_s, u_s)\mathrm{d}Y^u_s - \frac{1}{2}\int_0^t \frac{\ell^\intercal bb^\intercal\ell}{\sigma^2}(s, Y^u_s, u_s)\mathrm{d}s\Big)\mu(\mathrm{d}\ell) \\ &= \mathbb{E}^{\mathbb{Q}^u}\Big[\exp\Big(\int_0^t \frac{\lambda^\intercal b}{\sigma^2}(s, Y^u_s, u_s)\mathrm{d}Y^u_s - \frac{1}{2}\int_0^t \frac{\lambda^\intercal bb^\intercal\lambda}{\sigma^2}(s, Y^u_s, u_s)\mathrm{d}s\Big)\Big|\mathcal{Y}^u_t\Big] \\ &= \mathbb{E}^{\mathbb{Q}^u}[Z^u_T|\mathcal{Y}^u_t] = \mathbb{E}^{\mathbb{Q}^u}[Z^u_t|\mathcal{Y}^u_t],\end{aligned} \tag{3.1}$$

where we used that $\lambda \sim \mu$ and is independent of Y^u under $\mathbb{Q}^u$.

In order to avoid having to deal with matrix-valued SDEs,[1] we introduce the half-vectorization operator vech : $\mathcal{S}_m \to \mathbb{R}^{m(m+1)/2}$ (see, for example, [25]) such that

$$\mathcal{S}_m \ni A = (a_{i,j})_{1\le i,j\le m} \mapsto \mathrm{vech}A = (a_{1,1}, a_{2,1}, a_{2,2}, a_{3,1}, \dots, a_{m,m})^\intercal \in \mathbb{R}^{m(m+1)/2}.$$

[1] One could work with the matrix-valued version equally well, but vectorization allows us to 'stack' the components of our state variable into a single vector state, which is arguably more familiar from the PDE perspective; ultimately this is a matter of taste. We observe that the half-vectorization operator vech is linear and invertible, and so vech and vech^{-1} can be manipulated as fixed linear maps from the perspective of stochastic calculus and PDEs.

With this, we define two additional information state processes $\Upsilon^u = (\Upsilon^u_t)_{t\in[0,T]}$ and $\Gamma^u = (\Gamma^u_t)_{t\in[0,T]}$ by

$$\Upsilon^u_t := \int_0^t \frac{b}{\sigma^2}(s, Y^u_s, u_s)\mathrm{d}Y^u_s \quad \text{and} \quad \Gamma^u_t := \int_0^t \mathrm{vech}\Big(\frac{bb^\intercal}{\sigma^2}(s, Y^u_s, u_s)\Big)\mathrm{d}s, \tag{3.2}$$

taking values in $\mathbb{R}^m$ and $\mathbb{R}^{m(m+1)/2}$, respectively. Note that the two processes are related via $\langle \Upsilon^u\rangle = \mathrm{vech}^{-1}(\Gamma^u)$.

Using this notation, for $u \in \mathcal{A}$, we define the *unnormalized conditional distribution* of λ under $\mathbb{Q}^u$ given $\mathcal{Y}^u$, denoted by $\rho^u = (\rho^u_t)_{t\in[0,T]}$, as

$$\rho^u_t(A) := \mathbb{E}^{\mathbb{Q}^u}[Z^u_t \mathbb{1}_A(\lambda)|\mathcal{Y}^u_t] = \int_A \exp\Big(\ell^\intercal \Upsilon^u_t - \frac{1}{2}\ell^\intercal(\mathrm{vech}^{-1}\Gamma^u_t)\ell\Big)\mu(\mathrm{d}\ell) \quad \forall A \in \mathcal{B}(\mathbb{R}),$$

where $\mathcal{B}(\mathbb{R})$ is the Borel σ-field over $\mathbb{R}$. Using Bayes' rule [32, Lemma 3.5.3], we can identify the *normalized conditional distribution* of λ, denoted by $\pi^u = (\pi^u_t)_{t\in[0,T]}$, as

$$\pi^u_t(A) := \frac{\rho^u_t(A)}{\rho^u_t(\mathbb{R})} = \frac{\mathbb{E}^{\mathbb{Q}^u}[Z^u_t \mathbb{1}_A(\lambda)|\mathcal{Y}^u_t]}{\mathbb{E}^{\mathbb{Q}^u}[Z^u_t|\mathcal{Y}^u_t]} = \mathbb{P}(\lambda \in A|\mathcal{Y}^u_t).$$

In order to manipulate this (unnormalized) conditional distribution more simply, we define a function $F : \mathbb{R}^m \times \mathbb{R}^{m(m+1)/2} \to \mathbb{R}$ by

$$F(\upsilon, \gamma) := \int_{\mathbb{R}} \exp\Big(\ell^\intercal \upsilon - \frac{1}{2}\ell^\intercal(\mathrm{vech}^{-1}\gamma)\ell\Big)\mu(\mathrm{d}\ell) \tag{3.3}$$

and, for any sufficiently integrable function $\phi : \mathbb{R} \to \mathbb{R}$, the transformation $F[\phi] : \mathbb{R}^m \times \mathbb{R}^{m(m+1)/2} \to \mathbb{R}$ by

$$F[\phi](\upsilon, \gamma) := \int_{\mathbb{R}} \phi(\ell)\exp\Big(\ell^\intercal \upsilon - \frac{1}{2}\ell^\intercal(\mathrm{vech}^{-1}\gamma)\ell\Big)\mu(\mathrm{d}\ell). \tag{3.4}$$

We note that, by (3.1), we can express $\hat{Z}^u$ in terms of F evaluated along Υ^u, Γ^u, that is $\hat{Z}^u_t = F(\Upsilon^u_t, \Gamma^u_t)$, and similarly that $\rho^u_t(A) = F[\mathbb{1}_A](\Upsilon^u_t, \Gamma^u_t)$.

Remark 3.1. In [19], F as defined in (3.3) is referred to as the *Widder transform* of μ (due to [46], see also [32, Section 4.3 B]). This should not be confused with the Post–Widder transform common in the theory of Laplace transforms.

Noticing that the gradient of F with respect to υ is given by

$$F_\upsilon(\upsilon, \gamma) = \int_{\mathbb{R}} \ell \exp\Big(\ell^\intercal \upsilon - \frac{1}{2}\ell^\intercal(\mathrm{vech}^{-1}\gamma)\ell\Big)\mu(\mathrm{d}\ell) = F[\mathrm{id}](\upsilon, \gamma),$$

we find that the conditional mean of λ can be expressed in terms of the process $m^u = (m^u_t)_{t\in[0,T]}$ given by

$$m^u_t := G(\Upsilon^u_t, \Gamma^u_t) = \mathbb{E}[\lambda|\mathcal{Y}^u_t] = \int_{\mathbb{R}} \ell \pi^u_t(\mathrm{d}\ell), \tag{3.5}$$

where $G : \mathbb{R}^m \times \mathbb{R}^{m(m+1)/2} \to \mathbb{R}$ is defined as

$$G(\upsilon, \gamma) := \frac{F_\upsilon}{F}(\upsilon, \gamma). \tag{3.6}$$

By Assumption 2.1, we find that $|G|$ is bounded by the same constant $K > 0$ as λ.

Remark 3.2. In the above definitions, we have always constructed continuous versions of the conditional expectations, which are therefore measurable. The other identities then hold in a $\mathbb{P}$-a.s. sense.

LEMMA 3.3. *The function G is Lipschitz continuous.*

Proof. Since G is continuously differentiable, it suffices to show that its gradient is uniformly bounded. For this, we note that

$$|\mathrm{D}G|^2 = \Big|\frac{F_{vv}}{F} - G^2\Big|^2 + \Big|\frac{F_{v\gamma}}{F} - \frac{F_\gamma}{F}G\Big|^2 \le 2\Big\{\Big|\frac{F_{vv}}{F}\Big|^2 + \Big|\frac{F_v^2}{F^2}\Big|^2 + \Big|\frac{F_{v\gamma}}{F}\Big|^2 + \Big|\frac{F_\gamma F_v}{F^2}\Big|^2\Big\}.$$

As μ is compactly supported by Assumption 2.1, for the mixed derivative we obtain

$$|F_{v\gamma}(v,\gamma)| \le \int_{\mathbb{R}} K^3 \exp\Big(\ell^\intercal v - \frac{1}{2}\ell^\intercal\big(\mathrm{vech}^{-1}\gamma\big)\ell\Big)\mu(\mathrm{d}\ell) \le K^3 F(v,\gamma),$$

and similar estimates hold for the other terms. It follows that G is Lipschitz. ∎

Next, for all controls $u \in \mathcal{A}$, we now define the corresponding *innovations process* $V^u = (V_t^u)_{t\in[0,T]}$ by

$$\begin{aligned} V_t^u &:= \int_0^t \frac{1}{\sigma(s, Y_s^u, u_s)}\big(\mathrm{d}Y_s^u - (m_s^u)^\intercal b(s, Y_s^u, u_s)\mathrm{d}s\big) \\ &= W_t + \int_0^t (\lambda - m_s^u)^\intercal \frac{b}{\sigma}(s, Y_s^u, u_s)\mathrm{d}s. \end{aligned} \tag{3.7}$$

The following key lemma can be found in [39, Lemma 11.3] or [15, Lemma 22.1.7].

LEMMA 3.4. *For all controls $u \in \mathcal{A}$, the corresponding innovations process V^u is a standard $(\mathcal{Y}^u, \mathbb{P})$-Brownian motion.*

Using (3.5) and (3.7), we can rewrite Y^u with dynamics in the observation filtration $\mathcal{Y}^u$ as

$$Y_t^u = \int_0^t G(\Upsilon_s^u, \Gamma_s^u)^\intercal b(s, Y_s^u, u_s)\mathrm{d}s + \int_0^t \sigma(s, Y_s^u, u_s)\mathrm{d}V_s^u. \tag{3.8}$$

Similarly, the first auxiliary state Υ^u can be written as

$$\Upsilon_t^u = \int_0^t \Big[\frac{b}{\sigma^2}(s, Y_s^u, u_s)G(\Upsilon_s^u, \Gamma_s^u)^\intercal b(s, Y_s^u, u_s)\Big]\mathrm{d}s + \int_0^t \frac{b}{\sigma}(s, Y_s^u, u_s)\mathrm{d}V_s^u.$$

We now define the transformed running and terminal cost function $\tilde{k}$ and $\tilde{g}$ as

$$\tilde{k}(t, y, v, \gamma, u) := \frac{F[k(t, y, u, \cdot)](v, \gamma)}{F(v, \gamma)} \quad \text{and} \quad \tilde{g}(y, v, \gamma) := \frac{F[g(y, \cdot)](v, \gamma)}{F(v, \gamma)},$$

which are continuous as k and g are uniformly continuous in their last argument by combining Assumption 2.1 and Assumption 2.7. Indeed, with $z := (t, y, v, \gamma, u)$ and $z_n := (t_n, y_n, v_n, \gamma_n, u_n)$ such that $z_n \to z$, we immediately have

$$\begin{aligned} F[k(t_n, y_n, u_n, \cdot)](v_n, \gamma_n) &= \int_{\mathbb{R}} k(t_n, y_n, u_n, \ell) \exp\Big(\ell^\intercal v_n - \frac{1}{2}\ell^\intercal(\mathrm{vech}^{-1}\gamma_n)\ell\Big)\mu(\mathrm{d}\ell) \\ &\to F[k(s, y, u, \cdot)](v, \gamma) \end{aligned}$$

by dominated convergence. This is justified as μ is compactly supported, and the integrand is jointly continuous and as a result bounded along $(z_n)_{n\in\mathbb{N}}$. Thus, $\tilde{k}$ is continuous as a composition of continuous functions. Using Fubini's theorem and conditioning on $\mathcal{Y}^u$ in the cost functional, we obtain

$$\begin{aligned}\mathcal{J}(u) &= \int_0^T \mathbb{E}\Big[\mathbb{E}\big[k(t, Y_t^u, u_t, \lambda)\big|\mathcal{Y}_t^u\big]\Big]\mathrm{d}t + \mathbb{E}\Big[\big[g(Y_T^u, \lambda)\big|\mathcal{Y}_T^u\big]\Big] \\ &= \mathbb{E}\Big[\int_0^T \tilde{k}(t, Y_t^u, \Upsilon_t^u, \Gamma_t^u, u_t)\mathrm{d}t + \tilde{g}(Y_T^u, \Upsilon_T^u, \Gamma_T^u)\Big].\end{aligned}$$

This is now a control problem under full information, but with control-dependent noise and filtration, which we subsequently formulate dynamically. For ease of notation, we let $\tilde{m} := 1 + m + m(m+1)/2$ and define the drift and diffusion coefficient functions $f, \Sigma : \mathbb{S} \times \mathcal{U} \to \mathbb{R}^{\tilde{m}}$ with $x = (a, \upsilon, \gamma)$ as

$$f(t,x,u) := \begin{pmatrix} G(\upsilon,\gamma)^\intercal b(t,a,u) \\ \frac{b}{\sigma^2}(t,a,u) G(\upsilon,\gamma)^\intercal b(t,a,u) \\ \mathrm{vech}\big(\frac{bb^\intercal}{\sigma^2}(t,a,u)\big) \end{pmatrix} \quad \text{and} \quad \Sigma(t,x,u) := \begin{pmatrix} \sigma(t,a,u) \\ \frac{b}{\sigma}(t,a,u) \\ 0 \end{pmatrix}.$$

For each fixed control $u \in \mathcal{A}$, the coefficient functions $f(\,\cdot\,, u), \Sigma(\,\cdot\,, u) : \mathbb{S} \to \mathbb{R}^{\tilde{m}}$ are products of bounded, Lipschitz continuous functions, so they are also Lipschitz continuous. Hence, by a standard existence result (such as [36, Theorem 2.5.7]), for every initial condition in the *extended state space* $(t, x) \in \mathbb{S} := [0, T] \times \mathbb{R}^{\tilde{m}}$ and control $u \in \mathcal{A}$ there exists a pathwise unique $(\mathcal{Y}^u, \mathbb{P})$-strong solution $X^{u;t,x} := (A^{u;t,x}, \Upsilon^{u;t,x}, \Gamma^{u;t,x})$ to the $\tilde{m}$-dimensional *extended state* equation

$$\mathrm{d}X_s^{u;t,x} = f(s, X_s^{u;t,x}, u_s)\mathrm{d}s + \Sigma(s, X_s^{u;t,x}, u_s)\mathrm{d}V_s^u \tag{3.9}$$

for $s \in [t, T]$ with initial condition $X_t^{u;t,x} = x$. For the sake of clarity, we highlight that (3.9) can be seen simply as abbreviated notation for (3.8) and (3.2). Note that the diffusion coefficient Σ is degenerate, as the one-dimensional innovations process is the only driving noise and the third state variable is not even diffusive.

With this, we define the *extended cost functional* as

$$\mathcal{J}(u;t,x) := \mathbb{E}\Big[\int_t^T \tilde{k}(s, X_s^{u;t,x}, u_s)\mathrm{d}s + \tilde{g}(X_T^{u;t,x})\Big] \qquad \text{subject to (3.9)} \tag{3.10}$$

and the value function of the control problem as

$$V(t,x) := \inf_{u\in\mathcal{A}} \mathcal{J}(u;t,x).$$

The notion of (ε-)optimality given in Definition 2.8 applies for each $(t, x) \in \mathbb{S}$ in the obvious way. Further, the dynamic formulation of the value function is a (Markovian) ansatz, based on the idea that the (ε-)optimally controlled state should indeed be Markovian, which will be shown to be valid later on. We will see that the auxiliary processes Υ and Γ carry all information for the Bayesian estimation to lead to a Markovian state and hence can follow the idea of verification, where we propose a candidate (here V) for the value function.

Remark 3.5. 1) The extended control problem includes the original optimization problem (2.3). Specifically, for the solution $X^{u;0,0}$ of (3.9) we have $Y^u = A^{u;0,0}$ (as $Y_0^u = 0$) for all $u \in \mathcal{A}$, hence $\mathcal{J}(u) = \mathcal{J}(u; 0, 0)$. This means

that a $(0,0)$-optimal control for the extended cost functional (3.10) is also optimal for the original cost functional (2.3), as the minimization is over $\mathcal{A}$ in both cases.

2) In general V^u does not generate $\mathcal{Y}^u$ (this is the *innovations problem*, see e.g. [26]) but is only adapted to it. However, we will see that, for the ε-optimal controls u^ε constructed below, the corresponding innovations process generates the observation filtration, since the solution of the state equation is strong and Σ admits a left-inverse.
3) As the random variable λ only appears explicitly in the innovations process V^u, one might be inclined to expect that the full information problem depends on λ only via its (conditional) distribution. A priori, it is however not clear if this is indeed the case as swapping the innovations process with any other Brownian motion might affect the value function as the innovations process (and the filtration) is control-dependent. We argue below that V coincides with the value function in a weak formulation and construct ε-optimal controls by means of switching to another Brownian motion. Hence, a posteriori, V depends on λ only via its distribution.

Remark 3.6. Obtaining a finite dimensional description of the conditional distribution for a *time-dependent* hidden process $\lambda = (\lambda_t)_{t\in[0,T]}$ is known only in two cases, first in the conditionally Gaussian case (see [39, Chapter 12]) and second for a finite-state Markov chain (see [38, Chapter 9]) for which sufficient conditions for optimality are given in [12] in the form of a verification theorem. Since in our setting λ is static, we can work with a general distribution μ and still obtain an $(m + m(m+1)/2)$-dimensional description of the conditional distribution due to the separable structure of the drift.

In the following, for any $x \in \mathbb{R}^{\tilde{m}}$, we write $x = (a, \upsilon, \gamma)$ with $a \in \mathbb{R}$, $\upsilon \in \mathbb{R}^m$, and $\gamma \in \mathbb{R}^{m(m+1)/2}$. With this, the HJB equation for the extended control problem reads

$$\partial_t V + \inf_{u\in\mathcal{U}} \{\mathcal{L}^u V + \tilde{k}(\,\cdot\,, u)\} = 0, \qquad V(T,\,\cdot\,) = \tilde{g},$$

where for $u \in \mathcal{U}, b = b(\,\cdot\,, u), \sigma = \sigma(\,\cdot\,, u)$ the infinitesimal generator $\mathcal{L}^u$ is given by

$$\mathcal{L}^u = G^\intercal b\,\partial_a + \frac{b}{\sigma^2} G^\intercal b\,\mathrm{D}_\upsilon + \Big(\mathrm{vech}\frac{bb^\intercal}{\sigma^2}\Big)\mathrm{D}_\gamma + \frac{1}{2}\,\mathrm{tr}\,\Big[\Big(\sigma^2\mathrm{D}_{aa} + \frac{bb^\intercal}{\sigma^2}\mathrm{D}_{\upsilon\upsilon} + 2b\mathrm{D}_{a\upsilon}\Big)\Big]. \tag{3.11}$$

The HJB equation is fully nonlinear and degenerate as the second-order coefficient matrix is always of rank one, thus $\mathcal{L}^u$ is not uniformly elliptic.

Remark 3.7. To the best of our knowledge, no explicit solutions of the HJB equation have been obtained beyond the special cases of [9, 30] (infinite horizon) and [31] (finite horizon). Furthermore, there are no existence results which yield a solution sufficiently regular to apply classical verification. Such a classical verification theorem can nevertheless still be proved under the usual regularity assumptions. In [9, 30, 31], explicit classical solutions to the respective HJB equations were obtained. They also show that the optimally controlled state process does not admit a strong solution.

3.1. Connection of control and higher-order moments. In this subsection we briefly elaborate on the connection of the control and higher order conditional moments of λ. In particular, we draw connections to the conditional variance and the dual effect mentioned in the introduction. For simplicity, calculations are presented

for one-dimensional λ only. By definition of F, we see that with

$$G_k(\upsilon,\gamma) := \frac{\partial_\upsilon^k F}{F}(\upsilon,\gamma),$$

the conditional moments of k-th order are given by

$$m_t^{k;u} := G_k(\Upsilon_t^u, \Gamma_t^u) = \mathbb{E}[\lambda^k|\mathcal{Y}_t^u].$$

As a consequence the conditional variance in the initial problem is given by

$$\mathrm{var}_t^u := \mathbb{V}[\lambda|\mathcal{Y}_t^u] = G_2(\Upsilon_t^u,\Gamma_t^u) - G(\Upsilon_t^u,\Gamma_t^u)^2,$$

and straightforward computations detailed in Appendix A.1 show that

$$\mathrm{dvar}_t^u = -\Big[G^2(G_2+G^2)\Big](\Upsilon_t^u,\Gamma_t^u)\frac{b^2}{\sigma^2}(t,Y_t^u,u_t)\mathrm{d}t + G_{\upsilon\upsilon}(\Upsilon_t^u,\Gamma_t^u)\frac{b}{\sigma}(t,Y_t^u,u_t)\mathrm{d}V_t^u.$$

From this, we see that the controls exhibit the dual effect according to the definition given in [3]. There, the control is said to have *no dual effect of order k* (or *neutral* in the language of [21]), if all moments of higher order are independent of the control in the sense that they agree with the conditional moments given $\mathcal{Y}^0$ with $u \equiv 0$ (the "inactive control"). The control is said to *have a dual effect*, if it affects any higher order moment. Here, calculations similar to the above show that

$$\mathrm{d}m_t^{2;u} = G_{2,\upsilon}(\Upsilon_t^u,\Gamma_t^u)\frac{b}{\sigma}(t,Y_t^u,u_t)\mathrm{d}V_t^u,$$

i.e. the second moment is generally control-dependent and the dual effect is present.

4. Weak formulation. In this section we formulate the extended control problem in its weak formulation. We allow for the most general weak admissible controls in which the underlying filtered probability space and Brownian motion are part of the control, thereby introducing a control problem smaller in value than (3.10). The problem is then transformed into one under full information, exactly as in section 3. The purpose of this is to show that the values in strong and weak formulation agree and to allow comparison of these approaches from a modeling perspective.

DEFINITION 4.1. *For $(t,x) \in \mathbb{S}$, a* weak admissible control *is a seven-tuple $U^{t,x} = (\Omega^{t,x}, \mathfrak{F}^{t,x}, \mathcal{F}^{t,x}, \mathbb{P}^{t,x}, W^{t,x}, X^{t,x}, u^{t,x})$ such that*

1) *$(\Omega^{t,x}, \mathfrak{F}^{t,x}, \mathbb{P}^{t,x})$ is a probability space and the filtration $\mathcal{F}^{t,x}$ satisfies the usual conditions;*
2) *$W^{t,x}$ is a one-dimensional standard $(\mathcal{F}^{t,x}, \mathbb{P}^{t,x})$-Brownian motion;*
3) *$u^{t,x}$ is $\mathcal{F}^{t,x}$-progressively measurable and $\mathcal{U}$-valued;*
4) *$X^{t,x}$ is a continuous and $\mathcal{F}^{t,x}$-adapted process defined on $(\Omega^{t,x}, \mathfrak{F}^{t,x}, \mathbb{P}^{t,x})$ satisfying*

$$\mathrm{d}X_s^{t,x} = f(s,X_s^{t,x},u_s^{t,x})\mathrm{d}s + \Sigma(s,X_s^{u;t,x},u_s^{t,x})\mathrm{d}W_s^{t,x}, \quad s\in[t,T], \quad X_t^{t,x} = x.$$

That is, the tuple $(\Omega^{t,x}, \mathfrak{F}^{t,x}, \mathcal{F}^{t,x}, \mathbb{P}^{t,x}, W^{t,x}, X^{t,x})$ is a weak solution of the SDE (3.9). The components of $X^{t,x}$ are written as $X^{t,x} = (A^{t,x}, \Upsilon^{t,x}, \Gamma^{t,x})$. All objects above are understood with time index set $[t,T]$. The set of all weak admissible controls is denoted by $\mathcal{A}^{weak}(t,x)$.

With this, the value function over the set of weak controls is defined as

$$V^{weak}(t,x) := \inf_{U^{t,x}\in\mathcal{A}^{weak}(t,x)} \mathbb{E}^{t,x}\Big[\int_t^T \tilde{k}(s, X_s^{t,x}, u_s^{t,x})\mathrm{d}s + \tilde{g}(X_T^{t,x})\Big],$$

where $\mathbb{E}^{t,x}$ is the expectation under $\mathbb{P}^{t,x}$. From the fact that the state equation (3.9) admits a strong solution, it follows that the set of weak admissible controls is non-empty. As any strong solution of (3.9) is also a weak solution and the filtration is part of the control in this formulation, it is clear that we can embed $\mathcal{A} \hookrightarrow \mathcal{A}^{weak}(t,x)$, and hence

$$V \geq V^{weak} \qquad \text{on } \mathbb{S}. \tag{4.1}$$

This is the first key inequality which allows us to bound V from below by a "nicer" problem, where there is no dependence of the filtration and noise on the control.

Remark 4.2. The case for a weak formulation, like the one in this subsection, is the unobservability of the components of the state, in our case λ, W. Specifically, we cannot identify the Brownian motion W by observing only the state Y^u and, as a consequence, should be comfortable allowing the distribution and the driving Brownian motion to vary as part of the control, hence leading to the weak formulation. Furthermore, as long as the filtration generated by the state is only extended by independent "auxiliary" randomness, this does not violate the information pattern of basing decisions only on observations of the state.

The case against a weak formulation can also be made, as the noise process in the form of Brownian motion is generally control-independent and given "by nature", i.e. it is fixed. Furthermore, in order to make theoretical use of the above construction, one might have to work in a filtration strictly larger than the filtration generated by the state process Y^u, which in a sense violates a part of the idea behind the model, namely that the decision has to be made only on the basis of the information generated by the state Y^u. Another even more questionable point concerns the wide- and strict-sense admissible controls considered in [9, 30, 31]. There, the "observation filtrations" to which the controls are adapted are required to be independent of λ. But λ is part of the dynamics of the observable state, and thus should certainly *not* be independent of the observation filtration. In the strong formulation, $\mathcal{Y}^u$ is generally not independent of λ under $\mathbb{P}$.

5. Viscosity characterization and ε-optimal controls. Since obtaining a classical solution of the HJB equation is out of reach, as pointed out in Remark 3.7, we consider solutions in the viscosity sense (see [16]) instead. Recall that the HJB equation is given by

$$\partial_t V + \inf_{u\in\mathcal{U}}\{\mathcal{L}^u V + \tilde{k}(\,\cdot\,, u)\} = 0, \qquad V(T,\,\cdot\,) = \tilde{g} \tag{HJB}$$

on $\mathbb{S}$, where the infinitesimal generator $\mathcal{L}^u$ is defined in (3.11). We show that the value functions in the strong and weak formulation of the problem are equal to the unique viscosity solution of (HJB) using the stochastic Perron method. Moreover, we construct piecewise constant ε-optimal controls, which are also Markovian on a time-discretized grid. This allows us to link the strong and weak formulation in a clean way.

Our agenda for the remainder of this section is to apply a version of the stochastic Perron method [6]. More precisely, we

- prove a comparison principle for semicontinuous viscosity solutions of the HJB equation;
- using [4], show that the infimum of value functions of a family of auxiliary control problems with piecewise constant controls is a viscosity subsolution of the HJB equation;
- show that the supremum of stochastic subsolutions in the weak formulation with weak admissible controls is a viscosity supersolution.

We can then use the comparison principle and the auxiliary control problems to sandwich the value function V and show that it is itself a viscosity solution of (HJB).

5.1. The comparison principle. The comparison principle for the HJB equation (HJB) is a standard result. The main difficulty in the proof consists in controlling the viscosity sub- and supersolutions at infinity, which can be achieved by constructing a strict classical subsolution which grows sufficiently fast. The proof of the comparison principle is deferred to Appendix A.2.

THEOREM 5.1. *Let $U : \mathbb{S} \to \mathbb{R}$ be an upper semicontinuous viscosity subsolution and $W : \mathbb{S} \to \mathbb{R}$ be a lower semicontinuous viscosity supersolution of (HJB) for which there exist $C, q > 0$ such that*

$$0 \leq v(t,x) \leq C(1 + |x|^q) \qquad \forall v \in \{U, W\}, (t,x) \in \mathbb{S}.$$

If $U(T, \cdot) \leq W(T, \cdot)$ on $\mathbb{R}^{\tilde{m}}$, then $U \leq W$ everywhere on $\mathbb{S}$.

Proof. Let $\tilde{q} \geq 2$ such that $\tilde{q} > q$. A direct calculation shows that the function $\psi : \mathbb{S} \to (-\infty, 0)$ given by

$$\psi(t,x) := -|x|^{\tilde{q}} \exp\big(\zeta_1(T-t)\big) - \zeta_2(1 + T - t)$$

is a strict classical subsolution of the HJB equation provided that the constants $\zeta_1, \zeta_2 > 0$ are chosen sufficiently large. The comparison principle then follows using standard arguments as, e.g., in [7, Theorem 4.4]. ∎

5.2. The infimum of supersolutions. The most challenging step in our approach is the viscosity subsolution property of the value function V in the strong formulation. The main problem is that, in general, the set of admissible controls is not closed under pasting. That is, given two controls $u_1, u_2 \in \mathcal{A}$ and any $t \in [0,T]$, the pasted control $u := u_1 \mathbb{1}_{[0,t]} + u_2 \mathbb{1}_{(t,T]}$ can fail to be admissible. Since closedness under pasting is fundamental for the DPP to be valid, it is not immediately obvious if the value function V in the strong formulation can be linked to the HJB equation.

In a nutshell, our approach is based on the following two main ideas. First, by using the stochastic Perron method, we do not have to work with the value function directly, but can in fact resort to a sufficiently rich class of approximating functions from above as long as their pointwise infimum is a viscosity subsolution of the HJB equation. In what follows, this class of approximating functions is chosen to be the set of value functions with piecewise constant controls on a given time grid. The advantage of choosing these approximating functions is that it is relatively easy to show that they admit optimal controls in feedback form that are stable under pasting, which allows us to mitigate the problem of not being able to paste *arbitrary* controls.

Nevertheless, working with piecewise constant controls in our setting is still nontrivial as, in the strong formulation, the noise and filtration are still control-dependent. However, since the cost functional only depends on the distribution of the underlying noise and there are optimal controls in feedback form, we can first study an auxiliary

control problem with a fixed Brownian motion with respect to a fixed filtration to construct optimizers and then replace the driving Brownian motion and filtration with the appropriate control-dependent innovations process and filtration.

5.2.1. Piecewise constant controls. The control problem with piecewise constant controls is formulated with respect to the original Brownian motion W on our probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ and with respect to the filtration $\mathcal{F}^W$ generated by W and augmented by the $\mathbb{P}$-nullsets.

For each $n \in \mathbb{N}$ let $\delta_n := T2^{-n}$ be the dyadic step size of order n and define the associated time and space-time grid

$$\mathbb{T}^n := \{k\delta_n : k = 0, \dots, 2^n\} \quad \text{and} \quad \mathbb{S}^n := \mathbb{T}^n \times \mathbb{R}^{\tilde{m}}.$$

With this, the set of piecewise constant controls is given by

$$\mathcal{A}^n := \big\{u : [0,T] \times \Omega \to \mathcal{U} : u \text{ is } \mathcal{F}^W\text{-progressively measurable and constant on } ((k-1)\delta_n, k\delta_n] \text{ for all } k = 1, \dots, 2^n\big\}.$$

Observe that $\mathcal{A}^n \subset \mathcal{A}^{pre}$, but in general $\mathcal{A}^n \not\subseteq \mathcal{A}$ since we assume the piecewise constant controls to be $\mathcal{F}^W$-progressive. In any case, for $u \in \mathcal{A}^n$ and $(t,x) \in \mathbb{S}$, the associated state process $\hat{X}^{u;t,x} = (\hat{A}^{u;t,x}, \hat{\Upsilon}^{u;t,x}, \hat{\Gamma}^{u;t,x})$ given as the unique strong solution of

$$\mathrm{d}\hat{X}_s^{u;t,x} = f(s, \hat{X}_s^{u;t,x}, u_s)\mathrm{d}s + \Sigma(s, \hat{X}_s^{u;t,x}, u_s)\mathrm{d}W_s, \quad \hat{X}_t^{u;t,x} = x \tag{5.1}$$

with driving noise W is well-defined. With this, the cost functional for the piecewise constant control problem is defined as

$$\hat{\mathcal{J}}(u;t,x) := \mathbb{E}\Big[\int_t^T \tilde{k}(s, \hat{X}_s^{u;t,x}, u_s)\mathrm{d}s + \tilde{g}(\hat{X}_T^{u;t,x})\Big] \qquad \text{subject to (5.1)}$$

with associated value function $\hat{V}^n : \mathbb{S} \to \mathbb{R}$ given by

$$\hat{V}^n(t,x) := \inf_{u \in \mathcal{A}^n} \hat{\mathcal{J}}(u;t,x), \qquad (t,x) \in \mathbb{S}.$$

By Theorem 3.2.2 in [36], for each $t \in [0,T]$ fixed the mapping $x \mapsto \hat{\mathcal{J}}(u;t,x)$ is continuous, uniformly with respect to $u \in \mathcal{A}^n$, implying that also $x \mapsto \hat{V}^n(t,x)$ is continuous. With this and using the pseudo-Markov property for piecewise constant controls established in [36, Lemma 3.2.14] (see also [14] for a discussion of the importance of the pseudo-Markov property), it follows from classical arguments that the piecewise constant control problem satisfies the following version of the DPP, see Appendix A.3 for the proof.

PROPOSITION 5.2. *Let $(t,x) \in \mathbb{S}^n$ with $t < T$, and for each $u \in \mathcal{U}$ denote by $\hat{X}^{u;t,x}$ the state process with constant control u. Then it holds that*

$$\hat{V}^n(t,x) = \inf_{u \in \mathcal{U}} \mathbb{E}\Big[\int_t^{t+\delta_n} \tilde{k}(s, \hat{X}_s^{u;t,x}, u)\mathrm{d}s + \hat{V}^n(t+\delta_n, \hat{X}_{t+\delta_n}^{u;t,x})\Big]. \tag{5.2}$$

The advantage of the DPP is that it gives us a convenient way to construct optimal piecewise constant controls.

THEOREM 5.3. *For each $n \in \mathbb{N}$, there exists a measurable function $U_n^* : \mathbb{S} \to \mathcal{U}$ such that for each $(t,x) \in \mathbb{S}$ the SDE*

$$\mathrm{d}\hat{X}_s^{*;t,x} = f\big(s, \hat{X}_s^{*;t,x}, U_n^*(s, \hat{X}_s^{*;t,x})\big)\mathrm{d}s + \Sigma\big(s, \hat{X}_s^{*;t,x}, U_n^*(s, \hat{X}_s^{*;t,x})\big)\mathrm{d}W_s$$

with $\hat{X}_s^{*;t,x} = x$ for all $s \in [0,t]$ admits a unique strong solution and such that the control process

$$u_s^* := U_n^*(s, \hat{X}_s^{*;t,x}), \quad s \in [0,T]$$

is admissible and optimal for the piecewise constant control problem, that is

$$u^* \in \mathcal{A}^n \quad \text{and} \quad \hat{\mathcal{J}}^n(u^*;t,x) = \hat{V}^n(t,x).$$

Proof. It is sufficient to construct U_n^* on $\mathbb{S}^n$ and extend it as a piecewise constant function to $\mathbb{S}$. In particular, this guarantees that the control u^* is indeed piecewise constant. Let therefore $t \in \mathbb{T}^n$ with $t < T$. According to Corollary 3.2.8 in [36], the mapping

$$(x,u) \mapsto \mathbb{E}\Big[\int_t^{t+\delta_n} \tilde{k}(s, \hat{X}_s^{u;t,x}, u)\mathrm{d}s + \hat{V}^n(t+\delta_n, \hat{X}_{t+\delta_n}^{u;t,x})\Big]$$

is continuous on $\mathbb{R}^{\tilde{m}} \times \mathcal{U}$. We may therefore apply the measurable selection result [42, Theorem 2] to obtain a measurable optimizer $U_n^* : \mathbb{S}^n \to \mathcal{U}$ of the right-hand side of the DPP (5.2). Clearly, this function satisfies the desired properties. ∎

With our hands on an optimal feedback control for the piecewise constant control problem, we can now draw the connection to the full information problem in the strong formulation.

PROPOSITION 5.4. *Let $n \in \mathbb{N}$ and for $(t,x) \in \mathbb{S}$ let $u^* \in \mathcal{A}^n$ be the optimal control for $\hat{V}^n(t,x)$ constructed in Theorem 5.3. Then $u^* \in \mathcal{A}$ and*

$$V(t,x) \le \mathcal{J}(u^*;t,x) = \hat{\mathcal{J}}^n(u^*;t,x) = \hat{V}^n(t,x).$$

Proof. Let us first observe that $u^* \in \mathcal{A}^{pre}$ and denote by $\hat{X} = (\hat{A}, \hat{\Upsilon}, \hat{\Gamma})$ the state process associated with u^*. Since u^* is given in terms of a measurable function of $\hat{X}$, it follows that u^* is $\mathcal{F}^{\hat{X}}$-progressive. But $\mathcal{F}^{\hat{X}} = \mathcal{F}^{\hat{A}}$ and hence $u^* \in \mathcal{A}$. Finally, since the cost functional depends on the underlying Brownian motion only through its distribution, it follows that $\mathcal{J}(u^*;t,x) = \hat{\mathcal{J}}^n(u^*;t,x)$ from which we conclude. ∎

5.2.2. Convergence of the value functions. Up to this point, we have solved the piecewise constant control problem and argued that the constructed optimizer induces an admissible control in the full information problem in the strong formulation. It remains to argue that the value functions $\hat{V}^n$ converge to a viscosity subsolution of the HJB equation. This, however, is a standard argument since the DPP induces a monotone, consistent, and stable approximation scheme in the sense of [4].

To make this precise, let us fix $n \in \mathbb{N}$ and subsequently write $\mathfrak{M}(\mathbb{S}^n)$ for the space of real-valued measurable functions on $\mathbb{S}^n$. We introduce the approximation scheme at level n in terms of a mapping $S(n,\cdot) : \mathbb{S}^n \times \mathbb{R} \times \mathfrak{M}(\mathbb{S}^n) \to \mathbb{R}$ given by

$$S(n,t,x,v,w) := v - \inf_{u \in \mathcal{U}} \mathbb{E}\Big[\int_t^{t+\delta_n} \tilde{k}(s, \hat{X}_s^{u;t,x}, u)\mathrm{d}s + w(t+\delta_n, \hat{X}_{t+\delta_n}^{u;t,x})\Big], \quad t < T$$

and

$$S(n,T,x,v,w) := v - \tilde{g}(x).$$

Observe that the restriction of the piecewise constant value function $\hat{V}^n$ to $\mathbb{S}^n$ solves this scheme in the sense that

$$S\big(n,t,x,\hat{V}^n(t,x),\hat{V}^n\big)=0,\quad (t,x)\in\mathbb{S}^n.$$

Following Example 2 in [4], the scheme S is monotone, consistent, and stable and hence the relaxed limit $V^+:\mathbb{S}\to\mathbb{R}$ of the value functions $\hat{V}^n$, $n\in\mathbb{N}$, given by

$$V^+(t,x):=\limsup_{\substack{\mathbb{S}^n\ni(s,y)\to(t,x)\\ n\to\infty}}\hat{V}^n(s,y)$$

is an upper semicontinuous function and a viscosity subsolution of the HJB equation by Theorem 2.1 in [4]. Moreover, note that $V^+\geq V$ since $\hat{V}^n\geq V$ for all $n\in\mathbb{N}$.

THEOREM 5.5. *The relaxed limit $V^+:\mathbb{S}\to\mathbb{R}$ of the piecewise constant value functions $\hat{V}^n$, $n\in\mathbb{N}$, is an upper semicontinuous viscosity subsolution of the HJB equation satisfying $V^+(T,\cdot)=\tilde{g}$ and $V^+\geq V$. Moreover, there exist $C,q>0$ with*

$$0\leq V^+(t,x)\leq C(1+|x|^q)\qquad\forall(t,x)\in\mathbb{S}.$$

5.3. The supremum of subsolutions. It remains to show that the value function V^{weak} is bounded from below by a viscosity supersolution of the HJB equation. In order to achieve this, we rely on the notion of stochastic subsolutions associated with the weak control problem as formulated in section 4. These stochastic subsolutions are constructed in a way which guarantees that they are dominated by the value function V^{weak}, and their pointwise maximum is a viscosity supersolution of the HJB equation. Since the arguments leading to these results are standard and follow [6] very closely, we keep the exposition to a minimum.

DEFINITION 5.6. *The set of stochastic subsolutions of (HJB), denoted by $\mathcal{V}^-$, is the set of all lower semicontinuous functions $W:\mathbb{S}\to\mathbb{R}$ such that*

(1) there exist constants $C,q>0$ such that

$$W(T,x)\leq\tilde{g}(x)\quad\text{and}\quad 0\leq W(t,x)\leq C(1+|x|^q)\qquad\forall(t,x)\in\mathbb{S};$$

(2) for all $(t,x)\in\mathbb{S}$, any weak admissible control $U^{t,x}\in\mathcal{A}^{weak}$, and any pair of $\mathcal{F}^{t,x}$-stopping times $t\leq\tau\leq\rho\leq T$, we have

$$W(\tau,X_\tau^{t,x})\leq\mathbb{E}^{t,x}\Big[\int_\tau^\rho\tilde{k}(s,X_s^{t,x},u_s^{t,x})\mathrm{d}s+W(\rho,X_\rho^{t,x})\Big|\mathcal{F}_\tau^{t,x}\Big].$$

Since the function $W\equiv 0$ is clearly a stochastic subsolution, we see that $\mathcal{V}^-\neq\emptyset$. Moreover, the submartingale property and the terminal inequality directly show that

$$W(t,x)\leq\mathbb{E}^{t,x}\Big[\int_t^T\tilde{k}(s,X_s^{t,x},u_s^{t,x})\mathrm{d}s+\tilde{g}(X_T^{t,x})\Big]$$

for any weak control and hence $W\leq V^{weak}$. In particular, it follows that the pointwise supremum V^- of all stochastic subsolutions

$$V^-(t,x):=\sup_{W\in\mathcal{V}^-}W(t,x),$$

is dominated by V^{weak} and hence finite. Finally, as in [6, Theorem 4.1] with some minor but well-known adaptations as in [5] to account for our definition of stochastic subsolutions in terms of semicontinuous functions, we obtain the following key result.

THEOREM 5.7. *The supremum V^- of the set of stochastic subsolutions is a lower semicontinuous viscosity supersolution of the HJB equation satisfying $V^-(T,\cdot)=\tilde{g}$ and $V^-\leq V^{weak}$.*

5.4. Viscosity characterization and ε-optimal controls. It remains to piece together the results of the previous subsections to arrive at the main result of this article. Up to this point, we have argued that

$$V^- \leq V^{weak} \leq V \leq V^+,$$

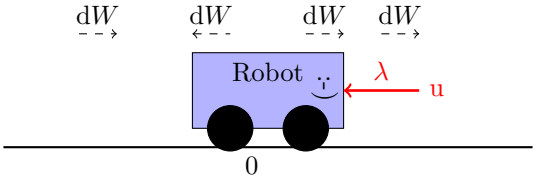
and V^-, V^+ are, respectively, viscosity super- and subsolutions of the HJB equation. Using the comparison principle, we therefore find that all functions above are in fact equal, and we have constructed ε-optimal controls.

THEOREM 5.8. *It holds that $V^- = V^{weak} = V = V^+$ is the unique continuous viscosity solution of the HJB equation in the class of nonnegative functions of polynomial growth with terminal value $\tilde{g}$. Moreover, for each $\varepsilon > 0$ and $(t,x) \in \mathbb{S}$, there exists $n \in \mathbb{N}$ such that $\hat{V}^n(t,x) \leq V(t,x) + \varepsilon$, and hence the optimal control associated with $\hat{V}^n(t,x)$ is ε-optimal for $V(t,x)$.*

Proof. We have $V^- \leq V^{weak} \leq V \leq V^+$ by construction. Moreover, V^- and V^+ are, respectively, lower and upper semicontinuous viscosity super- and subsolutions of the HJB equation satisfying $V^-(T,\cdot) = \tilde{g} = V^+(T,\cdot)$. The comparison principle hence applies, showing that $V^+ \leq V^-$, yielding the viscosity characterization. The existence of ε-optimal controls follows directly from the convergence $\hat{V}^n \to V^+ = V$ and Proposition 5.4. ∎

Remark 5.9. The approach used in this paper in fact gives a general way to construct ε-optimal controls for other control problems using stability of viscosity solutions and the stochastic Perron method.

6. Application to a robotics control problem. In this section, we present a toy application of our control methodology to a simple robotics control problem in a wind tunnel. The primary objective of this problem is to design an efficient and adaptive control strategy for a robot that is subjected to dynamic uncertain wind forces while moving on a horizontal one-dimensional plane. The goal is to maintain the robot's position as close to the center as possible while minimizing energy cost and adapting to the uncertainty of the motor's efficacy in the wind tunnel.



The robot is newly built and the one-dimensional efficacy λ of the motor is uncertain in this environment. It is subject to wind $\mathrm{d}W$ pushing it back and forth on the one-dimensional plane. Only the position Y^u of the robot on the horizontal plane can be observed, in particular we cannot directly observe the efficacy of the control u through the motor or the wind W. As a consequence, λ must be estimated online from the position Y^u of the robot. The energy cost is taken into account quadratically (cost of control) and the robot should be kept near the center. Deviation is penalized quadratically during the task and at the end. We choose coefficient and cost functions

$$b(t,y,u) = u, \qquad \sigma(t,y,u) = \sigma_0,$$
$$k(t,y,u,\ell) = cy^2 + \rho u^2, \qquad g(y,\ell) = Cy^2,$$

where the model parameters are given by

$$\sigma_0 = 1, \qquad T = 1, \qquad \rho = 2, \qquad c = 2, \qquad \text{and} \qquad C = 5.$$

In the case of an observable efficacy $\lambda \in \mathbb{R}$, the problem reduces to a standard stochastic linear-quadratic control problem which can be solved explicitly up to the solution of a system of Riccati differential equations. To be precise, the value function in the observable case is of the form $V^{\lambda}_{LQ}(t,a) = f^{\lambda}_1(t)a^2 + f^{\lambda}_2(t)$, where $f^{\lambda}_1, f^{\lambda}_2 : [0,T] \to \mathbb{R}$ solve

$$0 = \dot{f}^{\lambda}_1(t) - \frac{\lambda^2 (f^{\lambda}_1)^2(t)}{\rho^2} + c \qquad \text{and} \qquad 0 = \dot{f}^{\lambda}_2(t) + f^{\lambda}_1(t)$$

with terminal condition $f^{\lambda}_1(T) = C, f^{\lambda}_2(T) = 0$. The feedback map $u^{\lambda}_{LQ} : [0,T] \times \mathbb{R} \to \mathbb{R}$ for the optimal control in this problem is given by

$$u^{\lambda}_{LQ}(t,a) = -\frac{\lambda \partial_a V^{\lambda}_{LQ}(t,a)}{2\rho} = -\frac{\lambda f^{\lambda}_1(t)}{\rho} a.$$

The case of an unobservable λ does not admit a closed-form solution and has to be solved numerically. Here, we assume that λ is uniformly distributed over $[0,1]$ and compare the numerical approximation of an optimal control for this problem with two benchmark controls obtained from the problem with observable λ. The first one, the *naïve control*, is constructed by replacing the random λ by its mean $\bar{\lambda} := \mathbb{E}[\lambda] = 0.5$, that is by considering the problem with observable efficacy chosen as $\bar{\lambda}$. In other words, the naïve control in feedback form is given by

$$u^{naive}(t,a) := u^{\bar{\lambda}}_{LQ}(t,a) = -\frac{\bar{\lambda} f^{\bar{\lambda}}_1(t)}{\rho} a.$$

The naïve control does no updating of the estimate of λ and thus does not account for learning. The second benchmark control, the *certainty equivalent* (CE) control, is constructed by, at each time $t \in [0,T]$, acting as if the conditional mean was the true λ, that is by replacing λ by its conditional mean $\mathbb{E}[\lambda|\mathcal{Y}^u_t]$ in the problem with observable efficacy. Using the Markovian representation of the conditional mean via G, the CE control is hence given

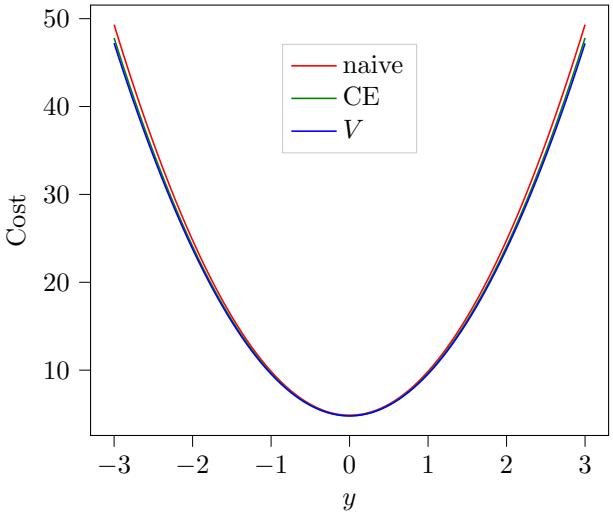
$$u^{CE}(t,x) := u^{G(\upsilon,\gamma)}_{LQ}(t,a).$$

The CE control is built on the idea that the conditional mean is the best approximation of λ but ignores the effect of the control on higher order moments, that is, it does not optimize for the dual effect. The expected cost V^{naive} and V^{CE} associated with the two benchmark controls u^{naive} and u^{CE} is computed by solving the linear PDE obtained by plugging the benchmark controls into the HJB equation.

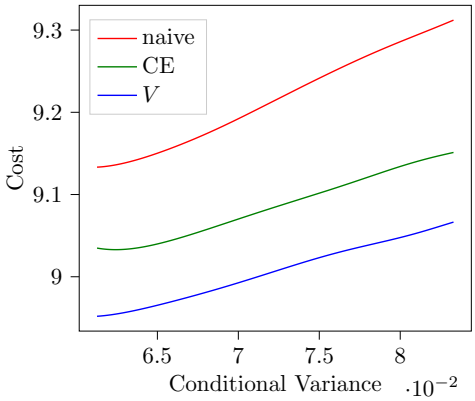
6.1. Numerical implementation and results. The controls and the associated expected costs are computed using a combination of the deep Galerkin method (DGM) and policy iteration on the respective PDEs. The choice of a deep learning method over classical finite difference methods comes from the observation that the latter methods are moderately inefficient due to the dimension of the state space being equal to $1 + 3$.

Regarding the implementation of the DGM, let us highlight that we do not approximate the value function directly, but rather approximate the value function by $(t,x) \mapsto (T-t)V_\theta(t,x) + \tilde{g}(x)$ where V_θ is a neural network parameterized by θ. This directly embeds the terminal condition into the approximating function. Second, we use the same DGM architecture as suggested in [43] with two layers for V_θ, and a simple two layer feedforward neural network for the approximating control. Each

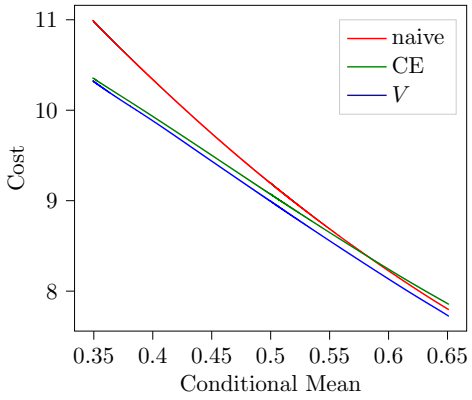
sub-layer has 512 nodes. We use the Adam optimizer with learning rate of 0.001 for value function and control and alternate gradient steps minimizing the infimum in the Hamiltonian and the DGM loss functional in a 1:1 relation. We use batches of 7500 points and obtain a terminal loss below 0.001 after approximately 16 000 training epochs. As an activation function, we use Sigmoid for both neural networks. The code is available on GitHub.[2]



(a) Cost vs. state y



(b) Cost vs. conditional variance



(c) Cost vs. conditional mean

Figure 6.1a compares the expected cost as functions of the initial state y, for fixed time and auxiliary states $(t, \upsilon, \gamma) = (0, 0, 0)$. It shows a small difference between the cost of the adaptive control and the cost induced by the naïve and CE control, with the adaptive control leading to the smallest total cost overall.

In Figure 6.1b, we fix time $t = 0.1$ and state $y = 1$, and a target conditional variance of 0.07. We then identify pairs (υ, γ) such that $G_\upsilon(\upsilon, \gamma) \approx 0.07$ and plot

[2] https://github.com/AlexanderMerkel/Optimal-adaptive-control-with-separable-drift-uncertainty

the expected cost with (t, y) fixed as a function of the conditional mean $G(\upsilon, \gamma)$. A similar process is used for Figure 6.1c, where we plot the expected cost as a function of the conditional variance with a target conditional mean of 0.52. The numerical results show that there is a *substantial* difference in the control actions and resulting costs, thus suggesting a significant advantage of using the adaptive control over the naïve and CE control.

More precisely, Figure 6.1b and Figure 6.1c show that the CE control and the adaptive control outperform the naïve control. We furthermore observe in Figure 6.1b that the expected cost is decreasing in conditional mean (which is expected, as we plot for state $y = 1$). More significantly, we see in Figure 6.1c that the adaptive control shows a substantial difference compared to the CE control. Moreover, for all three controls the expected cost is increasing in the conditional variance, illustrating the failure of the separation principle in this context.

Acknowledgements. This research was supported by the Deutsche Forschungsgemeinschaft through the Berlin–Oxford IRTG 2544: Stochastic Analysis in Interaction. SC also acknowledges the support of the UKRI Prosperity Partnership Scheme (FAIR) under EPSRC Grant EP/V056883/1, the Alan Turing Institute, and the Oxford–Man Institute for Quantitative Finance.

Appendix A. Additional Computations and Proofs.

A.1. Computations of Subsection 3.1. In this appendix we elaborate on the calculations of subsection 3.1 on the effect of the control of higher order conditional moments of λ in the one-dimensional case for simplicity. By Itô's formula, we have

$$\begin{aligned}\mathrm{d}m_t^u = \mathrm{d}G(\Upsilon_t^u, \Gamma_t^u) = \frac{b^2}{\sigma^2}(t, Y_t^u, u_t)\Big[G_\upsilon G + G_\gamma + \frac{1}{2}G_{\upsilon\upsilon}\Big](\Upsilon_t^u, \Gamma_t^u)\mathrm{d}t \\ + G_\upsilon(\Upsilon_t^u, \Gamma_t^u)\frac{b}{\sigma}(t, Y_t^u, u_t)\mathrm{d}V_t^u.\end{aligned}$$

Expressing the partial derivatives of G in terms of F, it follows that

$$G_\upsilon G + \frac{1}{2}G_{\upsilon\upsilon} + G_\gamma = \frac{1}{F}\frac{\partial}{\partial\upsilon}\Big(\frac{1}{2}F_{\upsilon\upsilon} + F_\gamma\Big) - \frac{F_\upsilon}{F^2}\Big(\frac{1}{2}F_{\upsilon\upsilon} + F_\gamma\Big).$$

A direct computation shows that F satisfies the backward heat equation $F_\gamma = -\frac{1}{2}F_{\upsilon\upsilon}$, and we conclude that $G_\upsilon G + \frac{1}{2}G_{\upsilon\upsilon} + G_\gamma = 0$ and therefore

$$\mathrm{d}m_t^u = G_\upsilon(\Upsilon_t^u, \Gamma_t^u)\frac{b}{\sigma}(t, Y_t^u, u_t)\mathrm{d}V_t^u.$$

Finally, again by Itô's formula, it follows that

$$\begin{aligned}\mathrm{dvar}_t^u &= \mathrm{d}G_\upsilon(\Upsilon_t^u, \Gamma_t^u) \\ &= \Big[G_{\upsilon\upsilon}G + G_{\upsilon\gamma} + \frac{1}{2}G_{\upsilon\upsilon\upsilon}\Big](\Upsilon_t^u, \Gamma_t^u)\frac{b^2}{\sigma^2}(t, Y_t^u, u_t)\mathrm{d}t \\ &\quad + G_{\upsilon\upsilon}(\Upsilon_t^u, \Gamma_t^u)\frac{b}{\sigma}(t, Y_t^u, u_t)\mathrm{d}V_t^u.\end{aligned}$$

Similarly to the calculations above, using the backward heat equation multiple times, we find that

$$G_{\upsilon\upsilon}G + G_{\upsilon\gamma} + \frac{1}{2}G_{\upsilon\upsilon\upsilon} = -G^2(G_2 + G^2),$$

and we conclude that

$$\mathrm{dvar}_t^u = -\Big[G^2(G_2 + G^2)\Big](\Upsilon_t^u, \Gamma_t^u)\frac{b^2}{\sigma^2}(t, Y_t^u, u_t)\mathrm{d}t + G_{\upsilon\upsilon}(\Upsilon_t^u, \Gamma_t^u)\frac{b}{\sigma}(t, Y_t^u, u_t)\mathrm{d}V_t^u.$$

A.2. Proof of Theorem 5.1. This appendix is dedicated to the proof of the comparison principle Theorem 5.1. We follow the classical line of argument based on perturbing the viscosity subsolution by a strict classical subsolution to control its behavior at infinity, an idea which appeared in the literature as early as [28].

Proof of Theorem 5.1. Fix $\tilde{q} \geq 2$ with $\tilde{q} > q$ and define $\Psi : \mathbb{S} \to (-\infty, 0]$ by

$$\psi(t,x) := -|x|^{\tilde{q}} \exp(\zeta(T-t)) - M(1+T-t), \qquad (t,x) \in \mathbb{S}.$$

If the constants $\zeta, M > 0$ are chosen sufficiently large, a direct computation shows that Ψ is a strict classical subsolution of (HJB). For $\rho > 1$, define the perturbation

$$U^\rho := \frac{\rho - 1}{\rho} U + \frac{1}{\rho}\psi.$$

As in [28], it follows that there exists a continuous function $\kappa : \mathbb{S} \to (0, \infty)$ such that U^ρ is a viscosity subsolution of the perturbed HJB equation

$$\mathrm{F}\big(\cdot, \partial_t U^\rho, \mathrm{D}_x U^\rho, \mathrm{D}_x^2 U^\rho\big) + \frac{\kappa}{\rho} = 0 \quad \text{on } [0,T) \times \mathbb{R}^{\tilde{m}},$$

where $\mathrm{F} : \mathbb{S} \times \mathbb{R} \times \mathbb{R}^{\tilde{m}} \times \mathcal{S}_{\tilde{m}} \to \mathbb{R}$ is given by

$$\mathrm{F}(t,x,p,q,Q) := -p - \inf_{u \in \mathcal{U}} \Big\{ f(t,x,u)q + \frac{1}{2}\operatorname{tr}\big[(\Sigma(t,x,u)\Sigma(t,x,u)^\intercal Q\big] + \tilde{k}(t,x,u) \Big\}$$

for all $(t,x,p,q,Q) \in \mathbb{S} \times \mathbb{R} \times \mathbb{R}^{\tilde{m}} \times \mathcal{S}_{\tilde{m}}$. We proceed to show that $U^\rho \leq W$, implying that $U \leq W$ by sending $\rho \to \infty$. We argue by contradiction by assuming that there exists $(t^*, x^*) \in \mathbb{S}$ with

$$U^\rho(t^*,x^*) - W(t^*,x^*) > 0. \tag{A.1}$$

Next, for all $k \in \mathbb{N}$, we then define a function

$$\varphi_k(t,x,\hat{x}) := U^\rho(t,x) - W(t,\hat{x}) - \frac{k}{2}|x - \hat{x}|^2$$

on the domain $\mathfrak{S} := [0,T] \times \mathbb{R}^{\tilde{m}} \times \mathbb{R}^{\tilde{m}}$ and set

$$\Theta_k := \sup_{(t,x,\hat{x}) \in \mathfrak{S}} \varphi_k(t,x,\hat{x}) \qquad \text{and} \qquad \Theta := \sup_{(t,x) \in \mathbb{S}} \varphi_0(t,x,x).$$

Using (A.1), we find that

$$0 < U^\rho(t^*,x^*) - W(t^*,x^*) \leq \Theta \leq \Theta_{k+1} \leq \Theta_k \leq \Theta_0, \qquad k \in \mathbb{N}. \tag{A.2}$$

Next, using that U and W are non-negative, the growth assumption on U, and $\tilde{q} > q$, we find that $\Theta_0 < \infty$. Now a standard argument as for example in step 1 of the proof of Theorem 5.4 in [8] shows that there exists a sequence $(t_k, x_k, \hat{x}_k) \in \mathfrak{S}$ such that $(t_k, x_k, \hat{x}_k)$ maximizes Θ_k, and a pair $(\bar{t}, \bar{x}) \in \mathbb{S}$ such that

$$\begin{aligned}
\lim_{k\to\infty} (t_k, x_k) &= \lim_{k\to\infty} (t_k, \hat{x}_k) = (\bar{t}, \bar{x}), & \lim_{k\to\infty} \frac{k}{2}|x_k - \hat{x}_k|^2 &= 0, \\
\lim_{k\to\infty} U^\rho(t_k, x_k) &= U^\rho(\bar{t}, \bar{x}), & \lim_{k\to\infty} W(t_k, \hat{x}_k) &= W(\bar{t}, \bar{x}), \\
\lim_{k\to\infty} \Theta_k &= \Theta = U^\rho(\bar{t}, \bar{x}) - W(\bar{t}, \bar{x}).
\end{aligned}$$

From this we also obtain that $\bar{t} < T$, as else we get the contradiction

$$0 < \Theta = U^\rho(\bar{t},\bar{x}) - W(\bar{t},\bar{x}) \le U(T,\bar{x}) - W(T,\bar{x}) + \frac{1}{\rho}\psi(T,\bar{x}) \le 0,$$

by the terminal inequality $U(T,\,\cdot\,) \le W(T,\,\cdot\,)$. Hence, without loss of generality, we may assume $t_k < T$, that is $(t_k,x_k),(t_k,\hat{x}_k) \in [0,T)\times\mathbb{R}^{\tilde{m}}$ for all $k\in\mathbb{N}$. From Ishii's lemma, see Theorem 8.3 in [16], for each k we find $M_k,\hat{M}_k \in \mathcal{S}_{\tilde{m}}$ satisfying[3]

$$\begin{pmatrix} M_k & 0 \\ 0 & -\hat{M}_k \end{pmatrix} \le \begin{pmatrix} \mathbf{I}_{\tilde{m}} & -\mathbf{I}_{\tilde{m}} \\ -\mathbf{I}_{\tilde{m}} & \mathbf{I}_{\tilde{m}} \end{pmatrix} \tag{A.3}$$

and constants $q_k = -\hat{q}_k$ such that[4]

$$\big(q_k, k(x_k-\hat{x}_k), M_k\big) \in \bar{J}^{2,+}U^\rho(t_k,x_k), \quad \big(\hat{q}_k, k(x_k-\hat{x}_k), \hat{M}_k\big) \in \bar{J}^{2,-}W(t_k,\hat{x}_k).$$

Using the subsolution property of U^ρ, the supersolution property of W, and letting $\kappa := \sup_{k\in\mathbb{N}}\max\{\kappa(t_k,x_k),\kappa(t_k,\hat{x}_k\} > 0$, it follows from (A.3) and Lipschitz-continuity of f and Σ that there exists a constant $C>0$ such that

$$\begin{aligned} 0 < \bar{\kappa} &\le \mathrm{F}\big(t_k,x_k,q_k,k(x_k-\hat{x}_k),M_k\big) - \mathrm{F}\big(t_k,\hat{x}_k,\hat{q}_k,k(x_k-\hat{x}_k),\hat{M}_k\big) \\ &\le Ck|x_k-\hat{x}_k|^2 + \sup_{u\in U}\big\{\tilde{k}(t_k,x_k,u) - \tilde{k}(t_k,\hat{x}_k,u)\big\}, \quad k\in\mathbb{N}. \end{aligned}$$

As the right-hand side tends to zero as $k\to\infty$, this is the desired contradiction. ∎

A.3. Proof of Proposition 5.2.

Proof of Proposition 5.2. Step 1: Let $(t,x)\in\mathbb{S}^n$ with $t<T$, $u\in\mathcal{U}$ and $\hat{u}\in\mathcal{A}^n$ such that $\hat{u}=u$ on $[0,t+\delta_n]$. According to Lemma 3.2.14 in [36], we have

$$\mathbb{E}\Big[\int_{t+\delta_n}^T \tilde{k}(s,\hat{X}_s^{\hat{u};t,x},\hat{u}_s)\mathrm{d}s + \tilde{g}(\hat{X}_s^{\hat{u};t,x})\Big|\mathcal{F}^W_{t+\delta_n}\Big] = \hat{\mathcal{J}}\big(\hat{u};t+\delta_n,\hat{X}^{u;t,x}_{t+\delta_n}\big).$$

But then the tower property of conditional expectation yields

$$\hat{V}(t,x) \ge \inf_{u\in\mathcal{U}}\mathbb{E}\Big[\int_t^{t+\delta_n}\tilde{k}(s,\hat{X}_s^{u;t,x},u)\mathrm{d}s + \hat{V}^n\big(t+\delta_n,\hat{X}^{u;t,x}_{t+\delta_n}\big)\Big].$$

Step 2: We fix $\varepsilon>0$. Since $\hat{\mathcal{J}}(u;t+\delta_n,\,\cdot\,)$ and $\hat{V}(t+\delta_n,\,\cdot\,)$ are continuous (uniformly in $u\in\mathcal{A}$) by Lemma 3.2.2 in [36], for any $y\in\mathbb{R}^{\tilde{m}}$ there exists $\rho>0$ with

$$|\hat{\mathcal{J}}(u;t+\delta_n,y) - \hat{\mathcal{J}}(u;t+\delta_n,\hat{y})| + |\hat{V}(t+\delta_n,y) - \hat{V}(t+\delta_n,\hat{y})| \le \frac{1}{3}\varepsilon$$

for all $u\in\mathcal{A}$ and $\hat{y}\in B_\rho(y)$, where $B_\rho(y)$ is the open ball of radius ρ centered around y. Observe that there exist sequences $\{y_k\}_{k\in\mathbb{N}}\subset\mathbb{R}^{\tilde{m}}$, $\{\rho_k\}_{k\in\mathbb{N}}\subset(0,\infty)$ and a Borel-partition $\{B_k\}_{k\in\mathbb{N}}$ of $\mathbb{R}^{\tilde{m}}$ such that $y_k\in B_k\subset B_{\rho_k}(y_k)$ for all $k\in\mathbb{N}$. Next, choose an $(\varepsilon/3)$-optimal control $u_k\in\mathcal{A}^n$ for $\hat{V}(t+\delta_n,y_k)$, so that is follows that

$$\hat{\mathcal{J}}(u_k;t+\delta_n,y) \le \hat{\mathcal{J}}(u_k;t+\delta_n,y_k) + \frac{1}{3}\varepsilon \le \hat{V}(t+\delta_n,y_k) + \frac{2}{3}\varepsilon \le \hat{V}(t+\delta_n,y) + \varepsilon$$

[3]Here, $\mathbf{I}_{\tilde{m}}$ denotes the identity matrix in $\mathcal{S}_{\tilde{m}}$.

[4]Here, $\bar{J}^{2,+}U^\rho(t_k,x_k)$ and $\bar{J}^{2,-}W(t_k,\hat{x}_k)$ denote the closures of second order super- and subjets of U^ρ and W, respectively.

for all $y \in B_k$ and $k \in \mathbb{N}$. Now fix $u \in \mathcal{U}$ and consider the control $\hat{u}$ given by

$$\hat{u} := u \quad \text{on } [0, t+\delta_n] \qquad \text{and} \qquad \hat{u} := \sum_{k=1}^{\infty} u_k \mathbb{1}_{\{\hat{X}^{u;t,x}_{t+\delta_n} \in B_k\}} \quad \text{on } (t+\delta_n, T].$$

Clearly, $\hat{u} \in \mathcal{A}^n$ and we conclude that

$$\begin{aligned}\hat{V}^n(t,x) &\leq \mathbb{E}\Big[\int_t^{t+\delta_n} \tilde{k}(s, \hat{X}^{u;t,x}_s, u)\mathrm{d}s + \sum_{k=1}^{\infty} \mathbb{1}_{\{\hat{X}^{u;t,x}_{t+\delta_n} \in B_k\}} \hat{\mathcal{J}}\big(u_k; t+\delta_n, \hat{X}^{u;t,x}_{t+\delta_n}\big)\Big] \\ &\leq \mathbb{E}\Big[\int_t^{t+\delta_n} \tilde{k}(s, \hat{X}^{u;t,x}_s, u)\mathrm{d}s + \hat{V}^n\big(t+\delta_n, \hat{X}^{u;t,x}_{t+\delta_n}\big)\Big] + \varepsilon.\end{aligned}$$

Sending $\varepsilon \downarrow 0$ and taking the infimum over all $u \in \mathcal{U}$ yields the result. ∎

REFERENCES

[1] A. Astolfi, D. Karagiannis, and R. Ortega, *Nonlinear and Adaptive Control with Applications*, Springer, 2008.

[2] E. Bandini, A. Cosso, M. Fuhrman, and H. Pham, *Backward SDEs for optimal control of partially observed path-dependent stochastic systems: A control randomization approach*, The Annals of Applied Probability, 28 (2018), pp. 1634 – 1678, https://doi.org/10.1214/17-AAP1340.

[3] Y. Bar-Shalom and E. Tse, *Dual effect, certainty equivalence, and separation in stochastic control*, IEEE Trans. Autom. Control, 19 (1974), pp. 494–500.

[4] G. Barles and P. E. Souganidis, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptot. Anal., 4 (1991), pp. 271–283.

[5] E. Bayraktar and M. Sîrbu, *Stochastic Perron"'s method and verification without smoothness using viscosity comparison: the linear case*, Proc. Amer. Math. Soc., 140 (2012), pp. 3645–3654.

[6] E. Bayraktar and M. Sîrbu, *Stochastic Perron's method for Hamilton–Jacobi–Bellman equations*, SIAM J. Control Optim., 51 (2013), pp. 4274–4294.

[7] C. Belak, A. Chen, C. Mereu, and R. Stelzer, *Optimal investment with time-varying stochastic endowments*, SIAM J. Financ. Math., 13 (2022), pp. 969–1003.

[8] C. Belak, L. Mich, and F. T. Seifried, *Optimal investment for retail investors*, Math. Finance, 32 (2022), pp. 555–594.

[9] V. E. Beneš, I. Karatzas, and R. Rishel, *The separation principle for a Bayesian adaptive control problem with no strict-sense optimal law*, Stoch. Monogr., 5 (1991).

[10] A. Bensoussan, *Stochastic Control of Partially Observable Systems*, Cambridge University Press, 1992.

[11] J.-M. Bismut, *Partially observed diffusions and their control*, SIAM Journal on Control and Optimization, 20 (1982), pp. 302–309, https://doi.org/10.1137/0320023.

[12] P. Caines and H. Chen, *Optimal adaptive LQG control for systems with finite state process parameters*, IEEE Trans. Autom. Control, 30 (1985), pp. 185–189.

[13] Á. Cartea, S. Jaimungal, and J. Penalva, *Algorithmic and High-Frequency Trading*, Cambridge University Press, 2015.

[14] J. Claisse, D. Talay, and X. Tan, *A pseudo-Markov property for controlled diffusion processes*, SIAM J. Control Optim., 54 (2016), pp. 1017–1029.

[15] S. N. Cohen and R. J. Elliott, *Stochastic Calculus and Applications*, vol. 2, Springer, 2015.

[16] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc., 27 (1992), pp. 1–67.

[17] T. E. Duncan, L. Guo, and B. Pasik-Duncan, *Adaptive continuous-time linear quadratic Gaussian control*, IEEE Trans. Autom. Control, 44 (1999), pp. 1653–1662.

[18] T. E. Duncan and B. Pasik-Duncan, *Adaptive control of continuous-time linear stochastic systems*, Math. Control Signals Syst., 3 (1990), pp. 45–60.

[19] E. Ekström, I. Karatzas, and J. Vaicenavicius, *Bayesian sequential least-squares estimation for the drift of a Wiener process*, Stoch. Proc. Appl., 145 (2022), pp. 335–352.

[20] N. El Karoui, D. H. Nguyen, and M. Jeanblanc-Picqué, *Existence of an optimal markovian filter for the control under partial observations*, SIAM Journal on Control and Optimization, 26 (1988), pp. 1025–1061, https://doi.org/10.1137/0326057.

[21] A. A. Feldbaum, *Dual control theory. i*, Avtom. i Telemekhanika, 21 (1960), pp. 1240–1249.
[22] W. H. Fleming and É. Pardoux, *Optimal control for partially observed diffusions*, SIAM J. Control Optim., 20 (1982), pp. 261–285.
[23] T. T. Georgiou and A. Lindquist, *The separation principle in stochastic control, redux*, IEEE Transactions on Automatic Control, 58 (2013), pp. 2481–2494.
[24] O. Guéant, *The Financial Mathematics of Market Liquidity: From optimal execution to market making*, CRC Press, 2016.
[25] H. V. Henderson and S. R. Searle, *Vec and vech operators for matrices, with some uses in jacobians and multivariate statistics*, The Canadian Journal of Statistics / La Revue Canadienne de Statistique, 7 (1979), pp. 65–81, http://www.jstor.org/stable/3315017 (accessed 2024-07-12).
[26] A. J. Heunis, *The innovations problem*, in Oxford Handbook of Nonlinear Filtering, Oxford University Press New York, 2011, pp. 425–449.
[27] K. Holkar and L. M. Waghmare, *An overview of model predictive control*, International Journal of control and automation, 3 (2010), pp. 47–63.
[28] K. Ishii, *Viscosity solutions of nonlinear second order elliptic PDEs associated with impulse control problems*, Funkcial. Ekvac, 36 (1993), pp. 123–141.
[29] D. Jiang, J. Sirignano, and S. N. Cohen, *Global convergence of deep Galerkin and PINNs methods for solving partial differential equations*. Preprint, available at https://arxiv.org/abs/2305.06000, 2023.
[30] I. Karatzas and D. L. Ocone, *The resolvent of a degenerate diffusion on the plane, with application to partially observed stochastic control*, Ann. Appl. Probab., (1992), pp. 629–668.
[31] I. Karatzas and D. L. Ocone, *The finite–horizon version for a partially–observed stochastic control problem of Beneš & Rishel*, Stoch. Anal. Appl., 11 (1993), pp. 569–605.
[32] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, 1998.
[33] I. Karatzas and X. Zhao, *Bayesian adaptive portfolio optimization*, in Option Pricing, Interest Rates and Risk Management, Cambridge University Press Cambridge, 2001, pp. 632–669.
[34] V. Krishnamurthy, *Partially observed Markov decision processes*, Cambridge university press, 2016.
[35] N. V. Krylov, *Approximating value functions for controlled degenerate diffusion processes by using piece-wise constant policies*, Electron. J. Probab., 4 (1999), pp. 1–19.
[36] N. V. Krylov, *Controlled Diffusion Processes*, Springer Science & Business Media, 2008.
[37] P. R. Kumar, *A survey of some results in stochastic adaptive control*, SIAM J. Control Optim., 23 (1985), pp. 329–380.
[38] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes I: General Theory*, vol. 5, Springer Science & Business Media, 2013.
[39] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes II: Applications*, vol. 6, Springer Science & Business Media, 2013.
[40] H. Mania, S. Tu, and B. Recht, *Certainty equivalence is efficient for linear quadratic control*, Adv. Neural Inf. Process. Syst., 32 (2019).
[41] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, Springer Science & Business Media, 2013.
[42] M. Schäl, *A selection theorem for optimization problems*, Arch. Math., 25 (1974), pp. 219–224.
[43] J. Sirignano and K. Spiliopoulos, *DGM: A deep learning algorithm for solving partial differential equations*, J. Comput. Phys., 375 (2018), pp. 1339–1364.
[44] L. Szpruch, T. Treetanthiploet, and Y. Zhang, *Exploration-exploitation trade-off for continuous-time episodic reinforcement learning with linear-convex models*. Preprint, available at https://arxiv.org/abs/2112.10264, 2021.
[45] K. T. Webster, *Handbook of Price Impact Modeling*, CRC Press, 2023.
[46] D. V. Widder, *Positive temperatures on an infinite rod*, Trans. Amer. Math. Soc., 55 (1944), pp. 85–95.
[47] W. M. Wonham, *On the separation theorem of stochastic control*, SIAM J. Control, 6 (1968), pp. 312–326.