An introduction to stochastic control and filtering

Adam Andersson
Chalmers University of Technology and the University of Gothenburg
Saab AB

June 28, 2025

Abstract

These are extended lecture notes for two lectures on stochastic control and filtering in the course *TMS165/MSA350 Stochastic Calculus* at Chalmers University of Technology and the University of Gothenburg. The notes act as an introduction to stochastic control and filtering, and for this purpose, both some technical details and more challenging settings have been set aside. Mathematical results are correct, but their proofs might contain gaps and in some cases need to be complemented with precise assumptions. In particular, measure-theoretical considerations are left out, and readers who know this can easily fill in such details. In this way, the principles are in focus, and the intended reader will be more ready to read the appropriate literature.

Contents

1 Introduction

These notes constitute a first draft of lecture notes in stochastic control and nonlinear filtering, and partially observed stochastic controls and games. In its current state, the presentation of the stochastic control paradigm of dynamic programming is about complete for finite- and infinite-horizon control. Nothing is written on the other paradigm of the stochastic maximum principle. Exact filtering with time-discrete observations is covered, but nothing on approximate filtering or filtering with time-continuous observations. The latter is mathematically interesting, but from an engineering perspective less so since modern digital sensors measure systems in discrete time. The notes are written as an introductory or complementary literature for anyone who wants to learn the covered topics rigorously but are intended to contribute with concepts and ideas rather than the whole rigorous theory. The reader is thus highly recommended to complement the reading with more complete texts, some of which are referenced. From an application point of view, the notes are written from the engineering (but still mathematical) perspective, although some comments on mathematical finance are provided. Moreover, the notes do not focus on numerical methods, although reformulations into optimization problems that are the foundation of deep learning based numerical methods are presented, and references to the respective methods are given.

The choice of including both stochastic control and filtering in the same notes is, from the control perspective, very natural, as system states are measured and estimated rather than known. However, from a pure filtering perspective, the choice might seem odd, even though stochastic control can be used as a tool, such as in the computational Doob-h transform; see [5]. This will be included in a future version. From a mathematical perspective, there are interesting similarities, not least between linear-quadratic control and linear filtering, the Kalman filter. These are both exactly solvable, and similar Riccati-type equations solve them.

Section 1.1 contains common notation. Sections 2 and 3 concern finite and infinite horizon control, respectively, with the dynamic programming paradigm. Section 4 introduces nonlinear filtering.

1.1 Notation

The space $C^{1,2}([0,T]\times\mathbb{R}^n)$ contains all real-valued continuous functions that have one continuous time derivative, two continuous space derivatives, and nonexisting mixed derivatives. For functions $\phi \in C^{1,2}([0,T]\times\mathbb{R}^n)$ we write ϕ_t for the partial derivative with respect to the time variable, ϕ_x and ϕ_{xx} the gradient and Hessian, respectively, in the space variable, respectively. For a function of only the space variable $\phi\colon \mathbb{R}^n \to \mathbb{R}$ we use the nabla notation $\nabla\phi$ and for a function of only the time variable $\psi\colon [0,T]\to V$ we use $\dot{\psi}$ for the time derivative. Here V is some normed vector space. In some calculations, we use $\frac{\partial}{\partial t}$ as an operator for the time derivative.

Throughout these notes, we use $(\|\cdot\|, \langle\cdot,\cdot\rangle)$ to denote the Euclidean norm and scalar product, i.e.,

$$\|v\| = \left(\sum_{i=1}^n v_i^2\right)^{\frac{1}{2}} \quad \text{and} \quad \langle u, v\rangle = \sum_{i=1}^n u_i v_i.$$

2 Finite horizon stochastic control

The objective of stochastic control is to control the trajectory of a noisy dynamical system in order to minimize a cost functional whose values depend on the trajectory. The noisy dynamical system is described by a stochastic differential equation. In this section, we consider stochastic control with a finite time horizon, which means that we consider an SDE on a finite time interval with terminal time $T > 0$. The cost functional is a sum of the cost at the terminal time and an integrated cost over the time interval, describing both the desired behavior of the system and the cost of controlling. In Section 2.1 we introduce the stochastic control problem.

The solution paradigm of dynamic programming, considered here, centers around the nonlinear Partial Differential Equation (PDE) called the Hamilton-Jacobi-Bellman (HJB) equation. In Section 2.2, we present a formal derivation of the HJB equation, and in Section 2.3 we prove a so-called *verification theorem*, solving the stochastic control problem. In certain settings, the HJB equation is analytically solvable, and Sections 2.4 and 2.5 concern two such settings. The first is the Linear-Quadratic control problem, a very important case, from both the theory and application point of view, with great structure being studied extensively. The second is the controlled Brownian motion, a smaller and less important class of problems, but introducing the famous Cole-Hopf transform, frequently used in the control literature. In Section 2.6 assumptions are listed from the PDE literature for uniform ellipticic problems, under which the HJB equation has classical and regular solutions. Uniform ellipticity of the HJB PDE is roughly equivalent to the controlled SDE having enough noise in every coordinate so that it nowhere degenerates into an Ordinary Differential Equation (ODE). There is a very important stochastic equivalent representation of the HJB equation in terms of Forward Backward Stochastic Differential Equations (FBSDE), the nonlinear Feynmann-Kac formula. It is introduced in Section 2.7 and Section 2.8 contains a digression into the SDE and FBSDE solution theory to explain the constructions and important and significant differences. Finally, Section 2.9 contains discounted stochastic control, which is important in mathematical finance and also in developing infinite-horizon stochastic control in Section 3.

2.1 The stochastic optimal control problem

Let $\mathcal{U}$ be a subset of $\mathbb{R}^k$-valued adapted stochastic processes $(u(t))_{t\in[0,T]}$. The processes in $\mathcal{U}$ are called *control processes* or *feasible controls*. Consider a controlled stochastic differential equation

$$\begin{cases} \mathrm{d}X^u(t) = f(t, X^u(t), u(t))\,\mathrm{d}t + \sigma(t, X^u(t), u(t))\,\mathrm{d}W(t), & t \in (0,T];\ u \in \mathcal{U}, \\ X^u(0) = x. \end{cases} \tag{2.1}$$

It is $\mathbb{R}^n$-valued and is controlled since it depends on the control process $u \in \mathcal{U}$. The Wiener process W is $\mathbb{R}^d$-valued. The objective of stochastic optimal control is to find a feasible control $u^* \in \mathcal{U}$ that minimizes the *cost functional*

$$J(u) = \mathbb{E}\left[\int_0^T \ell(t, X^u(t), u(t))\,\mathrm{d}t + g(X^u(T))\right].$$

Here ℓ and g are called the *running cost* and *terminal cost*, respectively. The special case with $g = 0$ is called a problem on *Lagrange form*, the case with $\ell = 0$ on *Meyer form* and the general case with $g \neq 0$ and $\ell \neq 0$ on *Bolza form*. A control process $u \in \mathcal{U}$ is a *Markov control policy* or *feedback control* if $u(t) = \pi(t, X^u(t))$ for some deterministic function π, called *Markov map* or *feedback map*. In the following, we sometimes refer to the minimization problem presented as *the control problem.*

The given setting can be used to model various problems in different fields. In engineering f can be the dynamics of a system described by Newton's laws of motion, and σ can represent vibrations or the fact that the controlling actuators, such as motors, are not ideal, i.e., that the control itself causes disturbances to the system. An example is a Segway that needs active control to stand and not fall. In mathematical finance, X can be the value of a portfolio and u an investment or consumption strategy. If the running cost ℓ includes a suitable *control cost* that penalizes the control, or the controls in $\mathcal{U}$ are restricted to some bounded set, then the controls are restricted from taking arbitrarily high values, and the problem is well-posed. Both properties are natural. In controlling a car, for example, the steering wheel, the gas and break pedals, and the gears all belong to a specified range. It is also natural that pressing the gas pedal has a cost in terms of fuel and that steering, breaking, and accelerating have costs in terms of passenger discomfort.

2.2 A formal derivation of the Hamilton–Jacobi–Bellman equation

The presentation of this section is based on [8, Sections III.6–III.7]. We turn to the synthesis of the control problem using an approach known as *dynamic programming*. We start by redefining the cost functional for varying times and initial values according to

$$J(t,x,u) = \mathbb{E}\left[\int_t^T \ell(s, X^u(s), u(s))\, \mathrm{d}s + g(X^u(T)) \,\middle|\, X^u(t) = x \right].$$

Then we define the *value function*

$$V(t,x) = \inf_{u\in\mathcal{U}} J(t,x,u), \tag{2.2}$$

which is central in dynamic programming.

With a formal calculation, we show that by assuming the so-called *Bellman's dynamic programming principle*, the value function satisfies a non-linear, second-order partial differential equation known as the *Hamilton–Jacobi–Bellman equation*, HJB equation. From the control problem, a Markov control policy is derived that depends on the solution to the HJB equation and its derivatives. Thus, solving the HJB equation solves the control problem. The calculation is formal and only serves an instructive purpose. The theory normally concerns an implication in the reverse direction, i.e., given that the HJB equation has a regular enough solution V, then this solution is the value function (2.2) and solves the control problem through the Markov control policy mentioned above. Such results are called *verification theorems* and we present two such theorems in these notes. Bellman's dynamic programming principle can be proved under the conditions of verification theorems, but is now an assumption. It is a sensible assumption considering the Markov property of the underlying SDE.

Bellman's Dynamic Programming Principle:

In words: Optimal control on a time interval $[t,T]$ is the same as optimal control in a subinterval $[t,\tau]$ and then continuing optimally in $[\tau,T]$ until the final time T.

In math: For all $t\in[0,T]$, $\tau\in(t,T]$ and $x\in\mathbb{R}$ we have

$$V(t,x) = \inf_{u\in\mathcal{U}} \mathbb{E}\left[\int_t^\tau \ell(s, X^u(s), u(s))\, \mathrm{d}s + V(\tau, X^u(\tau)) \,\middle|\, X(t) = x \right]. \tag{2.3}$$

Before we start the formal derivation, we state the Itô Formula.

Theorem 2.1 (The Itô Formula). *Let* $\varphi \in C^{1,2}([0,T] \times \mathbb{R}^n)$*,* $(W(t))_{t\in[0,T]}$ *be a Brownian motion,* $(f(t))_{t\in[0,T]}$*,* $(\sigma(t))_{t\in[0,T]}$ *be adapted (w.r.t. the filtration generated by* W*) stochastic processes satisfying standard assumptions and* $(X(t))_{t\in[0,T]}$ *satisfy*

$$X(t) = X(0) + \int_0^t f(t)\,\mathrm{d}t + \int_0^t \sigma(t)\,\mathrm{d}W(t), \quad t \in [0,T].$$

Then for all $0 \le t_0 \le t_1 \le T$ *it holds*

$$\begin{aligned}&\varphi(t_1, X(t_1)) - \varphi(t_0, X(t_0)) \\ &\quad = \int_{t_0}^{t_1} \varphi_t(s, X(s))\,\mathrm{d}s + \int_{t_0}^{t_1} (L\varphi)(s, X(s))\,\mathrm{d}s + \int_{t_0}^{t_1} \varphi_x(s, X(s))\sigma(s)\,\mathrm{d}W(s),\end{aligned}$$

where $L\varphi$ *is given by*

$$(L\varphi)(t,x) = \frac{1}{2}\,\mathrm{Tr}(\sigma^T(s)\sigma(s)\varphi_{xx}(t,x)) + \varphi_x(t,x)f(t).$$

We now formally derive the HJB equation from Bellman's dynamic programming principle. For notational simplicity, we take $d = k = n = 1$, i.e., 1-dimesional noise, control, and state. In a first step, we note that equality in (2.3) holds for the optimal control. Taking $\tau = t + h$ for $h > 0$ with $t + h \le T$ and replacing the optimal control with an arbitrary constant and deterministic control $u \equiv v \in \mathbb{R}$ in $[t, t+h]$ instead gives the inequality

$$V(t,x) \le \mathbb{E}\left[\int_t^{t+h} \ell(s, X(s), v)\,\mathrm{d}s + V(t+h, X(t+h)) \,\middle|\, X(t) = x\right]. \tag{2.4}$$

Subtracting $V(t,x)$ from both sides and dividing by h we have

$$\begin{aligned}0 &\le \frac{1}{h}\mathbb{E}\left[\int_t^{t+h} \ell(s, X(s), v)\,\mathrm{d}s + V(t+h, X(t+h)) - V(t,x) \,\middle|\, X(t) = x\right] \\ &= \frac{1}{h}\mathbb{E}\left[\int_t^{t+h} \ell(s, X(s), v)\,\mathrm{d}s \,\middle|\, X(t) = x\right] \\ &\quad + \frac{1}{h}\mathbb{E}\left[V(t+h, X(t+h)) - V(t,x) \,\middle|\, X(t) = x\right] \\ &= I_h^1(t,x,v) + I_h^2(t,x,v).\end{aligned} \tag{2.5}$$

First, changing the order of integration and expectation and since we condition on $X(t) = x$ we have that

$$\lim_{h\downarrow 0} I_h^1(t,x,v) = \lim_{h\downarrow 0} \frac{1}{h} \int_t^{t+h} \mathbb{E}\big[\ell(s, X(s), v) \,\big|\, X(t) = x\big]\,\mathrm{d}s = \ell(t,x,v). \tag{2.6}$$

Second, assuming that V is regular enough, using the Itô Formula and assuming the martingale property of the Itô integral, it holds that

$$
\begin{aligned}
&I_h^2(t,x,v) \\
&= \frac{1}{h}\mathbb{E}\bigg[\int_t^{t+h} \Big(\Big(\frac{\partial}{\partial t} + L^v\Big)V\Big)(s,X(s))\,\mathrm{d}s \,\Big|\, X(t)=x\bigg] \\
&\quad + \frac{1}{h}\mathbb{E}\bigg[\int_t^{t+h} V_x(s,X(s))\sigma(s,X(s),v)\,\mathrm{d}W(s) \,\Big|\, X(t)=x\bigg] \\
&= \frac{1}{h}\mathbb{E}\bigg[\int_t^{t+h} \Big(\Big(\frac{\partial}{\partial t} + L^v\Big)V\Big)(s,X(s))\,\mathrm{d}s \,\Big|\, X(t)=x\bigg] + 0.
\end{aligned}
$$

Here L^v is the second order partial differential operator

$$
(L^v\varphi)(t,x) = \frac{1}{2}\varphi_{xx}(t,x)\sigma^2(t,x,v) + \varphi_x(t,x)f(t,x,v),
$$

and φ_x and φ_{xx} are partial derivatives of φ. Similarly to the first limit in (2.6), changing the order of integration and conditional expectation, we have

$$
\begin{aligned}
\lim_{h\downarrow 0} I_h^2(t,x,v) &= \lim_{h\downarrow 0}\frac{1}{h}\int_t^{t+h} \mathbb{E}\bigg[\Big(\Big(\frac{\partial}{\partial t} + L^v\Big)V\Big)(s,X(s)) \,\Big|\, X(t)=x\bigg]\,\mathrm{d}s \\
&= \Big(\Big(\frac{\partial}{\partial t} + L^v\Big)V\Big)(t,x).
\end{aligned} \tag{2.7}
$$

Thus, for all $t\in[0,T]$, $x\in\mathbb{R}$, $v\in\mathbb{R}$ and all the assumptions that we made, we have from (2.5), (2.6) and (2.7) the differential inequality

$$
-V_t(t,x) \le (L^vV)(t,x) + \ell(t,x,v). \tag{2.8}
$$

Here we write for simplicity $V_t = \frac{\partial V}{\partial t}$.

We assume the existence of an optimal $u^*\in\mathcal{U}$ with an optimal Markov control policy π so that $u^*(t) = \pi(t,X(t))$, where

$$
\mathrm{d}X(t) = f(t,X(t),\pi(t,X(t)))\,\mathrm{d}t + \sigma(t,X(t),\pi(t,X(t)))\,\mathrm{d}W(t), \quad t\in(0,T].
$$

The dependence of u^* or equivalently π is suppressed for optimally controlled dynamics X. Then, by Bellman's dynamic programming principle, we have

$$
V(t,x) = \mathbb{E}\bigg[\int_t^{t+h} \ell(s,X(s),\pi(s,X(s)))\,\mathrm{d}s + V(t+h,X(t+h)) \,\Big|\, X(t)=x\bigg].
$$

By similar calculations as above, by the optimality of the Markov policy π, we have the equality

$$
-V_t(t,x) = \Big((L^vV)(t,x) + \ell(t,x,v)\Big)\Big|_{v=\pi(t,x)}. \tag{2.9}
$$

From (2.8) and (2.9) we conclude that

$$
-V_t(t,x) = \inf_{v\in\mathbb{R}}\Big((L^vV)(t,x) + \ell(t,x,v)\Big), \tag{2.10}
$$

and

$$
\pi(t,x) = \operatorname*{argmin}_{v\in\mathbb{R}}\Big((L^vV)(t,x) + \ell(t,x,v)\Big). \tag{2.11}
$$

In general, the argmin is not unique and is instead a set of minima. For cases with multiple minima, there should be $\in$ instead of $=$ in (2.11) and the reader can understand that the uniqueness of Markov control policies does not hold, and it is easy to construct discontinuous policies. We also notice that

$$J(T, x, u) = \mathbb{E}[g(X(T)) \mid X(T) = x] = g(x).$$

This gives a terminal condition to the HJB equation (2.10). Introducing the so-called Hamiltonian

$$\mathcal{H}(t, x, p, q) = \inf_{v \in \mathbb{R}} \Big(f(t, x, v)p + \frac{1}{2}\sigma^2(t, x, v)q + \ell(t, x, v) \Big),$$

allows us to write the HJB equation on the form

$$\begin{cases} -V_t(t, x) = \mathcal{H}(t, x, V_x, V_{xx}), \\ V(T, x) = g(x). \end{cases}$$

If σ depends on the control u, then the HJB equation is a fully nonlinear second order PDE, a quite non-trivial equation. If σ on the other hand does not depend on the control, then the HJB becomes a second order semi-linear equation, which is easier to work with by standard PDE methods. In this case, introducing the reduced Hamiltonian

$$\tilde{\mathcal{H}}(t, x, p) = \inf_{v \in \mathbb{R}} \Big(f(t, x, v)p + \ell(t, x, v) \Big),$$

the HJB equation can be written

$$\begin{cases} -V_t(t, x) = \frac{1}{2}V_{xx}\sigma(t, x)^2 + \tilde{\mathcal{H}}(t, x, V_x), \\ V(T, x) = g(x). \end{cases}$$

For many important applications (see Section 2.4 for one example) it is possible to analytically derive a function κ so that the optimal Markov control policy π is given by

$$\pi(t, x) = \kappa(t, x, V_x(t, x), V_{xx}(t, x)).$$

In such cases, we have

$$\mathcal{H}(t, x, p, q) = \Big(f(t, x, \kappa(t, x, p, q))p + \frac{1}{2}\sigma^2(t, x, \kappa(t, x, p, q))q + \ell(t, x, \kappa(t, x, p, q)) \Big).$$

With this form of the Hamiltonian, the PDE looks more standard without the infimum in the differential operator. We emphasize that deriving κ, when it exists and is analytically tractable, does not require solving the HJB equation itself.

2.3 A more general setting and a verification theorem

Here we lift the setting to multiple dimensions and make the problem a bit richer. The material is based on [8, Sections IV.2–IV.3] but is simplified by choosing a setting where weak solutions to SDE are not needed. First, the Brownian motion is d-dimensional. Second, the control processes take values in some set $U \subset \mathbb{R}^k$, third we consider an SDE that takes its values in some open and connected domain $O \subset \mathbb{R}^n$ with a boundary ∂O, and is stopped when $(t, X(t))_{t \geq 0}$ hits the boundary of the spatio-temporal domain $[0, T) \times O$. The stopping time when this occurs is called τ. The cost functional is modified to

$$J(u) = \mathbb{E}\left[\int_0^\tau \ell(t, X^u(t), u(t))\, \mathrm{d}t + g(\tau, X^u(\tau)) \right].$$

Thus, if the solution X^u leaves the domain O at a time $\tau < T$, then $g(\tau, X^u(\tau))$ is the penalty or cost of this (or reward if g is negative). If X stays in the domain O until the terminal time T, then the cost is $g(T, X^u(T))$. In the following, we also use the conditional cost functional

$$J(t,x,u) = \mathbb{E}\left[\int_t^\tau \ell(t, X^u(t), u(t))\,\mathrm{d}t + g(\tau, X^u(\tau)) \mid X^u(t)\right].$$

The set $\mathcal{U}$ contains all adapted stochastic processes $(u(t))_{t\in[0,T]}$ having all finite moments, i.e., processes which for all $m \geq 1$ satisfy

$$\int_0^T \mathbb{E}[\|u(s)\|^m]\,\mathrm{d}s < \infty.$$

The functions defining the problem are of the form

$$\begin{aligned}
&f\colon [0,T]\times O\times U \to \mathbb{R}^n,\\
&\sigma\colon [0,T]\times O\times U \to \mathbb{R}^{n\times d},\\
&\ell\colon [0,T]\times O\times U \to \mathbb{R},\\
&g\colon [0,T]\times O \to \mathbb{R},
\end{aligned}$$

and we introduce the function

$$a\colon [0,T]\times O\times U \to R^{n\times n},\ (t,x,v)\mapsto \sigma(t,x,v)\sigma^T(t,x,v).$$

In this setting, the Hamiltonian reads

$$\mathcal{H}(t,x,p,q) = \inf_{v\in U}\Big(\langle f(t,x,v), p\rangle + \frac{1}{2}\operatorname{Tr}(a(t,x,v)q) + \ell(t,x,v)\Big), \tag{2.12}$$

and the HJB equation

$$\begin{cases} -\Phi_t(t,x) = \mathcal{H}(t,x,\Phi_x,\Phi_{xx}), & (t,x)\in[0,T)\times O,\\ \Phi(t,x) = g(t,x), & (t,x)\in[0,T)\times\partial O\cup\{T\}\times\bar{O}. \end{cases} \tag{2.13}$$

By $\bar{O}$ we denote the closure of O, i.e., $\bar{O} = O\cup\partial O$. Notice that the HJB equation not only has a terminal value but also an inhomogeneous Dirichlet boundary condition that specifies the value of V on the boundary. We are using the letter Φ for the solution of the HJB equation and V be the value function defined by (2.2). For the purpose of showing that these coincide, so-called verification theorems are used.

We present one verification theorem that is applicable when the HJB equation has a classical solution and there exists a regular enough Markov control policy; see [8, Theorem 3.1]. To formalize the latter, we say that $\pi\colon [0,T]\times\bar{O}\to U$ belongs to $\mathcal{L}$ if π is locally Lipschitz continuous and has linear growth in x, uniformly in t. The space $C_{\mathrm{p}}([0,T],\bar{O})$ is the space of all real-valued continuous functions which satisfy a polynomial growth bound, i.e., all functions h satisfying for some $m\geq 0$ a bound

$$|h(t,x)| \leq C(1 + t^m + \|x\|^m),\quad t\in[0,T],\ x\in\bar{O}.$$

Theorem 2.2 (Verification theorem). *Let* $\Phi \in C^{1,2}([0,T) \times O) \cap C_{\mathrm{p}}([0,T], \bar{O})$ *be a solution to the HJB equation* (2.13). *Then for all* $(t,x) \in [0,T) \times O$ *and* $u \in \mathcal{U}$ *it holds that*

$$\Phi(t,x) \leq J(t,x,u).$$

If in addition there exists a Markov control policy $\pi \in \mathcal{L}$ *satisfying* $\pi(t,x) = \kappa(t,x,\Phi_x(t,x),\Phi_{xx}(t,x))$, *where*

$$\kappa(t,x,p,q) \in \operatorname*{argmin}_{v\in U}\Big(\langle f(t,x,v),p\rangle + \frac{1}{2}\operatorname{Tr}(a(t,x,v)q) + \ell(t,x,v)\Big), \tag{2.14}$$

then

(i) For all $(t,x) \in [0,T] \times O$ *it holds*

$$\Phi(t,x) = J(t,x,u^*),$$

and $\Phi = V$.

(ii) The dynamic programming principle (2.3) *is valid for any stopping time* τ.

Proof. The proof is essentially the derivation of the HJB equation, but the arguments are conducted in the reverse order and with better control of the different steps (since regularity of Φ is known). We start by proving the dynamic programming principle for Φ. By the Itô formula, for $t \in (0,T]$, an arbitrary control $u \in \mathcal{U}$ and a stopping time $\tilde{\tau} > t$, we have

$$\begin{aligned}\Phi(\tilde{\tau}, X^u(\tilde{\tau})) &- \Phi(t, X^u(t)) \\ &= \int_t^{\tilde{\tau}} \big(\Phi_t(s,X^u(s)) + (L^u\Phi)(s,X^u(s))\big)\,\mathrm{d}s \\ &\quad + \int_t^{\tilde{\tau}} \Phi_x(s,x)\sigma(s,x,u)\,\mathrm{d}W(s).\end{aligned} \tag{2.15}$$

The fact that Φ solves the HJB equation implies that

$$-\Phi_t(t,x) = \inf_{v\in U}\big((L^v\Phi)(t,x) + \ell(t,x,v)\big) \leq (L^u\Phi)(t,x) + \ell(t,x,u),$$

and thus the first integrand in (2.15) satisfy

$$\Phi_t(s,X^u(s)) + (L^u\Phi)(s,X^u(s)) \geq -\ell(s,X(s),u).$$

Using this fact, taking the conditional expectation in (2.15), using the martingale property of the stochastic integral and moving terms yields

$$\Phi(t,x) \leq \mathbb{E}\bigg[\int_t^{\tilde{\tau}} \ell(s,X^u(s),u(s))\,\mathrm{d}s + \Phi(\tilde{\tau}, X^u(\tilde{\tau}))\,\bigg|\, X^u(t) = x\bigg].$$

In particular, taking $\tilde{\tau} = \tau$ to be the exit time for the control problem and using the boundary condition $\Phi(t,x) = g(t,x)$ for $(t,x) \in [0,T) \times \partial O \cup \{T\} \times \bar{O}$, we have

$$\Phi(t,x) \leq \mathbb{E}\bigg[\int_t^{\tau} \ell(s,X^u(s),u(s))\,\mathrm{d}s + g(\tau, X^u(\tau))\,\bigg|\, X^u(t) = x\bigg] = J(t,x,u).$$

This proves the first statement of the theorem.

If π is an optimal Markov control policy and $u^*(s) = \pi(s, X^\pi(s))$, then instead it holds

$$\Phi_t(s, X^\pi(s)) + (L^\pi \Phi)(s, X^\pi(s)) = -\ell(s, X^\pi(s), u^*(s)). \tag{2.16}$$

Using this fact and by identical calculations as above we have

$$\Phi(t,x) = \mathbb{E}\left[\int_t^{\tilde{\tau}} \ell(s, X^\pi(s), \pi(s, X^\pi(s)))\,\mathrm{d}s + \Phi(\tilde{\tau}, X^\pi(\tilde{\tau})) \,\Big|\, X^\pi(t) = x \right] \tag{2.17}$$

and

$$\begin{aligned}\Phi(t,x) &= \mathbb{E}\left[\int_t^{\tau} \ell(s, X^\pi(s), u^*(s))\,\mathrm{d}s + g(\tau, X^\pi(\tau)) \,\Big|\, X^\pi(t) = x \right] \\ &= J(t,x,u^*) \\ &= V(t,x).\end{aligned}$$

This proves (i) and by (2.17) we conclude (ii). □

2.4 Linear Quadratic control

Here we consider an important special case that is analytically solvable and allows efficient implementations, also in real-time, for engineering applications. It was implemented, for instance, in the Apollo space shuttles to control the descent through the atmosphere of Earth. The considered setting has a linear unconstrained controlled SDE with $O = \mathbb{R}^n$ and a quadratic cost functional with unconstrained control $U = \mathbb{R}^\ell$, more precisely

$$\mathrm{d}X^u(t) = \big(AX^u(t) + Bu(t)\big)\,\mathrm{d}t + CX^u(t)\,\mathrm{d}W(t), \quad t \in (0,T],$$

and

$$J(u) = \mathbb{E}\left[\int_0^T \big(\|Ru(t)\|^2 + \|QX^u(t)\|^2\big)\,\mathrm{d}t + \|GX^u(T)\|^2 \right].$$

Here, A, B, G, R, Q are matrices in suitable dimensions and $C\colon \mathbb{R}^n \to \mathbb{R}^{n\times d}$ is a linear mapping of the form $(Cx)w = \sum_{\ell=1}^d (C_\ell x) w_\ell$, where $C_1, \ldots, C_d$ are matrices. In other words, the stochastic integral $\int_0^t CX(s)\,\mathrm{d}W(s)$ explicitly reads $\sum_{\ell=1}^d \int_0^t C_\ell X(s)\,\mathrm{d}W_\ell(s)$, where W_ℓ is the ℓ:th coordinate of W. The stated problem is called the *Linear Quadratic* (LQ) control problem. We solve the linear quadratic control problem in a 1-dimensional setting and state the solution to the multi-dimensional setting thereafter.

A 1-dimensional setting

To solve the linear quadratic stochastic control problem for $A, B, C, G, R, Q \in \mathbb{R}$ and $O = U = \mathbb{R}$, we study the corresponding HJB equation. We start by finding the minimum of the Hamiltonian. The function to minimize with respect to $v \in \mathbb{R}$ is

$$v \mapsto f(t,x,v)p + \frac{1}{2}a(t,x,v)q + \ell(t,x,v),$$

with

$$f(t,x,v) = Ax + Bv, \quad a(t,x,v) = C^2x^2, \quad \ell(t,x,v) = R^2v^2 + Q^2x^2.$$

By substitution it reads explicitly

$$v \mapsto (Ax+Bv)p + \frac{1}{2}C^2x^2q + R^2v^2 + Q^2x^2.$$

It is a simple quadratic function in v and it is for every $(t,x) \in [0,T] \times \mathbb{R}$ minimized by

$$v^* = -\frac{B}{2R^2}p.$$

Inserting v^* in the Hamiltonian (2.12), we have after a cancelation of terms

$$\mathcal{H}(t,x,p,q) = Axp - \frac{B^2p^2}{4R^2} + \frac{C^2x^2q}{2} + Q^2x^2.$$

Thus the HJB equation without the infimum reads

$$\begin{cases} -\Phi_t(t,x) = Ax\Phi_x(t,x) - \frac{B^2}{4R^2}\Phi_x^2(t,x) + \frac{C^2}{2}x^2\Phi_{xx}(t,x) + Q^2x^2, \\ \Phi(T,x) = G^2x^2. \end{cases} \tag{2.18}$$

We solve this equation by a quadratic ansatz were we search for a solution on the form

$$\Phi(t,x) = P(t)x^2, \tag{2.19}$$

where $P\colon [0,T] \to \mathbb{R}$ is an unknown differentiable function. By differentiating the proposed solution and inserting into the HJB equation (2.18) we have

$$-\dot{P}(t)x^2 = 2Ax^2P(t) - \frac{B^2}{R^2}x^2P(t)^2 + C^2x^2P(t) + Q^2x^2.$$

Since all terms contain a factor x^2 and since the equation holds for all x it necessarily holds that

$$\begin{cases} -\dot{P}(t) = 2AP(t) - \frac{B^2}{R^2}P(t)^2 + C^2P(t) + Q^2, \\ P(T) = G^2. \end{cases}$$

This is known as a Riccati equation. It has an explicit solution formula that we do not provide here. In multiple dimensions, this is no longer true, but the properties of the solution are known and there are efficient numerical schemes to approximate the solution. Thus, we have a solution Φ in the form (2.19) that satisfies the conditions of Theorem 2.2. From this we conclude that $\Phi = V$ is the value function, and the optimal Markov policy is given by

$$\pi(t,x) = -\frac{B}{2R^2}\Phi_x(t,x) = -\frac{B}{R^2}P(t)x.$$

This solves the linear quadratic problem in one dimension. Thus, the LQ-problem essentially reduces to solving the Riccati equation.

Multi-dimensional setting

In the multi-dimensional setting the Riccati equation is a matrix-valued ordinary differential equation. We state the synthesis of the control problem as a theorem.

Theorem 2.3 (Synthesis of the LQ-problem)**.** *For the linear quadratic control problem, the HJB equation (2.13) is solved by the quadratic form $\Phi(t,x) = x^TP(t)x$, where $P\colon [0,T] \to \mathbb{R}^{n\times n}$ solves the matrix Riccati equation*

$$\begin{cases} -\dot{P}(t) = A^TP(t) + P(t)A + C^TP(t)C + Q^TQ - P(t)B(R^TR)^{-1}B^TP(t), \quad t \in [0,T], \\ P(T) = G^TG, \end{cases}$$

where $C^TP(t)C := \sum_{k=1}^n C_k^TP(t)C_k$. Moreover, the optimal Markov control policy is given by

$$\pi(t,x) = -(R^TR)^{-1}B^TP(t)x.$$

Exercise 2.4. Prove Theorem 2.3.

Exercise 2.5. Here we present an alternative to solving the LQ-problem without the HJB-equation. Use the Itô formula on $\Phi(s, X(s))$ from t to T, where $\Phi(s, x) = x^T P(s)x$, to show that for all $t \in [0, T]$, $x \in \mathbb{R}^n$, $u \in \mathcal{U}$ we have

$$\begin{aligned} J(t, x, u) &= x^T P(t)x \\ &\quad + \mathbb{E}\left[\int_t^T \left\| (R^T R)^{\frac{1}{2}} u(s) + (R^T R)^{-\frac{1}{2}} B^T P(s) X^u(s) \right\|^2 \mathrm{d}s \,\middle|\, X(t) = x \right]. \end{aligned}$$

Use this to derive the Markov control policy π in Theorem 2.3.

Exercise 2.6. Let B and W be two independent Brownian motions and A, B, E be matrices, C, D, F be 3-tensors (like C above) and ξ be a vector in the suitable dimensions. Consider the linear controlled SDE

$$\begin{aligned} \mathrm{d}X^u(t) &= \big(A(X^u(t) - \xi) + Bu(t)\big)\,\mathrm{d}t \\ &\quad + (CX^u(t) + Du(t) + E)\,\mathrm{d}W(t) + Fu(t)\,\mathrm{d}B(t), \quad t \in (0, T], \end{aligned}$$

with the same quadratic cost functional as above. Solve the corresponding HJB-equation. A recommendation is to investigate the effect of introducing ξ, D, E, F one by one and then combining. It is also possible to start with or limit oneself to the 1-dimensional case. First hint: In the quadratic ansatz you might need to add first-order and zeroth-order terms. Second hint: Considering $F \neq 0$, define a Brownian motion $\widetilde{W} = (W^T, B^T)^T$ and formulate the diffusion coefficient in block structure. The answer for adding non-zero ξ and E is given in [1, Section 5.1].

2.5 Controlled Brownian motion and the Cole–Hopf transform

We next proceed with a control problem, the controlled Brownian motion, which admits an even simpler solution than the LQ-problem. It might not be general enough for many applications, but it provides nice reference solutions through Monte Carlo approximations, to other numerical methods, see, e.g., [2, Section 4.1.1]. The dynamics of the controlled Brownian motion is determined by the simple equation

$$\mathrm{d}X^u(t) = bu(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t), \quad t \in (0, T],$$

and the cost functional is given by

$$J(u) = \mathbb{E}\left[\int_0^T r^2 \|u(t)\|^2\,\mathrm{d}t + g(X^u(T)) \right].$$

Here, $b, r, \sigma \in \mathbb{R}$ are scalars and this is required for the given solution approach. On the other hand, the terminal function $g \colon \mathbb{R}^n \to \mathbb{R}$ is quite general and this allows the design of interesting control problems, from the academic point of view. The Hamitonian for the problem reads

$$\mathcal{H}(t, x, p, q) = \inf_{v \in U} \left(b\langle v, p \rangle + \frac{\sigma^2}{2} \mathrm{Tr}(q) + r^2 \|v\|^2 \right).$$

It is quadratic in v and differentiation yields a unique minimum

$$v^* = -\frac{bp}{2r^2}.$$

By substitution we have

$$\mathcal{H}(t,x,p,q) = b\Big\langle -\frac{bp}{2r^2}, p\Big\rangle + \frac{\sigma^2}{2}\mathrm{Tr}(q) + r^2\Big\| -\frac{bp}{2r^2}\Big\|^2 = \frac{\sigma^2}{2}\mathrm{Tr}(q) - \frac{b^2}{4r^2}\|p\|^2.$$

Thus the HJB equation for the control problem reads

$$\begin{cases} \Phi_t = \frac{\sigma^2}{2}\Delta\Phi - \frac{b^2}{4r^2}\|\Phi_x\|^2, & t \in [0,T), \\ \Phi(T,x) = g(x). \end{cases}$$

We transform this non-linear PDE to a linear PDE by means of the famous Cole–Hopf transformation. For some arbitrary constant $c \in \mathbb{R}$ we define the function $\Psi(t,x) = \exp(c\Phi(t,x))$. By the chain and product rules for differentiation we have

$$\begin{aligned} \Psi_t &= c\exp(c\Phi)\Phi_t \\ \Psi_x &= c\exp(c\Phi)\Phi_x \\ \Delta\Psi &= \mathrm{div}(\Psi_x) = c\exp(c\Phi)\Delta\Phi + c^2\exp(c\Phi)\|\Phi_x\|^2. \end{aligned}$$

From the expression of Ψ_t and since Φ_t satisfies the HJB-equation we have

$$\begin{aligned} \Psi_t = c\exp(c\Phi)\Phi_t &= c\exp(c\Phi)\left(\frac{\sigma^2}{2}\Delta\Phi - \frac{b^2}{4r^2}\|\Phi_x\|^2\right) \\ &= \frac{\sigma^2}{2}c\exp(c\Phi)\Delta\Phi - c\exp(c\Phi)\frac{b^2}{4r^2}\|\Phi_x\|^2. \end{aligned}$$

Further, from the expression for $\Delta\Psi$ above, we have

$$c\exp(c\Phi)\Delta\Phi = \Delta\Psi - c^2\exp(c\Phi)\|\Phi_x\|^2,$$

and thus

$$\Psi_t = \frac{\sigma^2}{2}\Delta\Psi - \frac{b^2 + 2\sigma^2 c r^2}{4r^2}c\exp(c\Phi)\|\Phi_x\|^2.$$

The choice

$$c = -\frac{b^2}{2\sigma^2 r^2}$$

yields the linear heat equation

$$\begin{cases} \Psi_t = \frac{\sigma^2}{2}\Delta\Psi, & t \in [0,T), \\ \Psi(T,x) = \exp(-\frac{b^2}{2\sigma^2 r^2}g(x)). \end{cases}$$

The unique solution to this equation is given by $\Psi(t,x) = \mathbb{E}[\Psi(T,\mathcal{X}(T)) \mid \mathcal{X}(t) = x]$, where $\mathrm{d}\mathcal{X}(t) = \sigma\,\mathrm{d}W(t)$. Since $p(\mathcal{X}(T) \mid \mathcal{X}(t) = x) = \mathcal{N}(x, \frac{\sigma^2}{2}(T-t)\mathrm{id})$ we have for $\xi \sim \mathcal{N}(0,\mathrm{id})$ that

$$\Psi(t,x) = \mathbb{E}\left[\exp\left(-\frac{b^2}{2\sigma^2 r^2}g\big(x + \sigma\sqrt{T-t}\,\xi\big)\right)\right]. \tag{2.20}$$

This allows us to conclude the first two statements of the following proposition. We leave the remaining proof to the reader.

Proposition 2.7 (Synthesis of the controlled Brownian motion)**.** *For the controlled Brownian motion, the HJB equation* (2.13) *is solved by*

$$\Phi(t,x) = -\frac{2\sigma^2 r^2}{b^2}\log(\Psi(t,x)),$$

where the function Ψ *is given by* (2.20)*. The optimal Markov control policy is given by*

$$\pi(t,x) = -\frac{b}{2r^2}\Phi_x(t,x),$$

where

$$\Phi_x(t,x) = -\frac{b^2}{2\sigma^2 r^2}\Psi(t,x)\Psi_x(t,x),$$

and

$$\Psi_x(t,x) = -\frac{b^2}{2\sigma^2 r^2}\mathbb{E}\left[\exp\left(-\frac{b^2}{2\sigma^2 r^2}g\big(x+\sigma\sqrt{T-t}\,\xi\big)\right)\nabla g\big(x+\sigma\sqrt{T-t}\,\xi\big)\right]$$

for $\xi \sim \mathcal{N}(0,1)$*.*

2.6 Non-degenerate control

In general the HJB equation (2.13) is difficult to study and requires the advanced concept of viscosity solutions. Also, due to discontinuous Markov control maps, the strong solution concept of the SDE is not sufficient, but weak solutions need to be considered. Fortunately, there is a condition under which classical PDE solutions and strong SDE solutions are guaranteed. This is when there is enough noise at all times in all space-dimensions that makes the solution regular. The intuition behind this is that the diffusion term introduces a local averaging effect of the solution which makes it smoother. This is formalized in terms of the following uniform ellipticity condition: There exists $C > 0$ such that for all $t \in [0,T]$, $x \in O$, $v \in U$, $\xi \in \mathbb{R}^n$ it holds that

$$\sum_{i,j=1}^{N} a_{ij}(t,x,v)\xi_i\xi_j \geq C\|\xi\|^2. \tag{2.21}$$

It is equivalent to the smallest eigenvalue of the positive matrix $a(t,x,v)$ being uniformly bounded from below by C. It implies that the noise at every time and every point in space is strong enough so that the diffusion regularizes the solution. Under the same condition, the density of the solution to the SDE exists. The next two theorems from [8, Section IV.4] are pure PDE results and their proofs do not belong to a course in stochastic calculus. They provide two sets of conditions for which the regularity of Φ satisfies that of Theorem 2.2. The first concern a bounded domain O with a smooth boundary ∂O, and the second concern an unbounded domain O. The latter require boundedness of all involved functions and bounded derivatives of the terminal cost function.

Theorem 2.8. *Let the following conditions hold:*

- *U is compact.*
- *O is bounded with ∂O being a manifold of class C^3.*
- *a, f, ℓ have one continuous time derivative and two continuous space derivatives.*
- *g has three continuous derivatives in both space and time.*
- *σ is uniformly elliptic, i.e., (2.21) holds.*

Then (2.13) has a unique solution $V \in C^{1,2}([0,T) \times O) \cap C([0,T], \bar{O})$.

Theorem 2.9. *Let the following conditions hold:*

- *U is compact.*
- *$O = \mathbb{R}^n$.*
- *a, f, ℓ are bounded and have one continuous time derivative and two continuous space derivatives.*
- *g has three bounded and continuous derivatives in space (since $O = \mathbb{R}^n$ the equation is not stopped and g does not depend on time).*
- *σ is uniformly elliptic, i.e., (2.21) holds.*

Then (2.13) has a unique solution $\Phi \in C_{\mathrm{b}}^{1,2}([0,T] \times \bar{O})$.

2.7 A stochastic representation of the HJB equation

While the HJB equation solves the control problem, it is only in very rare cases, such as the linear quadratic case, that analytical solutions are tractable. Also, classical numerical schemes do not scale well to higher dimensions, reducing its applicability. We now present a stochastic representation for the solution Φ of the HJB equation. Defining numerical schemes based on this representation, rather than the PDE representation, has shown to scale well to high dimensions, in particular schemes utilizing neural networks. We restrict ourselves to the case without control in the diffusion, i.e., with σ not depending on the control, and also to nonconstrained control $O = \mathbb{R}^n$ and $U = \mathbb{R}^k$.

Let the conditions of Theorem 2.2 hold, i.e., let $\Phi \in C^{1,2}([0,T] \times \mathbb{R}^n)$ be the solution to (2.13) and let $\pi \in \mathcal{L}$ and κ satisfy (2.14), i.e., be the minimizer of the Hamiltonian. Further let X be the solution to

$$\mathrm{d}X(t) = f(t, X(t), \kappa(t, X(t), \Phi_x(t, X(t))))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t), \quad t \in [0,T]. \tag{2.22}$$

It is the optimally controlled SDE. By the Itô formula we have

$$\begin{aligned}\Phi(t, X(t)) = \Phi(0, X(0)) + \int_0^t \big(\Phi_t(s, X(s)) + (L^\pi \Phi)(s, X(s))\big)\,\mathrm{d}s \\ + \int_0^t \Phi_x(s, X(s))\sigma(s, X(s))\,\mathrm{d}W(s).\end{aligned} \tag{2.23}$$

Since $\pi(t,x)$ attains the infimum in the Hamiltonian, the HJB equation reads

$$-\Phi_t(t,x) = (L^\pi \Phi)(t,x) + \ell(t, x, \pi(t,x)),$$

or

$$\Phi_t(t,x) + (L^\pi \Phi)(t,x) = -\ell(t,x,\pi(t,x)).$$

Using this in (2.23) yields

$$\begin{aligned}\Phi(t,X(t)) = \Phi(0,X(0)) &- \int_0^t \ell(s,X(s),\kappa(s,X(s),\Phi_x(s,X(s))))\,\mathrm{d}s \\ &+ \int_0^t \Phi_x(s,X(s))\sigma(s,X(s))\,\mathrm{d}W(s).\end{aligned} \tag{2.24}$$

It can be shown that Φ is the only function that satisfies (2.22), (2.24) and

$$\Phi(T,x) = g(x). \tag{2.25}$$

Thus, finding a function Φ that satisfies the HJB equation is equivalent to finding a function satisfying (2.22), (2.24), (2.25). This means that we have a second representation of the same problem, not involving a PDE. The relation is called the *nonlinear Feynmann-Kac formula.*

We will explain how to formulate (2.22), (2.24), (2.25) as an optimization problem suitable for numerical solvers. For this purpose, we introduce the stochastic processes $Y(t) = \Phi(t,X(t))$ and $Z(t) = \Phi_x(t,X(t))$ and notation

$$\begin{aligned}\tilde{f}(t,x,z) &= f(t,x,\kappa(t,x,z)),\\ \tilde{\ell}(t,x,z) &= \ell(t,x,\kappa(t,x,z)).\end{aligned} \tag{2.26}$$

With this at hand, we can write (2.22), (2.24), (2.25) as

$$\begin{cases} X(t) = X(0) + \int_0^t \tilde{f}(s,X(s),Z(s))\,\mathrm{d}s + \int_0^t \sigma(s,X(s))\,\mathrm{d}W(s),\\ Y(t) = Y(0) - \int_0^t \tilde{\ell}(s,X(s),Z(s))\,\mathrm{d}s + \int_0^t Z(s)\sigma(s,X(s))\,\mathrm{d}W(s),\\ Y(0) = \Phi(0,X(0)),\\ Y(T) = g(X(T)).\end{cases} \tag{2.27}$$

We remind the reader that Φ, its gradient, and initial value are unknown. In fact, the initial value $Y(0)$ is unknown while the terminal value $Y(T)$ is known although random, making (2.27) significantly different from the forward SDEs considered earlier in the course. It is an example of a so-called *Forward Backward Stochastic Differential Equation* (FBSDE) with a solution triple (X,Y,Z). The drift coefficient of the backward equation is called *driver.* We assume a setting where we know that $\Phi = V$ and thus Φ solves the optimal control problem. This means in particular that

$$Y(0) = V(0,X(0)) = J(0,X(0),u^*).$$

Here u^* is the optimal control defined by π and thus by κ (which is known) and V_x (which is unknown). An approximation of (2.27), replacing V_x by a possibly suboptimal and sufficiently regular control $\zeta\colon [0,T]\times\mathbb{R}^n \to \mathbb{R}^d$ gives

$$\begin{aligned} X^\zeta(t) &= X^\zeta(0) + \int_0^t \tilde{f}(s,X^\zeta(s),Z^\zeta(s))\,\mathrm{d}s + \int_0^t \sigma(s,X^\zeta(s))\,\mathrm{d}W(s),\\ Y^\zeta(t) &= Y^\zeta(0) - \int_0^t \tilde{\ell}(s,X^\zeta(s),Z^\zeta(s))\,\mathrm{d}s + \int_0^t \zeta(s,X^\zeta(s))\sigma(s,X^\zeta(s))\,\mathrm{d}W(s)\\ Y^\zeta(0) &= \mathbb{E}\left[\int_0^T \tilde{\ell}(t,X^\zeta(s),Z^\zeta(s))\,\mathrm{d}s + g(X^\zeta(T))\right]\\ Z^\zeta(t) &= \zeta(t,X^\zeta(t)).\end{aligned} \tag{2.28}$$

Given some fixed ζ, the equation for X^ζ does not depend on Y^ζ. Therefore, the initial value $Y^\zeta(0)$ is known and $Y^\zeta(t)$, $t \in (0,T]$ is integrated independently. We want two properties of the true solution $\zeta = V_x$. The first is that it should minimize $Y^\zeta(0)$, i.e., the cost functional. The second is that the terminal condition $Y^\zeta(T) = g(X^\zeta(T))$ should be satisfied. This proposes, for $\alpha \in [0,1]$, the following equivalent optimization problem

$$\underset{\zeta}{\text{minimize}} \quad \mathcal{J}(\zeta) = \alpha Y^\zeta(0) + (1-\alpha)\, \mathbb{E}\big[|Y^\zeta(T) - g(X^\zeta(T))|^2\big] \quad \text{subject to (2.28).} \tag{2.29}$$

For $\alpha = 0$ the problem is uniquely solvable and thus the first term is not necessary. It is included since it gives a convergent numerical scheme after discretization of (2.28) with the Euler–Maruyama method, while without it convergence fails in general; see [1]. Deep learning approximations of the function ζ, or rather functions $\zeta_0, \ldots, \zeta_N$ at each time step of an Euler–Maruyama method, have proved to work very well. This is the famous deep BSDE method introduced in [6].

Exercise 2.10 (Non-uniqueness of stochastic representation)**.** For $\psi\colon \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$, let FBSDE($\psi$) be the FBSDE (2.27) but with modified drift $f_\psi = \tilde{f} + \psi$ and driver $\ell_\psi = \tilde{\ell} - \langle z, \psi\rangle$. Prove that for any sufficiently regular ψ, FBSDE(ψ) is equivalent to the same HJB-equation and thus the same stochastic control problem.

An alternative optimization problem to (2.29), based on Exercise 2.10, that also converges after discretization with the Euler-Maruyama method was proposed in [2]. It relies on averaging the errors in the terminal values of a small number $K \geq 2$ of equivalent FBSDE, perturbed by $\psi_1, \ldots, \psi_K$ in the sense of Exercise 2.10. Thus, it reads

$$\underset{\zeta}{\text{minimize}} \quad \mathcal{J}(\zeta) = \sum_{k=1}^{K} \mathbb{E}\big[|Y^{\zeta,k}(T) - g(X^{\zeta,k}(T))|^2\big] \quad \text{subject to}$$

$$\begin{aligned}
X^{\zeta,k}(t) &= X^{\zeta,k}(0) + \int_0^t f_{\psi_k}(s, X^{\zeta,k}(s), Z^{\zeta,k}(s))\,\mathrm{d}s + \int_0^t \sigma(s, X^{\zeta,k}(s))\,\mathrm{d}W(s),\\
Y^{\zeta,k}(t) &= Y^{\zeta,k}(0) - \int_0^t \ell_{\psi_k}(s, X^{\zeta,k}(s), Z^{\zeta,k}(s))\,\mathrm{d}s + \int_0^t \zeta(s, X^{\zeta,k}a(s))\sigma(s, X^{\zeta,k}(s))\,\mathrm{d}W(s)\\
Z^{\zeta,k}(t) &= \zeta(t, X^{\zeta,k}(t)).
\end{aligned} \tag{2.30}$$

For $k=1$ and $\psi_1 = 0$ this optimization problem is equivalent to (2.29) with $\alpha = 0$, which again does not converge after discretization. Taking $K = 1$ and $\psi_1 \neq 0$ does not help, but for $k \geq 2$ it is shown experimentally that convergence is obtained and that the choice of $\psi_1, \ldots, \psi_K$ is not so sensitive [2].

Both optimization problems (2.29) and (2.30) are set up to find the gradient V_x, or an approximation of it, but not V. Given the gradient, we have the optimal control u^* through κ, or again an approximation of it. If the value function V is desired, it can be computed using

$$V(t,x) = \mathbb{E}\left[\int_t^T \ell(t, X(t), u^*(t))\,\mathrm{d}t + g(T, X(T))\right],$$

exactly in theory or approximately using a Monte-Carlo method in practice.

In the above, we have assumed that σ does not depend on the control. This is not necessary, but the HJB equation will instead have a representation in terms of a second-order FBSDE (2FBSDE), which also contains an equation for the Hessian $\Gamma(t) = V_{xx}(t, X(t))$. We refrain from further details, but we remark that there are successful numerical schemes using 2FBSDE and deep learning, [4].

2.8 A digression into solution theory for SDE and BSDE

Above, the process Z was defined as $Z(t) = V_x(t, X(t))$. This is sometimes seen in the literature, but it is more common to define it as $Z(t) = V_x(t, X(t))\sigma(t, X(t))$. With this choice, the functions $\tilde{f}$ and

$\tilde{\ell}$ need to be defined differently than in (2.26). Recall that the optimal control u^* is given by

$$u^*(t) = \pi(t, X(t)) = \kappa(t, x, V_x(t, X(t)))).$$

To make it a function of Z, we first see that

$$\begin{aligned} V_x(t, X(t)) &= V_x(t, X(t))\sigma(t, X(t))\sigma^T(t, X(t))\Big(\sigma(t, X(t))\sigma^T(t, X(t))\Big)^{-1} \\ &= Z(t)\sigma^T(t, X(t))\Big(\sigma(t, X(t))\sigma^T(t, X(t))\Big)^{-1}. \end{aligned}$$

Thus, with

$$\begin{aligned} \tilde{f}(t, x, z) &= f(t, x, \kappa(t, x, z\sigma^T(t, x)\big(\sigma(t, x)\sigma^T(t, x)\big)^{-1})), \\ \tilde{\ell}(t, x, z) &= \ell(t, x, \kappa(t, x, z\sigma^T(t, x)\big(\sigma(t, x)\sigma^T(t, x)\big)^{-1})) \end{aligned}$$

we have

$$\begin{aligned} X(t) &= X(0) + \int_0^t \tilde{f}(s, X(s), Z(s))\,\mathrm{d}s + \int_0^t \sigma(s, X(s))\,\mathrm{d}W(s), \\ Y(t) &= Y(0) - \int_0^t \tilde{\ell}(s, X(s), Z(s))\,\mathrm{d}s + \int_0^t Z(s)\,\mathrm{d}W(s), \\ Y(0) &= \Phi(0, X(0)), \\ Y(T) &= g(X(T)). \end{aligned}$$

Of course, this requires that the inverse is defined, but this is valid in the non-degenerate case assuming (2.21). Integrating Y backward in time instead gives

$$\begin{aligned} X(t) &= X(0) + \int_0^t \tilde{f}(s, X(s), Z(s))\,\mathrm{d}s + \int_0^t \sigma(s, X(s))\,\mathrm{d}W(s), \\ Y(t) &= g(X(T)) + \int_t^T \tilde{\ell}(s, X(s), Z(s))\,\mathrm{d}s - \int_t^T Z(s)\,\mathrm{d}W(s). \end{aligned} \tag{2.31}$$

This is a typical way of writing an FBSDE. For the specific problem $\tilde{f}$ and $\tilde{\ell}$ do not depend on Y, but in general they do.

To give the interested reader a taste of BSDE, we first discuss solutions to the forward SDE.

$$X(t) = x + \int_0^t f(s, X(s))\,\mathrm{d}s + \int_0^t \sigma(s, X(s))\,\mathrm{d}W(s). \tag{2.32}$$

Introducing the Banach space $\mathcal{S}_T^2$ of adapted (or rather predictable) stochastic processes equipped and defined by the norm

$$\|x\|_{\mathcal{S}_T^2} = \mathbb{E}\left[\sup_{t\in[0,T]} \|x(t)\|_{\mathbb{R}^n}^2\right],$$

the SDE can be written as a fixed point equation

$$X = \Gamma(X), \tag{2.33}$$

with $\Gamma : \mathcal{S}_T^2 \to \mathcal{S}_T^2$ given by

$$\Gamma(x) = \left(x(0) + \int_0^t f(s, x(s))\,\mathrm{d}s + \int_0^t \sigma(s, x(s))\,\mathrm{d}W(s)\right)_{t\in[0,T]}.$$

The existence and uniqueness of a solution $X \in \mathcal{S}_T^2$ to (2.33) are proved by the Banach's Fixed Point Theorem, or equivalently by Picard iterations. The latter is the process of defining a sequence $(X_n)_{n\in\mathbb{N}}$ by $X_n = \Gamma(X_{n-1})$, with $X_0 \equiv 0$ and $X = \lim_{n\to\infty} X_n$. The proof of Banach's Fixed Point Theorem is based on showing the convergence of this sequence, and using the theorem thus removes one step of the proof. Nevertheless, it is a matter of taste which proof to choose. Some communities prefer functional analytical arguments, while others including many probabilists do not. We state Banach's Fixed Point Theorem.

Theorem 2.11. *(Banach's Fixed Point Theorem) Let $(E, \|\cdot\|_E)$ be a Banach space. If $\Gamma\colon E \to E$ is a contraction, i.e., if there exists $\alpha \in (0,1)$ so that for all $x_1, x_2 \in E$ it holds*

$$\|\Gamma(x_1) - \Gamma(x_2)\|_E \le \alpha \|x_1 - x_2\|_E,$$

then there exists a unique fixed point $X \in E$ to Γ, i.e., an $X \in E$ satisfying $\Gamma(X) = X$.

Exercise 2.12. Show under Lipschitz and linear growth conditions on f and σ that for small enough $T > 0$ the conditions of Banach's Fixed Point Theorem are satisfied for Γ. In addition to showing the contraction property, it should be shown that for all $x \in \mathcal{S}_T^2$ we have $\Gamma(x) \in \mathcal{S}_T^2$.

Exercise 2.13. Show that SDE (2.32) has a unique solution for any $T > 0$ by iterative use of the result of Exercise 2.12, i.e., that SDE (2.32) has a unique solution.

Exercise 2.14. Introduce a family of equivalent norms on $\mathcal{S}_T^2$, defined by

$$\|x\|_\rho = \mathbb{E}\left[\sup_{t\in[0,T]} e^{\rho t}\|x(t)\|_{\mathbb{R}^n}^2\right], \qquad \rho \in \mathbb{R}.$$

Show under Lipschitz and linear growth conditions on f and σ that for any $T > 0$ and some ρ the conditions of Banach's Fixed Point Theorem are satisfied for Γ.

Inspired by considering the SDE as a fixed point problem, we now turn to how to set up a fixed point problem for a BSDE. For this purpose, we consider the equation

$$Y(t) = \xi + \int_t^T \ell(s, Y(s), Z(s))\,\mathrm{d}s - \int_t^T Z(s)\,\mathrm{d}W(s), \tag{2.34}$$

where ξ is $\mathcal{F}_T$-measurable and $(\ell(t,y,z))_{t\in[0,T]}$ is adapted for any fixed and deterministic y and z. This is a more general case than the FBSDE (2.31), where the randomness of f and ξ might or might not come from a forward SDE. Naively setting up a candidate fixed-point map for Y based on the right-hand side of (2.34), as we did for the SDE for X, fails for two reasons. First, the resulting process is not adapted in general and thus the map is not well defined $\mathcal{S}_T^2 \to \mathcal{S}_T^2$. Second, by such an approach we have completely ignored the process Z. The remedy is to use the martingale representation theorem.

We introduce the Hilbert space $\mathcal{H}_T^2$ of adapted (or rather predictable) stochastic processes, equipped with and defined by the norm

$$\|x\|_{\mathcal{H}_T^2} = \mathbb{E}\left[\int_0^T \|x(t)\|_{\mathbb{R}^n}^2\,\mathrm{d}t\right].$$

Theorem 2.15 (Martingale Representation Theorem). *Let $(M(t))_{t\geq 0} \in \mathcal{S}_T^2$ be a martingale. Then there exists a unique process $Z \in \mathcal{H}_T^2$ so that*

$$M(t) = \mathbb{E}[M(0)] + \int_0^t Z(s)\,\mathrm{d}W(s), \quad t \geq 0.$$

To arrive at a suitable fixed point map, we start by defining a mapping

$$\mathcal{S}_T^2 \times \mathcal{H}_T^2 \to \mathcal{S}_T^2,\ (y,z) \mapsto M^{y,z},$$

given by the square integrable martingale

$$M^{y,z}(t) = \mathbb{E}\left[\xi + \int_0^T \ell(s, y(s), z(s))\,\mathrm{d}s \,\middle|\, \mathcal{F}_t\right], \quad t \in [0,T]. \tag{2.35}$$

By the martingale representation theorem, there exists a unique adapted stochastic process $Z^{y,z}(t) \in \mathcal{H}_T^2$ so that

$$M^{y,z}(t) = M^{y,z}(0) + \int_0^t Z^{y,z}(s)\,\mathrm{d}W(s). \tag{2.36}$$

We define $Y^{y,z} \in \mathcal{S}_T^2$ given by

$$Y^{y,z}(t) = M^{y,z}(t) - \int_0^t \ell(s, y(s), z(s))\,\mathrm{d}s, \tag{2.37}$$

which by (2.35) is equivalent to

$$Y^{y,z}(t) = \mathbb{E}\left[\xi + \int_t^T \ell(s, y(s), z(s))\,\mathrm{d}s \,\middle|\, \mathcal{F}_t\right]. \tag{2.38}$$

Combining (2.36) with (2.37) and noticing that $Y^{y,z}(0) = M^{y,z}(0)$ we have

$$Y^{y,z}(t) = Y^{y,z}(0) - \int_0^t \ell(s, y(s), z(s))\,\mathrm{d}s + \int_0^t Z^{y,z}(s)\,\mathrm{d}W(s), \tag{2.39}$$

which is of the form of BSDE written as a forward equation. From (2.38) we have $Y^{y,z}(T) = \xi$, and integrating (2.39) backward in time we get

$$Y^{y,z}(t) = \xi + \int_t^T \ell(s, y(s), z(s))\,\mathrm{d}s - \int_t^T Z^{y,z}(s)\,\mathrm{d}W(s),$$

which is of the form of a BSDE written backward in time. From this it seems sensible to introduce the fixed point map Γ given by

$$\Gamma\colon \mathcal{S}_T^2 \times \mathcal{H}_T^2 \to \mathcal{S}_T^2 \times \mathcal{H}_T^2,\ (y,z) \mapsto (Y^{y,z}, Z^{y,z}).$$

The norm on the product space $\mathcal{S}_T^2 \times \mathcal{H}_T^2$ is the sum of the two norms. The solution to the BSDE is the unique pair of stochastic processes $(Y,Z) \in \mathcal{S}_T^2 \times \mathcal{H}_T^2$ satisfying

$$(Y,Z) = \Gamma(Y,Z).$$

The above sketch shows that BSDE are quite different to forward SDE. In particular, they have two solution components (Y,Z) and the martingale representation theorem is central in understanding the solution. As we saw above, when the BSDE is coupled to a forward SDE with solution X, in a Markovian way, then (Y,Z) are functions of X and these are determined by the solution Φ to a PDE. For more about BSDE, see [7].

2.9 A discounted optimal control problem

Here we extend the control problem slightly to account for a cost functional with so-called discount, see [10] for more details. It is theoretically interesting due to its natural connection to the Feynmann–Kac formula, it is used for the infinite horizon case in the next section and finally, in mathematical finance most problems are discounted and fit into this framework. The HJB equation for $\ell = 0$, f, σ not depending on the control and $O = \mathbb{R}^n$, reduces to the Kolmogorov backward equation

$$\begin{cases} \Phi_t = L\Phi, \\ \Phi(T, x) = g(x), \end{cases}$$

and we know that it is solved by $\Phi(t, x) = \mathbb{E}[g(X(T))|X(t) = x]$. Recall the Fenman–Kac formula, adding a term $\gamma\Phi(t, x)$ with $\gamma \in \mathbb{R}$ to the equation, i.e., considering

$$\begin{cases} \Phi_t(t, x) + \gamma\Phi(t, x) = (L\Phi)(t, x), \\ \Phi(T, x) = g(x), \end{cases}$$

gives the stochastic representation $\Phi(t, x) = \mathbb{E}[e^{-\gamma(T-t)}g(X(T))|X(t) = x]$. The reader familiar with mathematical finance recognizes this as the expected discounted return of a financial derivative with payoff function g. The extension in this section is similar and is based on adding a term $-\gamma\Phi(t, x)$ inside the Hamiltonian of the HJB equation and we see what control problem it solves. It turns out to be an important extension. We consider a general open domain O and thus a potentially stopped control problem.

We redefine

$$\mathcal{H}(t, x, o, p, q) = \inf_{v\in U} \Big(-\gamma o + \langle f(t, x, v), p\rangle + \frac{1}{2}\operatorname{Tr}(a(t, x, v)q) + \ell(t, x, v)\Big), \tag{2.40}$$

and consider the HJB equation

$$\begin{cases} -\Phi_t(t, x) = \mathcal{H}(t, x, \Phi, \Phi_x, \Phi_{xx}), & (t, x) \in [0, T) \times O, \\ \Phi(t, x) = g(t, x), & (t, x) \in [0, T) \times \partial O \cup \{T\} \times \bar{O}. \end{cases} \tag{2.41}$$

It turns out that it solves the control problem with cost functional

$$J(t, x, u) = \mathbb{E}\left[\int_t^\tau e^{-\gamma(s-t)}\ell(s, X^u(s), u(s))\,\mathrm{d}s + e^{-\gamma(T-t)}g(\tau, X^u(\tau)) \,\middle|\, X^u(t) = x \right]. \tag{2.42}$$

Exercise 2.16. Prove a modification of the verification Theorem 2.2 with the setting of the current subsection. Follow the steps of the proof of Theorem 2.2 but apply the Itô formula on $\Psi(s, X(s))$, where $\Psi(s, x) = e^{-\gamma(s-t)}\Phi(s, x)$.

Exercise 2.17. Prove a stochastic (FBSDE) representation of the discounted control problem similar to (2.27).

3 Infinite horizon stochastic control

So far in these notes, we have considered control on a finite time horizon $T > 0$ with a running cost and a terminal cost. If there is a clearly defined terminal time and transient behavior of the system, then this is the correct control problem. On the other hand, if we take the example of a Segway, i.e., an inverted pendulum being controlled around its unstable upward-pointing equilibrium point, then there is no natural terminal time, but the purpose of the controller is to control the Segway to remain standing. For this purpose infinite-horizon control of autonomous (not time-dependent coefficients) SDE is the suitable choice.

Here we consider the setting of [8, Section IV.5], an autonomous controlled SDE

$$\mathrm{d}X^u(t) = f(X^u(t), u(t))\,\mathrm{d}t + \sigma(X^u(t), u(t))\,\mathrm{d}W(t), \quad t > 0.$$

We let τ^u be the stopping time when X^u hits the boundary ∂O of the open domain O. The infinite horizon control problem with discount $\gamma \geq 0$ has the cost functional

$$J(x,u) = \mathbb{E}\left[\int_0^\tau e^{-\gamma s} \ell(X^u(s), u(s))\,\mathrm{d}s + e^{-\gamma\tau} g(X^u(\tau)) \,\Big|\, X^u(0) = x \right]. \tag{3.1}$$

The running and terminal costs do not depend on time and we remark that neither does the cost functional. The value function is given by

$$V(x) = \inf_{u \in \mathcal{U}} J(x,u), \quad x \in \bar{O}.$$

We notice the similarity between the discounted cost functional (2.42) and (3.1). It might therefore not come as a surprise that the Hamiltonians are the same, i.e. $\mathcal{H}$ is defined by (2.40). The corresponding HJB equation, considering that it is time independent and thus $\Phi_t \equiv 0$, is the elliptic PDE

$$\begin{cases} 0 = \mathcal{H}(x, \Phi, \Phi_x, \Phi_{xx}), & x \in O, \\ \Phi(x) = g(x), & x \in \partial O. \end{cases} \tag{3.2}$$

We next state a verification theorem.

Theorem 3.1 (Verification theorem). *Let $\Phi \in C^2(O) \cap C_{\mathrm{p}}(\bar{O})$ be a solution to the HJB equation (3.2). Then for all $x \in O$ and $u \in \mathcal{U}$ satisfying*

$$\lim_{t_1 \to \infty} e^{-\gamma t_1} \inf_{u \in \mathcal{U}} \mathbb{E}[\mathbf{1}_{\tau \geq t_1} \Phi(X^u(t_1))] \leq 0,$$

it holds that

$$\Phi(x) \leq J(x,u).$$

If there in addition exists a Markov control policy $\pi \in \mathcal{L}$ satisfying $\pi(x) = \kappa(x, \Phi_x(x), \Phi_{xx}(x))$, where

$$\kappa(x,p,q) \in \operatorname*{argmin}_{v \in U} \Big(\langle f(x,v), p \rangle + \frac{1}{2} \mathrm{Tr}(v(x,a)q) + \ell(x,v) \Big), \tag{3.3}$$

and if

$$\lim_{t_1 \to \infty} e^{-\gamma t_1} \inf_{u \in \mathcal{U}} \mathbb{E}[\mathbf{1}_{\tau \geq t_1} \Phi(X^\pi(t_1))] = 0,$$

then for all $x \in O$ it holds

$$\Phi(x) = J(x, u^*),$$

and $\Phi = V$.

Proof. Define $\Psi(t,x) = e^{-\beta t}\Phi(x)$, let $t_1 > 0$ and $X^u(0) = x$. Applying the Itô formula to $\Psi(t, X^u(t))$ from 0 to $\tau \wedge t_1$ yields

$$
\begin{aligned}
&e^{-\gamma(\tau\wedge t_1)}\Phi(X^u(\tau\wedge t_1)) - \Phi(x)\\
&\quad= \int_0^{\tau\wedge t_1} e^{-\gamma s}\big(-\gamma\Phi(X^u(s)) + (L^u\Phi)(X(s))\big)\,\mathrm{d}s\\
&\qquad+ \int_0^{\tau\wedge t_1} \Phi_x(X^u(s))\sigma(X^u(s), u(s))\,\mathrm{d}W(s).
\end{aligned}
$$

Using the fact that $\Phi(X^u(\tau)) = g(X^u(\tau))$ and rearranging terms, we have

$$
\begin{aligned}
\Phi(x) &= e^{-\gamma t_1}\mathbf{1}_{\tau\geq t_1}\Phi(X^u(t_1)) + e^{-\gamma\tau}\mathbf{1}_{\tau<t_1}g(X^u(\tau))\\
&\quad+ \int_0^{\tau\wedge t_1} e^{-\gamma s}\big(\gamma\Phi(X^u(s)) - (L^u\Phi)(X(s))\big)\,\mathrm{d}s\\
&\quad- \int_0^{\tau\wedge t_1} e^{-\gamma s}\Phi_x(X^u(s))\sigma(s, X^u(s))\,\mathrm{d}W(s).
\end{aligned}
$$

Similarly to the proof of Theorem 2.2, by taking expectations, using suboptimality of u and the fact that Φ satisfies the HJB equation, we have

$$
\begin{aligned}
\Phi(x) &\leq \mathbb{E}\Big[e^{-\gamma t_1}\mathbf{1}_{\tau\geq t_1}\Phi(X^u(t_1))\Big]\\
&\quad+ \mathbb{E}\left[\int_0^{\tau\wedge t_1} e^{-\gamma s}\ell(X^u(s), u(s))\,\mathrm{d}s + e^{-\gamma\tau}\mathbf{1}_{\tau<t_1}g(X^u(\tau))\right].
\end{aligned}
$$

In the limit as $t_1 \to \infty$, by the assumption on the first term, we have

$$
\Phi(x) \leq \mathbb{E}\left[\int_0^{\tau} e^{-\gamma s}\ell(X^u(s), u(s))\,\mathrm{d}s + e^{-\gamma\tau}g(X^u(\tau))\right] = J(x,u).
$$

For optimal u^* and by assumption it holds with equality, i.e.,

$$
\Phi(x) = \mathbb{E}\left[\int_0^{\tau} e^{-\gamma s}\ell(X(s), u^*(s))\,\mathrm{d}s + e^{-\gamma\tau}g(X(\tau))\right] = J(x,u^*).
$$

This completes the proof. □

Exercise 3.2. Consider the infinite horizon linear quadratic control problem

$$
\mathrm{d}X^u(t) = \big(AX^u(t) + Bu(t)\big)\,\mathrm{d}t + CX^u(t)\,\mathrm{d}W(t), \quad t > 0,
$$

with $O = \mathbb{R}^n$ and cost functional

$$
J(x,u) = \mathbb{E}\left[\int_0^{\infty} e^{-\gamma t}\big(\|Ru(t)\|^2 + \|QX^u(t)\|^2\big)\,\mathrm{d}t \,\Big|\, X^u(0) = x\right]. \tag{3.4}
$$

Show the following:

(i) The HJB-equation for this problem is solved by $\Phi(x) = x^T Px$, were $P \in \mathbb{R}^{n\times n}$ solves the stationary Riccati equation

$$
\gamma P = A^T P + PA + C^T PC + Q^T Q - PB(R^T R)^{-1}B^T P,
$$

where $C^T PC := \sum_{k=1}^n C_k^T PC_k$.

(ii) The optimal Markov control policy is $\pi(x) = -B^T(R^T R)^{-1}Px$.

4 Filtering with discrete observations

Here we turn to the filtering problem, which, compared to the control problem, is a statistical problem. From the control perspective, filtering is highly relevant as full knowledge of the system to be controlled is seldom available but only provided through noisy sensors at discrete times. Filtering allows estimation of the hidden or latent system state in terms of posteriori distributions, in the Bayesian sense. Filtering is, however, also highly relevant for various applications outside optimal control, such as parameter estimation and diagnostics in medicine.

The filtering problem is introduced in Section 4.1. In the SDE setting of these notes, the filtering problem is solved exactly by repeated uses of the Bayes formula and prediction with the Fokker-Planck equation or the Forward Kolmogorov equation; see Section 4.2. Having the exact filter in theory does not for most problems imply a practically implementable filter. One reason is that the Fokker-Planck equation is nontrivial to solve, and numerical solutions with classical methods do not scale in state dimension. The second reason is that the nominator in the use of the Bayes formula is hard to evaluate. In Section 4.3 we present a nonnormalized exact filter where the nominator of the Bayes formula is neglected. The resulting filter is exact at every time after normailization. In more general Bayesian filtering, not necessarily stemming from SDE, the so-called filter equations are central. They are briefly introduced and discussed in Section 4.4. Finally, in Section 4.5 we derive the very important Kalman filter. It is the exact filter for the Ornstein-Uhlenbeck process under linear observations with Gaussian additive measurement noise.

4.1 The filtering problem

In order to formalize the filtering problem, consider $\mathbb{R}^j$-dimensional sensor measurements or data $(\widetilde{Y}_k)_{k=0}^K$, collected at times $(t_k)_{k=0}^K$ with $t_0 = 0$ and $t_k < t_{k+1}$, $k \in \{0, 1, \dots, K\}$. In filtering, measurements are modeled as noisy non-linear transformations of the solution to an SDE with a $\mathbb{R}^n$-dimensional solution $(X(t))_{t\geq 0}$. More precisely, data $(\widetilde{Y}_k)_{k=0}^K$ are assumed to be equal in distribution to $(Y_k)_{k=0}^K$, or close enough in practice since no model is perfect, where

$$Y_k = h(X(t_k)) + r_k, \quad k = 0, 1, 2, \dots, K. \tag{4.1}$$

Here, $(r_k)_{k=0}^\infty$ are independent identically distributed centered Gaussian random variables with covariance R and $h\colon \mathbb{R}^n \to \mathbb{R}^j$ is a measurement function. The unknown, hidden, or latent state $(X(t))_{t\geq 0}$ is assumed to satisfy the stochastic differential equation.

$$\mathrm{d}X(t) = f(X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t),\ t > 0; \quad X(0) \sim p_0. \tag{4.2}$$

The Wiener process $(W(t))_{t\geq 0}$ is d-dimensional. We are ready to state the filtering problem.

The filtering problem: Given noisy measurements $(\widetilde{Y}_k)_{k=0}^K$, measured from some physical, biological, or technical process through some sensor, under the assumption that $\widetilde{Y}$ satisfy in distribution the measurement and state models (4.1)–(4.2), estimate for all times $t \geq 0$ the probability distribution of the unknown state $X(t)$, given all observations $\widetilde{Y}_k$, up to and including time t. If there exists a density of this distribution, then we can search for a family $(\pi(t))_{t\geq 0}$ conditional densities satisfying

$$\mathbb{P}(X(t) \in A \mid Y_0 = \widetilde{Y}_0, \dots, Y_k = \widetilde{Y}_k) = \int_A \pi(t, x)\,\mathrm{d}x, \quad t \in [t_k, t_{k+1}),$$

for (measurable) sets $A \subset \mathbb{R}^n$.

We distinguish data $(\widetilde{Y}_k)_{k=0}^K$ and model output $(Y_k)_{k=0}^K$ in order to emphasize the statistical nature of the problem. Otherwise, in the literature, only $(Y_k)_{k=0}^K$ is most often introduced, as if the data were generated by (4.1)–(4.2), or as if the statistical interpretation was tacitly known to all readers. From here on, we only consider the model output $(Y_k)_{k=0}^K$, but keep in mind that real data only approximately satisfy it in distribution.

4.2 The exact filter

The filtering problem is solved in a Bayesian way with predictions between observation times and an update using the Bayes rule. This is why it is commonly referred to as the Bayesian filtering problem. To formalize the filter, we start by defining the likelihood function $\ell(x, y) = p(y|x) = \mathcal{N}(y; h(x), R)$, where $\mathcal{N}(\cdot; \mu, R)$ is the Gaussian density function with mean μ and covariance R. Here we use the common convention in statistics to let p denote probability densities and let the arguments show what density is meant. At the initial time $t = 0$, we have the unconditional distribution $X(0) \sim p_0$. Taking into account the measurement Y_0 at this time and applying the Bayes rule, we have

$$\pi(0, x) = p(X_0 = x \mid Y_0) = \frac{p(Y_0 \mid X_0 = x)p(X_0 = x)}{p(Y_0)} = \frac{\ell(x, Y_0)p_0(x)}{p(Y_0)}. \tag{4.3}$$

Next in the time interval (t_0, t_1) there is no measurement, and only prediction is needed to compute $\pi(t)$ in this interval. But we know that the density of an SDE evolves with the Kolmogorov Forward Equation (KFE), also called the Fokker–Planck equation, and thus the filtering density satisfies, in the given time interval, the equation

$$\pi_t = L^*\pi, \tag{4.4}$$

where for $a^{ij} = (\sigma^T\sigma)_{i,j}$ we have that

$$L^*\varphi = \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}(a^{ij}\varphi) - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}(f_i\varphi),$$

is the formal adjoint of the operator L introduced in Theorem 2.1. We remind the reader that the fact that L^* is the formal adjoint of L means that for all sufficiently regular $\varphi, \psi \colon \mathbb{R}^n \to \mathbb{R}$ it holds that

$$\int_{\mathbb{R}^n} (L\varphi)(x)\psi(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} \varphi(x)(L^*\psi)(x)\,\mathrm{d}x.$$

With (4.3) and (4.4) we have the filtering distribution $\pi(t)$ for $t \in [0, t_1)$ and we notice that $\pi(t_1^-, x) = \lim_{t \uparrow t_1} \pi(t, x) = p(X(t_1) = x|Y_0)$. Using this fact, we do the update at time step t_1 with the Bayes rule, i.e.,

$$\begin{aligned}
\pi(t_1, x) &= p(X(t_1) = x \mid Y_0, Y_1) \\
&= \frac{p(Y_1 \mid X(t_1) = x, Y_0)p(X(t_1) = x \mid Y_0)}{p(Y_1 \mid Y_0)} \\
&= \frac{p(Y_1 \mid X(t_1) = x)p(X(t_1) = x \mid Y_0)}{p(Y_1 \mid Y_0)} \\
&= \frac{\ell(x, Y_1)\pi(t_1^-, x)}{p(Y_1 \mid Y_0)}.
\end{aligned}$$

In the third equality, we use the fact that when we know the value of $X(t_1)$, then knowing Y_0 does not contribute with information in the prediction of Y_1 and thus $p(Y_1 \mid X(t_1) = x, Y_0) = p(Y_1 \mid X(t_1) = x)$. In the vocabulary of Bayesian networks or graphical models, Y_1 is said to be conditionally independent of Y_0 given $X(t_1)$. The given procedure can be iterated and this gives an exact filter. In order to formulate the filter on $[0, \infty)$, we define $t_{K+1} := \infty$ and let $\pi(t)$ for $t > t_K$ be pure prediction. We also let $Y_{0:k}$ denote the $j \times (k+1)$-matrix given by $Y_{0:k} = (Y_0, Y_1, \dots, Y_k)$. We formulate the exact filter as a theorem.

Theorem 4.1 (The exact filter with discrete time observations)**.** *Let X be the weak solution to (4.2) and let $Y_0, \dots, Y_K \in \mathbb{R}^j$ be measurements. Then the solution $(\pi(t))_{t \geq 0}$ to the filtering problem satisfies the initial value*

$$\pi(0, x) = \frac{\ell(x, Y_0) p_0(x)}{p(Y_0)}, \quad x \in \mathbb{R}^n,$$

and the recursion

$$\begin{aligned} \pi_t(t, x) &= (L^* \pi)(t, x), && t \in [t_k, t_{k+1}),\ x \in \mathbb{R}^n,\ k = 0, \dots, K, \\ \pi(t_k, x) &= \frac{\ell(x, Y_k) \pi(t_k^-, x)}{p(Y_k \mid Y_{0:k-1})}, && x \in \mathbb{R}^n,\ k = 1, \dots, K. \end{aligned}$$

In settings where the KFE is analytically solvable and $p(Y_k \mid Y_{0:k-1})$ is easily computable, the theorem provides a computationally feasible and exact filter. Unfortunately, in general, neither is easy. However, if one wants to define an approximate filtering method by applying some numerical method to the KFE, then one can remove the problem of computing $p(Y_k \mid Y_{0:k-1})$ by considering an unnormalized filter instead, that is, exact up to a normalizing constant.

4.3 The unnormalized exact filter

The Kolmogorov Forward Equation is a linear equation. In particular, this means that any multiplicative constant to the initial value can equally well be multiplied to the solution instead. For example, if we define $\rho(0, x) = \ell(x, Y_0) p_0(x)$ and

$$\rho_t(t, x) = (L^* \rho)(t, x), \quad t \in [0, t_1), \ \in \mathbb{R}^n,$$

then

$$\pi(t, x) = \frac{\rho(t, x)}{p(Y_0)}.$$

Iterating this principle, we get an unnormalized filter which can, if desired, be normalized at any instance of time. This is used in [3], where the KFE is solved with a numerical scheme based on deep learning known as deep splitting. Deep learning is attractive because it scales very well in increasing dimensions compared to, e.g., finite element methods or finite difference methods that suffer under the so-called curse of dimensionality and cannot be applied for $n \geq 4$ and are only highly efficient for $n = 1$, that is, in one dimension.

Theorem 4.2 (The exact unnormalized filter with discrete time observations)**.** *Let X be the weak solution to (4.2), $Y_0, \dots, Y_K \in \mathbb{R}^j$ be measurements and $(\rho(t))_{t \geq 0}$ satisfy the initial value*

$$\rho(0, x) = \ell(x, Y_0) p_0(x), \quad x \in \mathbb{R}^n,$$

and the recursion

$$\rho_t(t, x) = (L^* \rho)(t, x), \qquad t \in [t_k, t_{k+1}),\ x \in \mathbb{R}^n,\ k = 0, \dots, K, \tag{4.5}$$

$$\rho(t_k, x) = \ell(x, Y_k) \rho(t_k^-, x), \quad x \in \mathbb{R}^n,\ k = 1, \dots, K. \tag{4.6}$$

Then we have

$$\pi(t, x) = \frac{\rho(t, x)}{\int \rho(t, x') \, \mathrm{d}x'} = \frac{\rho(t, x)}{p(Y_0) \prod_{i=1}^{k} p(Y_i \mid Y_{0:i-1})}, \quad t \in [t_k, t_{k+1}),\ x \in \mathbb{R}^n,\ k = 0, \dots, K.$$

4.4 The filtering equations

In this section we step away from the specific SDE setting and consider a broader filtering problem with model inputs

$p(X_0)$,	Initial value prior,
$p(X_{k+1} \mid X_k)$,	Transition probabilities,
$p(Y_k \mid X_k)$,	Measurement likelihood.

We notice that an SDE model gives rise to transition probabilities, via the Kolmogorov Forward Equation, and is thus a special case. The so-called filtering equations, which in fact are formulas and not equations are:

$$p(X_{k+1} \mid Y_{0:k}) = \int p(X_{k+1} \mid X_k) p(X_k \mid Y_{0:k}) \, \mathrm{d}X_k, \qquad \text{Prediction equation,}$$

$$p(X_k \mid Y_{0:k}) = \frac{p(Y_k \mid X_k) p(X_k \mid Y_{0:k})}{\int p(Y_k \mid X_k') p(X_k' \mid Y_{0:k}) \, \mathrm{d}X_k}, \qquad \text{Update equation.}$$

The prediction equation is obtained by simply factorizing the joint probability

$$p(X_{k+1}, X_k \mid Y_{0:k}) = p(X_{k+1} \mid X_k, Y_{0:k}) p(X_k \mid Y_{0:k}) = p(X_{k+1} \mid X_k) p(X_k \mid Y_{0:k}),$$

and marginalize, also called the Chapman–Kolmogorov equation. In the last equality, we used the fact that the probability of the new state X_{k+1} given the previous state X_k and the measurements $Y_{0:k}$, the state contains all the necessary information. In the language of Bayesian networks, X_{k+1} is conditionally independent of $Y_{0:k}$ given X_k. The update equation is the same as we have seen before. For an exposition of filtering in this more general setting, see [9].

4.5 The Kalman filter

The only practically interesting filtering problem that can be solved exactly is the case with linear state equation and linear measurement model, leading to the famous Kalman filter. We thus consider the linear state equation

$$\mathrm{d}X(t) = FX(t) \, \mathrm{d}t + \sigma(t) \, \mathrm{d}W(t),\ t > 0; \quad X(0) \sim \mathcal{N}(\xi_0, P_0), \tag{4.7}$$

and the linear measurement model

$$Y_k = H_k X(t_k) + r_k,\ k \geq 0; \quad r_k \sim \mathcal{N}(0, R). \tag{4.8}$$

For the filter prediction step, we use the transition probabilities of X and therefore solve (4.7). For this purpose, we use the matrix exponential e^{-tF} in an integrating factor argument. The matrix exponential is given by the Taylor series

$$e^{-tF} = \sum_{\ell=0}^{\infty} \frac{(-tF)^\ell}{\ell!}, \quad t \in \mathbb{R}.$$

Multiplying both sides of (4.7) with e^{-tF} and using the property that $\frac{\mathrm{d}}{\mathrm{d}t} e^{-tF} = -Fe^{-tF}$, we have by the product rule

$$\mathrm{d}(e^{-tF} X(t)) = e^{-tF}(\,\mathrm{d}X(t) - FX(t)\,\mathrm{d}t) = e^{-tF}\sigma(t)\,\mathrm{d}W(t).$$

Integrating from s to t we have

$$e^{-tF} X(t) = e^{-sF} X(s) + \int_s^t e^{-rF}\sigma(r)\,\mathrm{d}W(r),$$

and multiplying both sides by e^{tF} and using the fact that $e^{tF}e^{-tF} = \mathrm{id}$ gives

$$X(t) = e^{(t-s)F} X(s) + \int_s^t e^{(t-r)F}\sigma(r)\,\mathrm{d}W(r).$$

From this we first conclude that

$$\mathbb{E}\big[X(t) \mid X(s)\big] = e^{(t-s)F} X(s).$$

We recall that the covariance matrix of a random vector Z is the unique nonnegative definite matrix Q satisfying $\mathbb{E}[\langle Z - \mathbb{E}[Z], u\rangle\langle Z - \mathbb{E}[Z], v\rangle] = \langle Qu, v\rangle$ for all vectors u, v. To compute the covariance matrix of $X(t)$ given $X(s)$ we use the Itô Isometry and some elementary operations to obtain

$$\begin{aligned}
&\mathbb{E}\Big[\big\langle X(t) - \mathbb{E}[X(t) \mid X(s)], u\big\rangle\big\langle X(t) - \mathbb{E}[X(t) \mid X(s)], v\big\rangle \mid X(s)\Big] \\
&= \mathbb{E}\left[u^T \int_s^t e^{-(t-r)F}\sigma(r)\,\mathrm{d}W(r) v^T \int_s^t e^{-(t-r)F}\sigma(r)\,\mathrm{d}W(r)\right] \\
&= \mathbb{E}\left[\int_s^t u^T e^{-(t-r)F}\sigma(r)\,\mathrm{d}W(r) \int_s^t v^T e^{-(t-r)F}\sigma(r)\,\mathrm{d}W(r)\right] \\
&= \int_s^t \big\langle u^T e^{-(t-r)F}\sigma(r), v^T e^{-(t-r)F}\sigma(r)\big\rangle\,\mathrm{d}r \\
&= \int_r^t \big\langle \sigma(r)^T e^{-(t-r)F^T} u, \sigma(r)^T e^{-(t-r)F^T} v\big\rangle\,\mathrm{d}r \\
&= \int_s^t \big\langle e^{-(t-r)F}\sigma(r)\sigma(r)^T e^{-(t-r)F^T} u, v\big\rangle\,\mathrm{d}r \\
&= \left\langle \int_s^t e^{-(t-r)F}\sigma(r)\sigma(r)^T e^{-(t-r)F^T}\,\mathrm{d}r\, u, v\right\rangle.
\end{aligned}$$

Thus, we conclude that

$$X(t) \mid X(s) \sim N\left(e^{-(t-s)F} X(s), \int_s^t e^{-(t-r)F}\sigma(r)\sigma(r)^T e^{-(t-r)F^T}\,\mathrm{d}r\right).$$

Assuming that $X(s) \sim \mathcal{N}(\mu_s, \Sigma_s)$ we have

$$\begin{aligned}
X(t) &\sim \mathcal{N}\left(e^{-(t-s)F}\mu_s, e^{-(t-s)F}\Sigma_s e^{-(t-s)F^T} + \int_s^t e^{-(t-r)F}\sigma(r)\sigma(r)^T e^{-(t-r)F^T}\,\mathrm{d}r\right) \\
&=: \mathcal{N}(\mu(t), \Sigma(t)).
\end{aligned}$$

The matrix exponential $e^{(t-s)F}$ is the solution (operator) to the Ordinary Differential Equation (ODE)

$$\dot{\mu} = F\mu,\ t > s; \quad \mu(s) = \mu_s,$$

meaning that $\mu(t) = e^{(t-s)F}\mu_s$, $t \geq s$. Thus, the mean $\mu(t)$ of $X(t)$ for $t \geq s$ is solved by this equation. Moreover, we claim that the covariance matrix $\Sigma(t)$ of $X(t)$ for $t \geq s$ is the solution to the ODE

$$\dot{\Sigma} = F\Sigma + \Sigma F^T + \sigma\sigma^T,\ t \geq s; \quad \Sigma(s) = \Sigma_s.$$

It is called a Lyapunov equation and is in fact a Riccati equation without quadratic term. The claim is proved by multiplying integrating factors from both the left- and the right-hand sides of the Lyapunov equation, and by the product rule. More precisely,

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{-tF}\Sigma(t)e^{-tF^T} = e^{-tF}\big(\dot{\Sigma}(t) - F\Sigma(t) - \Sigma(t)F^T\big)e^{-tF^T} = e^{-tF}\sigma(t)\sigma(t)^T e^{-tF^T}.$$

Integrating from s to t yields

$$e^{-tF}\Sigma(t)e^{-tF^T} = e^{-sF}\Sigma(s)e^{-sF^T} + \int_s^t e^{-rF}\sigma(r)\sigma(r)^T e^{-rF^T}\,\mathrm{d}r,$$

and multiplying by e^{tF} and e^{tF^T} from left and right, respectively, yields

$$\Sigma(t) = e^{(t-s)F}\Sigma(s)e^{(t-s)F^T} + \int_s^t e^{(t-r)F}\sigma(r)\sigma(r)^T e^{(t-r)F^T}\,\mathrm{d}r.$$

This was the claim. We now have two ODE for prediction mean and covariance.

Theorem 4.3 (Unconditional distribution)**.** *Let X be the solution to the Ornstein–Uhlenbeck process solving the SDE* (4.7)*. We have that $X(t) \sim \mathcal{N}(\xi(t), P(t))$, $t \geq 0$, where*

$$\begin{aligned} \dot{\xi} &= F\xi, \\ \dot{P} &= FP + PF^T + \sigma\sigma^T, \end{aligned}$$

with initial values $\xi(0) = \xi_0$ and $P(0) = P_0$.

In the Kalman filter, the predictions follow between the update times, the same two ODE as in Theorem 4.3. At the update times they have discontinuities from the use of Bayes' formula. We next state the Kalman filter and motivate the update step in the proof.

Theorem 4.4 (Kalman filter). *Let X be the solution to the Ornstein–Uhlenbeck process solving the SDE (4.7) and let Y be the measurements given by (4.8). The filter is exactly solvable and is given by $\pi(t,x) = \mathcal{N}(x;\xi(t),P(t))$, where for $k \geq 0$, $t \in [t_k, t_{k+1})$, we have predictions*

$$\begin{aligned} \dot{\xi} &= F\xi, \\ \dot{P} &= FP + PF^T + \sigma\sigma^T, \end{aligned}$$

and updates

$$\begin{aligned} \xi(t_k) &= \xi(t_k^-) + K_k Z_k, \\ P(t_k) &= (I - K_k H_k)P(t_k^-), \end{aligned}$$

where

$$\begin{aligned} Z_k &= Y_k - H_k\xi(t_k^-), \\ S_k &= H_k P(t_k^-)H_k^T + R, \\ K_k &= P(t_k^-)H_k^T S_k^{-1}, \end{aligned}$$

and $\xi(t_0^-) := \xi_0$, $P(t_0^-) := P_0$. Moreover, $p(Y_k \mid Y_{0:k-1}) = \mathcal{N}(H\xi(t_k^-), S_k)$.

The proof of the update step of the Kalman filter is based on the following general well-known lemma on jointly Gaussian random variables.

Lemma 4.5. *Let the random variables x and y have the joint probability distribution*

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} A & C \\ C^T & B \end{bmatrix} \right).$$

Then the conditional probability of x given y is given by

$$p(x|y) = \mathcal{N}(a + CB^{-1}(y-b), A - CB^{-1}C^T).$$

Proof of Theorem 4.4. At the initial time $t_0 = 0$ we have

$$p(X(t_0), Y_0) = \mathcal{N}(m_0, C_0),$$

where

$$m_0 = \begin{bmatrix} \xi_0 \\ H_0\xi_0 \end{bmatrix}, \quad C_0 = \begin{bmatrix} P_0 & P_0 H_0^T \\ H_0 P_0 & H_0 P_0 H_0^T + R \end{bmatrix}.$$

Lemma (4.5) applied to $x = X(t_0)$ and $y = Y_0$ yields the distribution

$$\begin{aligned} \pi(t_0) &= p(X(t_0)|Y_0) \\ &= N(\xi_0 + P_0 H_0^T (H_0 P_0 H_0^T + R)^{-1}(Y_0 - H_0\xi_0), P_0 - P_0 H_0^T (H_0 P_0 H_0^T + R)^{-1} H_0 P_0), \end{aligned}$$

and the update formulas can be identified. By the arguments prior to the theorem we have for $t \in [0, t_1)$ the predictions $\pi(t) = \mathcal{N}(\xi(t), P(t))$. At time t_1 we similarly have $p(X(t_1), Y_1|Y_0) = \mathcal{N}(m_1, C_1)$, where

$$m_1 = \begin{bmatrix} \xi(t_1^-) \\ H_1\xi(t_1^-) \end{bmatrix}, \quad C_1 = \begin{bmatrix} P(t_1^-) & P(t_1^-)H_1^T \\ H_1 P(t_1^-) & H_1 P(t_1^-) H_1^T + R \end{bmatrix},$$

and the update formulas are obtained just as for time t_0. The procedure is repeated inductively, and this completes the proof. ∎

The derived Kalman filter propagates the mean and covariance of the filter distribution over time. It is also possible to derive, at least at the measurement times, a propagation of the mean and precision, where the latter is the inverse of the covariance matrix. This is the information Kalman filter. Depending on the dimensions of the measurement and state spaces, one or the other is more advantegous from the computational perspective.

The information Kalman filter: Derive a Kalman filter at the measurement times, where the precision matrix $P_k^{-1} := P(t_k)^{-1}$ is updated instead of $P(t_k)$. Use the unnormalized update formula (4.6) and use completion of squares.

A Acknowledgement

The author thanks Kristoffer Andersson, Kasper Bågmark, Karl Hammar, Per Ljung, Benjamin Svedung Wettervik, and Johan Ulander for great feedback that has helped to improve the text.

References

[1] K. Andersson, A. Andersson, and C. Oosterlee, *Convergence of a robust deep FBSDE method for stochastic control*, SIAM Journal on Scientific Computing (2023).

[2] ———, *The deep multi-FBSDE method: a robust deep learning method for coupled FBSDEs*, arXiv preprint arXiv:2503.13193 (2025).

[3] K. Bågmark, A. Andersson, S. Larsson, and F. Rydin, *A convergent scheme for the Bayesian filtering problem based on the Fokker–Planck equation and deep splitting*, arXiv preprint arXiv:2409.14585 (2024).

[4] C. Beck, W. E, and A. Jentzen, *Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations*, Journal of Nonlinear Science **29** (2019), 1563–1619.

[5] N. Chopin, A. Fulop, J. Heng, and A. H Thiery, *Computational doob h-transforms for online filtering of discretely observed diffusions*, International conference on machine learning, 2023, pp. 5904–5923.

[6] W. E, J. Han, and A. Jentzen, *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*, Commun. Math. Stat **5** (November 2017), 349–380.

[7] N. El Karoui, S. Peng, and M. C. Quenez, *Backward stochastic differential equations in finance*, Mathematical finance (1997).

[8] W. H Fleming and H. M. Soner, *Controlled Markov processes and viscosity solutions*, Vol. 25, Springer Science & Business Media, 2006.

[9] S. Särkkä and L. Svensson, *Bayesian filtering and smoothing*, Vol. 17, Cambridge university press, 2023.

[10] N. Touzi, *Stochastic control problems, viscosity solutions and application to finance*, Scuola normale superiore, 2004.