

Bayesian sequential least-squares estimation for the drift of a Wiener process

Erik Ekström^{a,*}, Ioannis Karatzas^{b,c,2}, Juozas Vaicenavicius^d

^a *Department of Mathematics, Uppsala University, 75106 Uppsala, Sweden*

^b *Department of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027, USA*

^c *Intech Investment Management, One Palmer Square, Suite 441, Princeton, NJ 08542, USA*

^d *Sensmetry UAB, J. Jasinskio 16A, 03163 Vilnius, Lithuania*

Received 15 January 2019; received in revised form 17 June 2019; accepted 12 September 2019

Available online 21 September 2019

Abstract

Given a Wiener process with unknown and unobservable drift, we try to estimate this drift as effectively but also as quickly as possible, in the presence of a quadratic penalty for the estimation error and of a fixed, positive cost per unit of observation time. In a Bayesian framework, where the unobservable drift is assumed to have a known “prior” distribution, this question reduces to choosing judiciously a stopping time for an appropriate diffusion process in natural scale. We establish structural properties of the solution for the corresponding problem of optimal stopping. In particular, we show that, regardless of the prior distribution, the continuation region is monotonically shrinking in time. Moreover, we provide conditions on the prior distribution that guarantee a one-sided stopping region. Lastly, some concrete prior distributions are studied to illustrate the theoretical results.

© 2019 Elsevier B.V. All rights reserved.

MSC: primary 62L12; secondary 62L10; 60G40

Keywords: Sequential estimation; Sequential analysis; Optimal stopping

* Corresponding author.

E-mail addresses: ekstrom@math.uu.se (E. Ekström), ikl@columbia.edu, ikaratzas@intechjanus.com

(I. Karatzas), juozas.vaicenavicius@sensmetry.com (J. Vaicenavicius).

¹ Support from the Swedish Research Council is gratefully acknowledged.

² Support from the National Science Foundation, USA under NSF grant DMS-14-05210 and DMS-20-04997, is gratefully acknowledged.

1. Introduction

Imagine trying to estimate a quantity about which there is considerable uncertainty, and which cannot be observed directly. Instead, one has access to a stream of observations that this unobservable quantity affects and, based on this stream, tries to find an estimator of the unobservable quantity which is “optimal” in the sense of least-squares. However, access to the stream of information is costly: one pays a fixed, positive cost per unit of time, for as long as information is being obtained. How does one then resolve the dilemma inherent in this situation, which calls for balancing the conflicting requirements of fidelity in estimation and of cost minimization?

We study here an instance of this problem in a highly idealized and stylized form, and in a Bayesian setting. Namely, we assume that the unobservable quantity is a random variable X with known distribution, and that one observes sequentially the process

$$Y(t) = Xt + W(t), \quad 0 \leq t < \infty. \quad (1.1)$$

Here W is a standard Wiener process, independent of the random variable X . Moreover, we assume that the known, “prior” distribution μ of X has finite second moment and is *non-atomic*; that is, we exclude the trivial case where μ is a one-point distribution. We posit that, at any given time $t \in [0, \infty)$, we have access to the observations

$$\sigma\{Y(s), 0 \leq s \leq t\}.$$

The right-continuous augmentation $\mathbb{F}^Y = \{\mathcal{F}^Y(t)\}_{0 \leq t < \infty}$ of the family of σ -algebras $(\sigma\{Y(s), 0 \leq s \leq t\})_{0 \leq t < \infty}$ is called the *observations filtration*, and we set

$$\mathcal{F}^Y(\infty) := \sigma\left(\bigcup_{0 \leq t < \infty} \mathcal{F}^Y(t)\right).$$

We denote by $\mathcal{T}^Y$ the collection of stopping times of this filtration $\mathbb{F}^Y$, to wit, the collection of random variables $\tau : \Omega \rightarrow [0, \infty)$ with $\{\tau \leq t\} \in \mathcal{F}^Y(t)$ for every $t \in [0, \infty)$.

Based on the flow of information $\mathbb{F}^Y$, we construct the least-squares estimate

$$\hat{X}(t) = \mathbb{E}[X | \mathcal{F}^Y(t)], \quad 0 \leq t < \infty \quad (1.2)$$

of the unobserved variable. In this work we seek to compute the minimal expected cost

$$C_* = \inf_{\tau \in \mathcal{T}^Y} C(\tau), \quad C(\tau) := \mathbb{E}[(X - \hat{X}(\tau))^2 + c\tau], \quad (1.3)$$

and to determine whether it is attained by some stopping time $\tau_* \in \mathcal{T}^Y$. Here $c > 0$ is a given real constant, representing the cost of one unit of delay in the estimation procedure. The positivity of this constant, along with the obvious bound $C_* \leq C(0) = \text{Var}(X) < \infty$, implies that we may restrict attention in (1.3) to stopping times τ with

$$\mathbb{E}[\tau] < \infty. \quad (1.4)$$

Preview: We show in Section 2 that the least-squares estimate $\hat{X}$ in (1.2) is an Itô diffusion in natural scale, and describe its dynamics in detail. The question of (1.3) is cast in Section 3 as a problem of optimal stopping for this process $\hat{X}$, with a cost criterion that involves only the constant $c > 0$ and the local variance function of $\hat{X}$. This piece of serendipity allows us to obtain quite general qualitative properties of the solution, as developed in Section 5. In particular, we show that the continuation region contracts in time for any given prior distribution, and we provide conditions under which the continuation region and the stopping

region are separated by a single curve. Some explicit results for the Gaussian and Bernoulli cases are presented in Sections 3 and 4, respectively. Finally, Section 6 presents an additional example in which the prior distribution is symmetric.

2. Preliminaries on the conditional mean and variance processes

In this section we recall a general result regarding the conditional mean and variance processes from the theory of filtering. We build then on this result in order to unveil the structure of our problem at hand, as well as the stochastic dynamics of the processes that are crucial for its analysis.

2.1. Projecting onto the observations filtration

We first recall the conditional distribution of X given observations on the process Y . For a proof we refer to [1, Proposition 3.16].

Proposition 2.1. *Consider a function $q : \mathbb{R} \rightarrow \mathbb{R}$ satisfying the integrability condition*

$$\int_{\mathbb{R}} |q(u)| \mu(du) < \infty.$$

Then, for any $t \geq 0$, we have

$$\mathbb{E}[q(X) | \mathcal{F}^Y(t)] = \frac{\int_{\mathbb{R}} q(u) \exp\{uy - u^2t/2\} \mu(du)}{\int_{\mathbb{R}} \exp\{uy - u^2t/2\} \mu(du)} \Big|_{y=Y(t)}.$$

On the strength of Proposition 2.1, the conditional expectation of X given the observations up to time $t \in (0, \infty)$ is given as

$$\widehat{X}(t) := \mathbb{E}[X | \mathcal{F}^Y(t)] = G(t, Y(t)), \quad (2.1)$$

where for each $(t, y) \in [0, \infty) \times \mathbb{R}$ we set

$$G(t, y) := \frac{\int_{\mathbb{R}} u \exp\{uy - u^2t/2\} \mu(du)}{\int_{\mathbb{R}} \exp\{uy - u^2t/2\} \mu(du)} = \int_{\mathbb{R}} u \mu_{t,y}(du) \quad (2.2)$$

and

$$\mu_{t,y}(A) := \frac{\int_A \exp\{uy - u^2t/2\} \mu(du)}{\int_{\mathbb{R}} \exp\{uy - u^2t/2\} \mu(du)}, \quad A \in \mathcal{B}(\mathbb{R}). \quad (2.3)$$

This measure $\mu_{t,y}$ is the conditional (“posterior”) distribution of X at time t , given the values $Y(s)$, $0 \leq s < t$, and $Y(t) = y$ of the observation process up to that time.

We have a similar computation for the conditional variance

$$\text{Var}(X | \mathcal{F}^Y(t)) = \mathbb{E}[(X - \widehat{X}(t))^2 | \mathcal{F}^Y(t)] = \mathbb{E}[X^2 | \mathcal{F}^Y(t)] - \widehat{X}^2(t) = H(t, Y(t))$$

of X given the observations up to time $t \in (0, \infty)$, where

$$\begin{aligned} H(t, y) &:= \frac{\int_{\mathbb{R}} u^2 \exp\{uy - u^2t/2\} \mu(du)}{\int_{\mathbb{R}} \exp\{uy - u^2t/2\} \mu(du)} - \left(\frac{\int_{\mathbb{R}} u \exp\{uy - u^2t/2\} \mu(du)}{\int_{\mathbb{R}} \exp\{uy - u^2t/2\} \mu(du)} \right)^2 \\ &= \int_{\mathbb{R}} u^2 \mu_{t,y}(du) - \left(\int_{\mathbb{R}} u \mu_{t,y}(du) \right)^2 = \int_{\mathbb{R}} (u - G(t, y))^2 \mu_{t,y}(du). \end{aligned} \quad (2.4)$$

It is straightforward from (2.2);(2.4) that the quantities $G(t, y)$, $H(t, y)$ are, respectively, the center of gravity and the second central moment of the measure $\mu_{t,y}$ in (2.3); to wit,

$$\begin{aligned} G(t, y) &= \mathbb{E}[X \mid Y(s), 0 \leq s < t; Y(t) = y], \\ H(t, y) &= \text{Var}[X \mid Y(s), 0 \leq s < t; Y(t) = y]. \end{aligned}$$

The function G of (2.2) is strictly increasing in its spatial variable, has partial derivatives of all orders on $(0, \infty) \times \mathbb{R}$, and satisfies on this strip the *Backwards Burgers equation*

$$\partial G + \frac{1}{2} D^2 G + G \cdot DG = 0. \quad (2.5)$$

The gradient of the function G , i.e., the function

$$H = DG$$

of (2.4), is positive on the strip $(0, \infty) \times \mathbb{R}$ and satisfies there the equation

$$\partial H + \frac{1}{2} D^2 H + G \cdot DH + H^2 = 0. \quad (2.6)$$

Here and throughout, we are denoting by $\partial \equiv \partial/\partial t$ the partial derivative with respect to the temporal argument t , and by $D^k \equiv \partial^k/\partial y^k$ the partial derivative of order $k = 1, 2, \dots$ with respect to the spatial argument, in this case y .

Remark 2.2 (Bijections). Let now $\mathcal{I}_\mu$ denote the interior of the smallest closed interval containing the support of the probability measure μ , i.e.,

$$\mathcal{I}_\mu = (\inf(\mathcal{S}_\mu), \sup(\mathcal{S}_\mu)) \quad \text{with} \quad \mathcal{S}_\mu := \text{supp}(\mu). \quad (2.7)$$

Then, for any given $t \in [0, \infty)$, the function

$$G_t(\cdot) \equiv G(t, \cdot) : \mathbb{R} \rightarrow \mathcal{I}_\mu$$

defined in (2.2) is a strictly increasing, continuous bijection (see also [3]). The strict increase of this function $G_t(\cdot)$ implies that $Y(t) = G_t^{-1}(\widehat{X}(t))$ holds for $0 \leq t < \infty$.

To wit, the observation processes Y and the least-squares estimate process $\widehat{X}$ are bijections of each other pointwise in time, and thus generate the same filtration. In particular, $G_t^{-1}(x)$ is the unique value of the observation process $Y(t)$ at time t , that yields $\widehat{X}(t) = x$.

Remark 2.3 (The Widder Transform). The derivation of the parabolic backwards partial differential equations (2.5), (2.6) is facilitated by the observation that G is itself the logarithmic gradient $G = D \log F$ of the function

$$F(t, y) := \int_{\mathbb{R}} \exp\left\{uy - \frac{t}{2} u^2\right\} \mu(du), \quad (t, y) \in (0, \infty) \times \mathbb{R} \quad (2.8)$$

that appears in the denominators of (2.2), (2.4). It is checked easily that this function, the so-called *Widder Transform* of the prior distribution μ , solves the backward heat equation

$$\partial F + \frac{1}{2} D^2 F = 0.$$

Conversely, as shown by Widder [14] and Robbins & Siegmund [10], every positive solution of this backward heat equation can be written in the form (2.8), in terms of an appropriate measure μ on $\mathcal{B}(\mathbb{R})$. For a probabilistic treatment and development of the relevant theory, see Section 3.4.B in [7].

For technical convenience, we shall impose henceforth the following integrability condition:

Assumption 2.4. For some real number $a > 0$, the prior distribution μ satisfies

$$\int_{\mathbb{R}} \exp\{au^2\} \mu(du) < \infty. \quad (2.9)$$

Assumption 2.4 is a rather mild requirement in the sense that, for any $t > 0$, the integrability condition (2.9) is satisfied by the posterior distribution $\mu_{t,y}$ of any prior μ . The assumption allows us to extend the definition of $\mu_{t,y}$ in (2.3) to include points of the type $(0, y)$; and the resulting $\mu_{0,y}$ coincides with the posterior distribution in a scenario with prior distribution

$$\xi(A) := \frac{\int_A \exp\{au^2/2\} \mu(du)}{\int_{\mathbb{R}} \exp\{au^2/2\} \mu(du)}, \quad A \in \mathcal{B}(\mathbb{R}),$$

conditional on observing $Y(a) = y$. Consequently, the points $(0, y)$ can be regarded as interior points for a shifted problem started instead at time $-a$; it is therefore clear that, for instance, (2.5) holds then on the whole domain $[0, \infty) \times \mathbb{R}$.

2.2. Dynamics under the observations filtration

The process

$$\widehat{W}(t) := Y(t) - \int_0^t \widehat{X}(s) ds = \int_0^t (X - \widehat{X}(s)) ds + W(t), \quad 0 \leq t < \infty, \quad (2.10)$$

known as the *innovation process* in the theory of filtering, is clearly adapted to the observations filtration $\mathbb{F}^Y$. It is also a Wiener process of this filtration, as it is continuous, an $\mathbb{F}^Y$ -martingale, and has the right quadratic variation; for instance, see [1, Proposition 2.30 on p. 33].

We write $\mathbb{F}^{\widehat{W}} = (\mathcal{F}^{\widehat{W}}(t))_{0 \leq t < \infty}$ for the right-continuous augmentation of the filtration $\sigma\{\widehat{W}(s) : 0 \leq s \leq t\}$, $0 \leq t < \infty$, that this process generates; similarly, we shall use the notation $\mathbb{F}^{\widehat{X}} = (\mathcal{F}^{\widehat{X}}(t))_{0 \leq t < \infty}$ for the right-continuous augmentation of the filtration generated by the conditional expectation process $\widehat{X}$ in (2.1), and $\mathbb{F}^Y = (\mathcal{F}^Y(t))_{0 \leq t < \infty}$ for the right-continuous augmentation of the filtration generated by the observation process Y . Clearly, and in light of Remark 2.2, we have the comparisons $\mathbb{F}^{\widehat{W}} \subseteq \mathbb{F}^Y = \mathbb{F}^{\widehat{X}}$.

We deduce now from (2.1), (2.10) the representation for the observations process

$$dY(t) = \widehat{X}(t) dt + d\widehat{W}(t) = G(t, Y(t)) dt + d\widehat{W}(t) \quad (2.11)$$

as the solution of a stochastic differential equation driven by the innovations process $\widehat{W}$, with initial condition $Y(0) = 0$. Because of the smoothness of the function G , this equation admits a pathwise unique, strong solution, so we deduce the filtration identities

$$\mathbb{F}^{\widehat{W}} = \mathbb{F}^Y = \mathbb{F}^{\widehat{X}}. \quad (2.12)$$

On the other hand, with the notation $G_t(\cdot) = G(t, \cdot)$ already introduced in Remark 2.2, we set

$$\Psi(t, x) := DG(t, G_t^{-1}(x)) = H(t, G_t^{-1}(x)). \quad (2.13)$$

An application of Itô's formula to (2.1) yields, in conjunction with (2.5) and (2.11), a stochastic differential equation for the conditional mean process $\widehat{X}$ of (2.1), namely,

$$d\widehat{X}(t) = \Psi(t, \widehat{X}(t)) d\widehat{W}(t), \quad \widehat{X}(0) = \mathbb{E}(X) = \int_{\mathbb{R}} u \mu(du). \quad (2.14)$$

The function Ψ of (2.13), the dispersion function of the stochastic differential equation right above, can be expressed as

$$\Psi(t, x) = H(t, G_t^{-1}(x)) = \text{Var}[X \mid \widehat{X}(s), 0 \leq s < t; \widehat{X}(t) = x],$$

and a bit more generally as

$$\Psi(t, \widehat{X}(t)) = H(t, G_t^{-1}(\widehat{X}(t))) = H(t, Y(t)) = \text{Var}[X \mid \mathcal{F}^Y(t)]. \quad (2.15)$$

Furthermore, it is checked with the help of (2.5), (2.6) that the function $\Psi > 0$ of (2.13) satisfies on the strip $(0, \infty) \times \mathcal{I}_\mu$ the fully nonlinear, backwards parabolic equation

$$\partial \Psi + \Psi^2 \left(\frac{1}{2} D^2 \Psi + 1 \right) = 0. \quad (2.16)$$

Once again, we denote differentiation with respect to the temporal argument t by ∂ , and differentiation with respect to the spatial argument (in this case x) by D .

Finally, we recall from [3, Proposition 3.6] the following result about the function Ψ .

Proposition 2.5 (*Properties of the Dispersion Function Ψ*).

1. $\partial \Psi \leq 0$; consequently, by (2.16), we have $D^2 \Psi \geq -2$.
2. If μ is compactly supported, then the function Ψ is bounded.

3. Optimal stopping

The above considerations show that the optimal stopping problem (1.3) can be cast in the form

$$\inf_{\tau \in \mathcal{T}} \mathbb{E} \left[\Psi(\tau, \widehat{X}(\tau)) + c\tau \right]. \quad (3.1)$$

Here Ψ is the function of (2.13), the process $\widehat{X}$ satisfies the dynamics of (2.14), and $\mathcal{T}$ stands for the collection of stopping times of the filtration $\mathbb{F}^{\widehat{X}} = \mathbb{F}^{\widehat{W}} = \mathbb{F}^Y$, as in (2.12).

It is a noteworthy feature of this problem, that the same function Ψ of (2.13) appears both as the dispersion of the diffusion $\widehat{X}$ in (2.14), and as the cost function for the new formulation of the optimal stopping problem in (3.1). This feature makes the problem rather special, and aids considerably its analysis in Sections 4, 5.

Proposition 3.1. *For every stopping time $\tau \in \mathcal{T}$ we have*

$$\mathbb{E}[\Psi(\tau, \widehat{X}(\tau))] = \text{Var}(X) - \mathbb{E} \left[\int_0^\tau \Psi^2(s, \widehat{X}(s)) \, ds \right]. \quad (3.2)$$

Proof. From (1.1) and the strong law of large numbers for the Wiener process, we have $\lim_{t \rightarrow \infty} (Y(t)/t) = X$, a.e.; in other words, the random variable X is $\mathcal{F}^Y(\infty)$ -measurable. As a result, the P.Lévy martingale convergence theorem gives $\lim_{t \rightarrow \infty} \mathbb{E}(X^k | \mathcal{F}^Y(t)) = \mathbb{E}(X^k | \mathcal{F}^Y(\infty)) = X^k$ a.e., for $k = 1, 2$; therefore also

$$\lim_{t \rightarrow \infty} \Psi(t, \widehat{X}(t)) = \lim_{t \rightarrow \infty} \left(\mathbb{E}(X^2 | \mathcal{F}^Y(t)) - \left(\mathbb{E}(X | \mathcal{F}^Y(t)) \right)^2 \right) = 0$$

on the strength of (2.15). Now it follows from the dynamics in (2.14), the partial differential equation in (2.16), and elementary stochastic calculus, that the positive process

$$M(t) := \Psi(t, \widehat{X}(t)) + \int_0^t \Psi^2(s, \widehat{X}(s)) \, ds$$

$$= \text{Var}(X) + \int_0^t \Psi(s, \widehat{X}(s)) D\Psi(s, \widehat{X}(s)) d\widehat{W}(s), \quad 0 \leq t < \infty$$

is a local martingale. It is thus also a supermartingale, and consequently

$$\mathbb{E} \left(\Psi(\tau, \widehat{X}(\tau)) + \int_0^\tau \Psi^2(s, \widehat{X}(s)) ds \right) \leq \text{Var}(X) < \infty \quad (3.3)$$

holds by the optional sampling theorem for every stopping time $\tau \in \mathcal{T}$; this includes $\tau = \infty$, so we have also

$$\mathbb{E} \left(\int_0^\infty \Psi^2(s, \widehat{X}(s)) ds \right) \leq \text{Var}(X) < \infty. \quad (3.4)$$

We shall show presently that, as claimed in (3.2), the first inequalities in (3.3) and (3.4) hold actually as equalities.

In order to see these things, let us start from the observation that the representation

$$\widehat{X}(\tau) = \mathbb{E}(X) + \int_0^\tau \Psi(s, \widehat{X}(s)) d\widehat{W}(s) \quad (3.5)$$

from (2.14) holds for every stopping time $\tau \in \mathcal{T}$, including $\tau = \infty$: the martingale $\widehat{X}$ and the submartingale $\widehat{X}^2$ are both uniformly integrable. Thus, the representation

$$X - \widehat{X}(\tau) = \int_\tau^\infty \Psi(s, \widehat{X}(s)) d\widehat{W}(s)$$

holds, as does the analogue

$$\begin{aligned} \text{Var}(X \mid \mathcal{F}^Y(\tau)) &= \mathbb{E}[(X - \widehat{X}(\tau))^2 \mid \mathcal{F}^Y(\tau)] = \mathbb{E} \left(\int_\tau^\infty \Psi^2(s, \widehat{X}(s)) ds \mid \mathcal{F}^Y(\tau) \right) \\ &= \Psi(\tau, \widehat{X}(\tau)) \end{aligned}$$

of (2.15); for the second equality we have used the finite upper bound of (3.4). This, in turn, leads to

$$\mathbb{E}[\text{Var}(X \mid \mathcal{F}^Y(\tau))] = \mathbb{E}(X - \widehat{X}(\tau))^2 = \mathbb{E} \Psi(\tau, \widehat{X}(\tau)) = \mathbb{E} \int_\tau^\infty \Psi^2(s, \widehat{X}(s)) ds$$

upon taking expectations. In addition, (3.5) gives

$$\text{Var}(\mathbb{E}(X \mid \mathcal{F}^Y(\tau))) = \text{Var}(\widehat{X}(\tau)) = \mathbb{E} \int_0^\tau \Psi^2(s, \widehat{X}(s)) ds,$$

once again using the finite upper bound in (3.4). From these computations, and from a classical identity about variances, we deduce

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[\text{Var}(X \mid \mathcal{F}^Y(\tau))] + \text{Var}(\mathbb{E}(X \mid \mathcal{F}^Y(\tau))) \\ &= \mathbb{E} \left(\Psi(\tau, \widehat{X}(\tau)) + \int_0^\tau \Psi^2(s, \widehat{X}(s)) ds \right), \end{aligned}$$

that is, our claim (3.2). With the choice $\tau = \infty$, these considerations give the identity $\mathbb{E} \int_0^\infty \Psi^2(s, \widehat{X}(s)) ds = \text{Var}(X)$, as claimed. $\square$

The identity (3.2) allows us now to cast the optimal stopping problem of (1.3)/(3.1) in the equivalent form

$$v := \inf_{\tau \in \mathcal{T}} \mathbb{E} \left[\int_0^\tau (c - \Psi^2(s, \widehat{X}(s))) ds \right]. \quad (3.6)$$

3.1. Markovian framework

To study the optimal stopping problem in its new form (3.6), we first embed it into a Markovian framework, by allowing the diffusion $\widehat{X}$ of (2.14) to start at any given point $(t, x) \in [0, \infty) \times \mathcal{I}_\mu$. More precisely, we define the function $v : [0, \infty) \times \mathcal{I}_\mu \rightarrow (-\infty, 0]$ via

$$v(t, x) := \inf_{\tau \in \mathcal{T}} \mathbb{E} \left[\int_0^\tau \left(c - \Psi^2(t + s, \widehat{X}^{t,x}(t + s)) \right) ds \right], \quad (3.7)$$

where the dynamics of the process $\widehat{X} = \widehat{X}^{(t,x)}$ are given by

$$\begin{cases} d\widehat{X}(t + s) = \Psi(t + s, \widehat{X}(t + s)) d\widehat{W}(s) \\ \widehat{X}(t) = x \end{cases}$$

and $\widehat{W}$ is again standard scalar Brownian motion. Since $\tau = 0$ is an admissible stopping time, the value function v in (3.7) is non-positive: $v \leq 0$. On the other hand, it is clear from (3.1), (3.2) that $v(t, x) \geq -\text{Var}(X) > -\infty$, so v is also real-valued, as indicated.

In accordance with standard optimal stopping theory for Markov processes with continuous paths, we introduce the so-called continuation region

$$\mathcal{C} := \{(t, x) \in [0, \infty) \times \mathcal{I}_\mu : v(t, x) < 0\}$$

and its complement, the stopping region

$$\mathcal{D} := \{(t, x) \in [0, \infty) \times \mathcal{I}_\mu : v(t, x) = 0\}.$$

Moreover, for any given starting point (t, x) , we denote by

$$\tau^{(t,x)} := \inf\{s \geq 0 : (t + s, \widehat{X}^{(t,x)}(t + s)) \in \mathcal{D}\}$$

the first hitting time of the stopping region. Then we know (for instance, [4,5,12,13]) that the function $v : [0, \infty) \times \mathcal{I}_\mu \rightarrow [-\text{Var}(X), 0]$ of (3.7) is upper-semicontinuous, and that for each $(t, x) \in [0, \infty) \times \mathcal{I}_\mu$ the stopping time $\tau^{(t,x)}$ attains the infimum there, i.e.,

$$v(t, x) = \mathbb{E} \left[\int_0^{\tau^{(t,x)}} \left(c - \Psi^2(t + s, \widehat{X}^{(t,x)}(t + s)) \right) ds \right].$$

Remark 3.2. It is clear from the formulation (3.7) that immediate stopping ($\tau^{(t,x)} = 0$) is optimal if the inequality

$$c \geq \sup_{(t,x) \in [0, \infty) \times \mathcal{I}_\mu} \Psi^2(t, x) = \sup_{x \in \mathcal{I}_\mu} \Psi^2(0, x) \quad (3.8)$$

holds; here, the equality follows from Proposition 2.5. A bit more generally, if

$$c \geq \sup_{x \in \mathcal{I}_\mu} \Psi^2(T_c, x)$$

holds for some $T_c \in (0, \infty)$, then the strip $[T_c, \infty) \times \mathcal{I}_\mu$ belongs to the stopping region $\mathcal{D}$.

3.2. A very simple special case: The Gaussian distribution

As the simplest illustration, let us consider the Gaussian prior distribution μ with mean $m \in \mathbb{R}$ and variance $\sigma^2 \in (0, \infty)$, i.e.,

$$\mu(du) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(u - m)^2}{2\sigma^2}\right\} du,$$

a special case of the Kalman–Bucy filter. Here we have $\mathcal{I}_\mu = \mathbb{R}$, and the functions F , G , H and Ψ take the respective forms

$$F(t, y) = \frac{1}{\sqrt{1 + \sigma^2 t}} \exp \left\{ -\frac{1}{2\sigma^2} \left(\frac{(m + \sigma^2 y)^2}{1 + \sigma^2 t} - m^2 \right) \right\},$$

$$G(t, y) = \frac{m + \sigma^2 y}{1 + \sigma^2 t}, \quad H(t, y) = \Psi(t, x) = \frac{\sigma^2}{1 + \sigma^2 t} =: \xi(t).$$

Now, the function $t \mapsto c - \xi^2(t)$ is negative for $t \in [0, \frac{1}{\sqrt{c}} - \frac{1}{\sigma^2})$ if $\sqrt{c} < \sigma^2$, and it is everywhere non-negative if $\sqrt{c} \geq \sigma^2$. With

$$\tau_* := \left(\frac{1}{\sqrt{c}} - \frac{1}{\sigma^2} \right)^+,$$

it follows from [Remark 3.2](#) that the above constant τ_* is an optimal (albeit trivial!) stopping time in (3.6).

In [2], a similar result is obtained in the case when W is a fractional Brownian motion.

4. A time-homogeneous case: the Bernoulli distribution

As our second example, let us consider now the Bernoulli prior distribution

$$\mu = (1 - p)\delta_{-\beta} + p\delta_\beta$$

with symmetric support, where $p \in (0, 1)$ and $\beta \in (0, \infty)$. In this case we have $\mathcal{I}_\mu = (-\beta, \beta)$, as well as

$$G(t, y) = \beta \frac{p e^{\beta y} - (1 - p) e^{-\beta y}}{p e^{\beta y} + (1 - p) e^{-\beta y}}, \quad \text{thus} \quad H(t, y) = \beta^2 - G^2(t, y)$$

and

$$\Psi(t, x) = \beta^2 - x^2 =: \psi(x).$$

We note that we are here at the opposite extreme of the example in [Section 3.2](#): All these are functions of only the spatial variable; and the last of them does not even depend on the parameter $p \in (0, 1)$.

The stopping problem (3.6) thus takes the form

$$v(x) = \inf_{\tau \in \mathcal{T}} \mathbb{E} \left[\int_0^\tau \left(c - \psi^2(\widehat{X}(t)) \right) dt \right] \quad (4.1)$$

where $\widehat{X}$ is a diffusion in natural scale, with state-space $\mathcal{I}_\mu = (-\beta, \beta)$ and initial condition $x \in \mathcal{I}_\mu$:

$$\begin{cases} d\widehat{X}(t) = \psi(\widehat{X}(t)) d\widehat{W}(t), \\ \widehat{X}(0) = \beta(2p - 1) =: x \in \mathcal{I}_\mu. \end{cases} \quad (4.2)$$

We note that for $\beta^4 \leq c$, the integrand in (4.1) is non-negative, and hence the trivial stopping time $\tau_* \equiv 0$ is optimal.

Thus, we assume from now onwards that

$$\beta^4 > c;$$

then $c - \psi^2(x)$ is negative for $|x| < \gamma$ with

$$\gamma := \sqrt{\beta^2 - \sqrt{c}},$$

zero for $|x| = \gamma$, and positive for $\gamma < |x| < \beta$. Conjecturing that an optimal stopping rule is of the type

$$\tau_a^* := \inf\{t \geq 0 : |\widehat{X}(t)| \geq a\} \quad (4.3)$$

for some constant $a \in (\gamma, \beta)$, general optimal stopping theory leads to the following free-boundary problem:

To find a constant $a \in (\gamma, \beta)$ and an evenly symmetric function $u : (-\beta, \beta) \rightarrow (-\infty, 0]$ of class $\mathcal{C}^1((-\beta, \beta)) \cap \mathcal{C}^2((-\beta, \beta) \setminus \{-a, a\})$, such that

$$\begin{cases} u(x) < 0, & (\psi^2(x)/2)u''(x) + c - \psi^2(x) = 0; & x \in [0, a), \\ u(x) = 0, & c - \psi^2(x) > 0; & x \in [a, \beta); \end{cases} \quad (4.4)$$

and then to argue that the function u coincides with the minimum expected cost v in (4.1).

In the two paragraphs that follow we shall show that this problem admits a unique solution, which coincides with the value function v of (4.1) and can be computed explicitly.

4.1. Verification

Indeed, if such a function u with the above properties exists, the process

$$N := u(\widehat{X}) - u(x) - \int_0^\cdot \frac{1}{2}(\psi^2 u'')(\widehat{X}(t))dt = \int_0^\cdot (\psi u')(\widehat{X}(t))d\widehat{W}(t)$$

is a local martingale. The function $\psi u'$ is continuous, and supported on the compact interval $[-a, a]$, thus bounded. Therefore, for any stopping time $\tau \in \mathcal{T}$ with $\mathbb{E}(\tau) < \infty$ as in (1.4), we have

$$\mathbb{E}(N^2(\tau)) = \mathbb{E} \int_0^\tau (\psi u')^2(\widehat{X}(t))dt \leq \|\psi u'\|_\infty^2 \mathbb{E}(\tau) < \infty.$$

As a consequence, $N(\cdot \wedge \tau)$ is a square-integrable martingale, and $\mathbb{E}(N(\tau)) = 0$ holds, leading to

$$u(x) = \mathbb{E}[u(\widehat{X}(\tau))] - \mathbb{E} \int_0^\tau \frac{1}{2}(\psi^2 u'')(\widehat{X}(t))dt \leq \mathbb{E} \int_0^\tau (c - \psi^2(\widehat{X}(t)))dt \quad (4.5)$$

on account of the inequalities $u \leq 0$, $(\psi^2/2)u'' + c - \psi^2 \geq 0$ from (4.4).

We repeat now the above reasoning for the stopping time τ_a^* defined in (4.3). This satisfies the property $\mathbb{E}(\tau_a^*) < \infty$, as is checked by considering the diffusion process $\widehat{X}$ of (4.2) on the interval $[-a, a]$ as its state-space, and recalling Proposition 5.5.32 (i) in [7]. For this stopping time, both inequalities summoned to justify the last comparison in (4.5) hold as equalities, and thus so does (4.5) itself:

$$u(x) = \mathbb{E} \int_0^{\tau_a^*} (c - \psi^2(\widehat{X}(t)))dt. \quad (4.6)$$

Now (4.5) and (4.6) show that the stopping time τ_a^* is optimal for the problem of (4.1), among all stopping times with finite expectation. As we argued in the discussion following (1.3), these are the only relevant times for the stopping problem under consideration, and we are done: $u(x) = v(x)$ holds for every $x \in \mathcal{I}_\mu$.

In particular, there can exist at most one solution to the free-boundary problem.

4.2. Construction

For a given constant $a \in (0, \infty)$, the recipe

$$u(x) := 2 \int_x^a \left(\int_y^a \frac{\psi^2(\xi) - c}{\psi^2(\xi)} d\xi \right) dy, \quad 0 \leq x \leq a, \quad (4.7)$$

defines a function that satisfies the equation $(\psi^2(x)/2)u''(x) + c - \psi^2(x) = 0$ in (4.4), as well as the “smooth-fit” conditions” $u(a) = u'(a-) = 0$.

We extend this function by even symmetry to all of $[-a, a]$. For the resulting extension to have the claimed smoothness, we need the condition $u'(0+) = 0$, namely

$$\int_0^a \frac{d\xi}{\psi^2(\xi)} = \frac{a}{c}. \quad (4.8)$$

Now, the function $Q(x) := \int_0^x \psi^{-2}(y)(c - \psi^2(y)) dy$, $0 < x < \beta$, satisfies $Q(0) = 0$, decreases strictly on $(0, \gamma)$, and increases strictly to infinity on (γ, β) . It attains its overall minimum at $x = \gamma$, namely,

$$Q(\gamma) = \int_0^\gamma \frac{c - \psi^2(\xi)}{\psi^2(\xi)} d\xi < 0.$$

Therefore, there exists a unique number $a \in (\gamma, \beta)$ that satisfies $Q(a) = 0$, i.e., (4.8).

With the constant a thus chosen, $c - \psi^2(x) > 0$ holds for every $x \in [a, \beta)$; setting

$$u(x) := 0, \quad x \in (a, \beta) \quad (4.9)$$

and extending again by even symmetry, we obtain a function u defined via (4.7), (4.9) on all of $\mathcal{I}_\mu = (-\beta, \beta)$; this function satisfies all the requirements of the free-boundary problem in (4.4). From what we have proved so far, the function u emerges as the unique solution of this problem, as well as the minimum expected cost in (4.1); that is, $v \equiv u$.

5. Structural properties

In contrast to the two examples just discussed, the typical situation is that the stopping and continuation regions cannot be described so easily. Thus, general methods to determine their structural properties are of considerable interest.

For this purpose, the following monotonicity result will prove useful. It is based on the observation that the term ψ^2 appearing in the integrand in (3.7) coincides with the instantaneous quadratic variation rate of the underlying process $\widehat{X}$ (we extend the function Ψ to be equal to zero outside $[0, \infty) \times \mathcal{I}_\mu$). This suggests a time-change of the martingale $\widehat{X}$ in the manner of Dambis–Dubins–Schwarz (e.g., Theorem 3.4.6 and Problem 3.4.7 in [7]). We follow the construction given by Janson and Tysk in [6], where two diffusion processes in natural scale with the same starting point, but with different dynamics, are constructed as time-changes of the *same* Brownian motion.

Theorem 5.1. *Assume that two distributions μ_i , $i = 1, 2$, are given and that the corresponding variance functions Ψ_i in (2.13) satisfy $\Psi_1(t, x) \geq \Psi_2(t, x)$ for all $(t, x) \in [0, \infty) \times \mathbb{R}$. Then the corresponding value functions v_i , $i = 1, 2$, of (3.7) satisfy $v_1(t, x) \leq v_2(t, x)$ for all $(t, x) \in [0, \infty) \times \mathbb{R}$.*

Proof. It suffices to show that $v_2(0, x) \geq v_1(0, x)$; the case of a general time variable is similar. For $x \in \mathbb{R}$, let B be a one-dimensional Brownian motion with $B(0) = x$. For each $i = 1, 2$, let $\tau_i(\cdot)$, $i = 1, 2$, be the unique stopping time solution of the integral equation

$$\tau_i(t) = \int_0^t \Psi_i^2(\theta, B(\tau_i(\theta))) d\theta, \quad 0 \leq t < \infty,$$

with the terminology and construction developed by Janson and Tysk in [6]. Then the process $X_i(t) := B(\tau_i(t))$, $0 \leq t < \infty$, satisfies the stochastic integral equation

$$X_i(s) = x + \int_0^s \Psi_i(u, X_i(u)) dB_i(u), \quad 0 \leq s < \infty,$$

for some Brownian motion B_i . Consequently, the distribution of $\{X_i(s), s \geq 0\}$ coincides with the distribution of $\{\widehat{X}_i^{(0,x)}(s), s \geq 0\}$, for $i = 1, 2$.

Furthermore, it follows from [6, Lemma 10] that

$$\tau_1(t) \geq \tau_2(t) \tag{5.1}$$

holds for all $t \geq 0$. Now, let γ_2 be a stopping time (of the right-continuous augmentation of the filtration generated by the process X_2) which minimizes

$$\mathbb{E} \left[c\gamma - \int_0^\gamma \Psi_2^2(s, X_2(s)) ds \right] = \mathbb{E}[c\gamma - \tau_2(\gamma)]$$

over all stopping times γ . Define

$$\gamma_1 := \inf\{s \geq 0 : \tau_1(s) > \tau_2(\gamma_2)\}$$

so that $\tau_1(\gamma_1) = \tau_2(\gamma_2)$, and note that (5.1) implies that $\gamma_1 \leq \gamma_2$. Then γ_1 is a stopping time for the process X_1 , though not necessarily an optimal one for the stopping problem under consideration. Consequently,

$$\begin{aligned} v_2(0, x) &= \mathbb{E} \left[c\gamma_2 - \int_0^{\gamma_2} \Psi_2^2(s, X_2(s)) ds \right] = \mathbb{E}[c\gamma_2 - \tau_2(\gamma_2)] \\ &\geq \mathbb{E}[c\gamma_1 - \tau_1(\gamma_1)] = \mathbb{E} \left[c\gamma_1 - \int_0^{\gamma_1} \Psi_1^2(s, X_1(s)) ds \right] \geq v_1(0, x), \end{aligned}$$

which completes the proof. $\square$

As we have seen in the Gaussian and Bernoulli cases, the structure of the stopping region $\mathcal{D}$ depends crucially on the prior distribution μ ; however, we note the following consequence of Theorem 5.1, which provides a very general structural result with respect to the temporal parameter.

Corollary 5.2 (Contracting Continuation Region). *The function $t \mapsto v(t, x)$ is non-decreasing, for every fixed $x \in \mathcal{I}_\mu$. Consequently, the t -section of the stopping region, namely,*

$$\mathcal{D}_t := \{x \in \mathcal{I}_\mu : (t, x) \in \mathcal{D}\},$$

is increasing in time: $\mathcal{D}_{t_1} \subseteq \mathcal{D}_{t_2}$ for $0 \leq t_1 \leq t_2$.

Proof. Consider two time points t_i , $i = 1, 2$, with $t_1 < t_2$, and define $\Psi_i(t, x) = \Psi(t_i + t, x)$ for $(t, x) \in [0, \infty) \times \mathcal{I}_\mu$. Since, on the strength of Proposition 2.5, the function $\Psi(\cdot, x)$ is decreasing, we have $\Psi_1(t, x) \geq \Psi_2(t, x)$ for each $(t, x) \in [0, \infty) \times \mathcal{I}_\mu$. It then follows

from [Theorem 5.1](#) that the corresponding value functions satisfy $v_1(0, x) \leq v_2(0, x)$, which is equivalent to $v(t_1, x) \leq v(t_2, x)$.

Thus $x \in \mathcal{D}_{t_1}$ (i.e., $v(t_1, x) = 0$) leads to $v(t_2, x) = 0$, i.e., to $x \in \mathcal{D}_{t_2}$. $\square$

Corollary 5.3 (*Comparison with the Bernoulli Distribution*). For $\beta > 0$, let $a = a(\beta) > 0$ be the optimal stopping boundary-point for the Bernoulli distribution with support $\{-\beta, \beta\}$ as determined in [Section 4](#). For a “prior” distribution μ , recall the notation of [\(2.7\)](#):

- (i) Assume that $\mathcal{S}_\mu \subseteq [-\beta, \beta]$. Then $\mathcal{C} \subseteq [0, \infty) \times (-a, a)$.
- (ii) Assume that $\mathcal{S}_\mu \subseteq (-\infty, -\beta] \cup [\beta, \infty)$, with $\mathcal{S}_\mu \cap (-\infty, -\beta] \neq \emptyset$ and $\mathcal{S}_\mu \cap [\beta, \infty) \neq \emptyset$. Then $\mathcal{C} \supseteq [0, \infty) \times (-a, a)$.

Proof. It is straightforward to check that among all distributions with support contained in $[-\beta, \beta]$ and expected value $x \in [-\beta, \beta]$, the Bernoulli distribution

$$\frac{\beta - x}{2\beta} \delta_{-\beta} + \frac{\beta + x}{2\beta} \delta_{\beta} \quad (5.2)$$

is the one with the largest variance. Consequently, if $\mathcal{S}_\mu \subseteq [-\beta, \beta]$, then $\Psi(0, x) \leq \beta^2 - x^2 = \psi(x)$. Thus, by [Proposition 2.5](#) we have $\Psi(t, x) \leq \Psi(0, x) \leq \psi(x)$, and (i) follows from [Theorem 5.1](#).

Similarly, among all distributions μ with $\mathcal{S}_\mu \cap (-\beta, \beta) = \emptyset$ and with expected value $x \in (-\beta, \beta)$, the one with the smallest variance is the Bernoulli distribution in [\(5.2\)](#). Consequently, $\Psi(t, x) \geq \beta^2 - x^2$ for all $x \in \mathcal{I}_\mu$, and (ii) follows as above, on account of [Theorem 5.1](#). $\square$

We restrict now attention to sub-classes of prior distributions, for which further structural properties can be derived. We first recall the following well-known result from optimal stopping theory (see, for instance, [\[8, Remark, page 217\]](#)).

Lemma 5.4. Assume that $\Psi^2(t, x) > c$ at some point $(t, x) \in [0, \infty) \times \mathcal{I}_\mu$. Then $(t, x) \in \mathcal{C}$.

Proof. By the continuity of the function Ψ^2 , there exist a real number $\varepsilon > 0$ and a rectangle $\mathcal{R} = [t_1, t_2] \times (a, b) \subseteq [0, \infty) \times \mathcal{I}_\mu$ with $(t, x) \in \mathcal{R}$, and $\Psi^2 - c > \varepsilon$ on $\mathcal{R}$. Denoting by

$$\tau_{\mathcal{R}} := \inf\{s \geq 0 : (t + s, \widehat{X}(t + s)) \notin \mathcal{R}\},$$

we have

$$v(t, x) \leq \mathbb{E} \left[\int_0^{\tau_{\mathcal{R}}} (c - \Psi^2(t + s, \widehat{X}(t + s))) \, ds \right] \leq -\varepsilon \mathbb{E}[\tau_{\mathcal{R}}] < 0,$$

which shows that $(t, x) \in \mathcal{C}$. $\square$

Our next task is to provide conditions under which the stopping region is one-sided. We shall use the notation $\overline{\mathcal{I}}_\mu = \mathcal{I}_\mu \cup \{a, b\}$, where $a = \inf(\mathcal{S}_\mu)$ and $b = \sup(\mathcal{S}_\mu)$ are the (possibly infinite) boundary points of $\mathcal{I}_\mu$, as in [\(2.7\)](#).

Proposition 5.5 (*One-sided Stopping Region*). Assume that, for every fixed time $t \geq 0$, the function $x \mapsto \Psi(t, x)$ (equivalently, the function $y \mapsto H(t, y)$) is non-decreasing. Then the following statements hold:

- (i) There exists a non-decreasing function $b : [0, \infty) \rightarrow \overline{\mathcal{I}}_\mu$ such that the optimal continuation region is of the form

$$\mathcal{C} = \{(t, x) \in [0, \infty) \times \mathcal{I}_\mu : x > b(t)\}.$$

(ii) With $\Psi(t, \infty) := \lim_{x \rightarrow \infty} \Psi(t, x)$, let

$$T := \inf\{t \geq 0 : \Psi^2(t, \infty) \leq c\}.$$

Then $b(t) \in \mathcal{I}_\mu$ for all $t < T$, and $b(t) = \sup(\mathcal{S}_\mu)$ for $t \geq T$, as in (2.7).

(iii) If $x \mapsto \Psi(t, x)$ is strictly increasing for all $t \geq 0$, then the function $b : [0, \infty) \rightarrow \bar{\mathcal{I}}_\mu$ is continuous.

Proof. (i). Without loss of generality, we consider the initial time $t = 0$. We consider two points $(0, x_1)$ and $(0, x_2)$ with $x_1, x_2 \in \mathcal{I}_\mu$ and $x_1 < x_2$. By comparison results for solutions of stochastic integral equations (see, for instance, [9, Theorem IX.3.7]), we obtain $\hat{X}^{0, x_1}(s) \leq \hat{X}^{0, x_2}(s)$ for all times $s \geq 0$. Therefore,

$$\mathbb{E} \left[\int_0^\tau (c - \Psi^2(s, \hat{X}^{0, x_1}(s))) \, ds \right] \geq \mathbb{E} \left[\int_0^\tau (c - \Psi^2(s, \hat{X}^{0, x_2}(s))) \, ds \right]$$

holds for any stopping time τ . Taking the infimum over all stopping times τ yields $v(0, x_1) \geq v(0, x_2)$. In particular, if $v(0, x_1) < 0$, then also $v(0, x_2) < 0$, which shows that $\mathcal{C}$ has the claimed form. The monotonicity of b is immediate from Corollary 5.2.

(ii). With $t \geq T$, we have $\Psi^2(t + s, \cdot) \leq c$ for all $s \geq 0$ by Proposition 2.5 1, and the claim follows from Remark 3.2. For $t < T$, on the other hand, there are points $x \in \mathcal{I}_\mu$ with $\Psi^2(t, x) > c$, so the respective claim follows from Lemma 5.4.

(iii). The upper semi-continuity of v and the monotonicity of b imply $b(t) = b(t+)$ for all $t \geq 0$.

Next assume that $x \mapsto \Psi(t, x)$ is strictly increasing, and that $b(t_1-) < b(t_1)$ for some $t_1 > 0$. Since $(t_1, b(t_1)) \in \mathcal{D}$, it follows from Lemma 5.4 that $\Psi^2(t_1, b(t_1)) \leq c$. Consequently, there exist an $\varepsilon > 0$ and a rectangle $\mathcal{R} = (t_0, t_1) \times (x_1, x_2)$ with $t_0 < t_1$ and $b(t_1-) \leq x_1 < x_2 \leq b(t_1)$ such that $\mathcal{R} \subseteq \mathcal{C}$ and $\Psi^2 \leq c - \varepsilon$ on $\mathcal{R}$. Moreover, $v(t_0, x_1) \leq v < 0$ on $\mathcal{R}$. For a starting point $(t, x) \in \mathcal{R}$, define

$$\tau_{\mathcal{R}}^{t, x} := \inf\{s \geq 0 : (t + s, \hat{X}^{t, x}(t + s)) \notin \mathcal{R}\}$$

to be the first exit time from $\mathcal{R}$. Since $\mathcal{R} \subseteq \mathcal{C}$, the process

$$\begin{aligned} & v(t + (s \wedge \tau_{\mathcal{R}}^{t, x}), \hat{X}^{t, x}(t + (s \wedge \tau_{\mathcal{R}}^{t, x}))) \\ & + \int_0^{s \wedge \tau_{\mathcal{R}}^{t, x}} (c - \Psi^2(t + \theta, \hat{X}^{t, x}(t + \theta))) \, d\theta, \quad s \geq 0, \end{aligned}$$

is a martingale by optimal stopping theory, and

$$\begin{aligned} v(t, x) &= \mathbb{E} \left[v(t + \tau_{\mathcal{R}}^{t, x}, \hat{X}^{t, x}(t + \tau_{\mathcal{R}}^{t, x})) + \int_0^{\tau_{\mathcal{R}}^{t, x}} (c - \Psi^2(t + s, \hat{X}^{t, x}(t + s))) \, ds \right] \\ &\geq \mathbb{E} \left[\mathbf{1}_{\{\tau_{\mathcal{R}}^{t, x} \geq t_1 - t\}} \int_0^{t_1 - t} (c - \Psi^2(t + s, \hat{X}^{t, x}(t + s))) \, ds \right] \\ &\quad + v(t_0, x_1) \mathbb{P}(\tau_{\mathcal{R}}^{t, x} < t_1 - t) \\ &\geq \varepsilon(t_1 - t) \mathbb{P}(\tau_{\mathcal{R}}^{t, x} \geq t_1 - t) + v(t_0, x_1) \mathbb{P}(\tau_{\mathcal{R}}^{t, x} < t_1 - t). \end{aligned}$$

Here the first term is of size $\varepsilon(t_1 - t)$ for t close to t_1 , whereas the probability $\mathbb{P}(\tau_{\mathcal{R}}^{t, x} < t_1 - t)$ is of order $o(t_1 - t)$ as $t \rightarrow t_1$. Consequently, for each $x \in (x_1, x_2)$ there exists t close to t_1 such that $v(t, x) > 0$, which is a contradiction. This proves that $b(t_1-) = b(t_1)$, so b is continuous. $\square$

Remark 5.6. There is an analogue of [Proposition 5.5](#) for problems in which the function $x \mapsto \Psi(t, x)$ is non-increasing for every fixed $t \geq 0$. Arguing exactly as above, this condition implies the existence of a non-increasing boundary b such that

$$\mathcal{C} = \{(t, x) : x < b(t)\}.$$

5.1. A case with a one-sided stopping region: the absolute value of a normal distribution

Let us consider a case where the prior belief is represented by the absolute value of a normally distributed random variable with mean 0 and variance σ^2 , i.e.,

$$\mu(du) = \sqrt{\frac{2}{\pi\sigma^2}} \exp\left\{-\frac{u^2}{2\sigma^2}\right\} du, \quad u \geq 0.$$

Then $\mathcal{I} = (0, \infty)$, and determined computation gives

$$H(t, y) = \frac{\sigma^2}{1 + \sigma^2 t} \left(1 - z \frac{\varphi(z)}{\Phi(z)} - \frac{\varphi^2(z)}{\Phi^2(z)}\right) \Big|_{z=Z(t,y)}, \quad \text{for } Z(t, y) := \frac{\sigma y}{\sqrt{1 + \sigma^2 t}}$$

and

$$\varphi(b) = \frac{1}{\sqrt{2\pi}} \exp\{-b^2/2\}, \quad \Phi(a) = \int_{-\infty}^a \varphi(b) db$$

for the function of [\(2.4\)](#). Note that this function satisfies

$$\lim_{y \rightarrow \infty} H(t, y) = \frac{\sigma^2}{1 + \sigma^2 t}$$

for $t \geq 0$, very much in accordance with [Section 3.2](#). Furthermore, $\Psi(t, \cdot)$ is increasing if and only if $H(t, \cdot)$ is increasing, and

$$DH(t, y) = \frac{\sigma^3 \varphi(z)}{(1 + \sigma^2 t)^{3/2} \Phi(z)} \left(z^2 - 1 + 3z \frac{\varphi(z)}{\Phi(z)} + 2 \frac{\varphi^2(z)}{\Phi^2(z)}\right) \Big|_{z=Z(t,y)}.$$

To see that $DH \geq 0$, we follow an argument from [\[11\]](#). It suffices to check that

$$f(z) := z^2 + 3z \frac{\varphi(z)}{\Phi(z)} + 2 \frac{\varphi^2(z)}{\Phi^2(z)} = \left(z + 2 \frac{\varphi(z)}{\Phi(z)}\right) \left(z + \frac{\varphi(z)}{\Phi(z)}\right) \geq 1.$$

Straightforward calculations give

$$\begin{aligned} f'(z) &= 2 \left(\frac{\varphi(z)}{\Phi(z)} + z\right) \left(1 - z \frac{\varphi(z)}{\Phi(z)} - \frac{\varphi^2(z)}{\Phi^2(z)}\right) + \frac{\varphi(z)}{\Phi(z)} (1 - f(z)) \\ &> \frac{\varphi(z)}{\Phi(z)} (1 - f(z)) \end{aligned} \tag{5.3}$$

at all points z . However, it is clear that $\lim_{z \rightarrow \infty} f(z) = \infty$, and using the expansion

$$\Phi(z) = \frac{\varphi(z)}{-z} \left(1 - \frac{1}{z^2} + o(1/z^2)\right)$$

for $z < 0$ yields $\lim_{z \rightarrow -\infty} f(z) = 1$. Therefore, if there is a finite root of the equation $f(z) = 1$, then there exists a finite z_0 with $f(z_0) \leq 1$ and $f'(z_0) = 0$, which contradicts [\(5.3\)](#). Therefore, $f \geq 1$, so $DH \geq 0$.

It now follows from [Proposition 5.5](#) that the continuation region is one-sided and given by

$$\mathcal{C} = \{(t, x) \in [0, T) \times (0, \infty) : x > b(t)\}$$

for some continuous, non-decreasing function $b : [0, T) \rightarrow [0, \infty)$ with $b(T) := \lim_{t \uparrow T} b(t) = \infty$, where $T = \left(\frac{1}{\sqrt{c}} - \frac{1}{\sigma^2}\right)^+$.

6. Symmetric prior distributions

In this section we consider the special case when μ is symmetric around the origin with $\mathcal{I}_\mu = (-a, a)$ in (2.7), for some $a \in (0, \infty]$. Then the functions Ψ and v are also symmetric around the origin, in the sense that $\Psi(t, x) = \Psi(t, -x)$ and $v(t, x) = v(t, -x)$. Consequently, the optimal stopping problem can be re-written in terms of the reflected diffusion $Z = Z^{t,x} = |\widehat{X}^{t,x}|$ as

$$v(t, x) = \inf_{\tau \in \mathcal{T}} \mathbb{E} \left[\int_0^\tau \left(c - \Psi^2(t+s, Z^{t,x}(t+s)) \right) ds \right], \quad (t, x) \in [0, \infty) \times [0, a). \quad (6.1)$$

Proposition 6.1. *Assume that μ is symmetric around the origin.*

- (i) *Assume that, for every fixed time $t \geq 0$, the function $\Psi(t, \cdot) : [0, a) \rightarrow [0, \infty)$ is non-decreasing. Then there exist a point $t_0 \geq 0$ and a non-decreasing boundary $b : [t_0, \infty) \rightarrow [0, a]$ such that*

$$\mathcal{C} = ([0, t_0) \times \mathcal{I}_\mu) \cup \{(t, x) \in [t_0, \infty) \times \mathcal{I}_\mu : |x| > b(t)\}.$$

- (ii) *Assume that, for every fixed time $t \geq 0$, the function $\Psi(t, \cdot) : [0, a) \rightarrow [0, \infty)$ is non-increasing. Then there exists a non-increasing boundary $b : [0, \infty) \rightarrow [0, a]$ such that*

$$\mathcal{C} = \{(t, x) \in [0, \infty) \times \mathcal{I}_\mu : |x| < b(t)\}.$$

Proof. Without loss of generality, we consider the initial time $t = 0$. For $x \geq 0$, let $(\widetilde{Z}, L)$ be the unique continuous process such that $L(0) = 0$, L is non-decreasing, $Z(0) = x$, $Z(s) \geq 0$ and

$$\begin{cases} d\widetilde{Z}(s) = \Psi(s, \widetilde{Z}(s)) d\widehat{W}(s) + dL(s), \\ \int_0^t \mathbf{1}_{\{\widetilde{Z}(s)=0\}} dL(s) = L(t). \end{cases}$$

Then $\widetilde{Z}$ is the reflected version of $\widehat{X}$, and the processes $\{Z(s), s \geq 0\}$ and $\{\widetilde{Z}(s), s \geq 0\}$ coincide in law. Moreover, by comparison we have that $x_1 \leq x_2$ implies that $\widetilde{Z}^{x_1}(t) \leq \widetilde{Z}^{x_2}(t)$ for all t . The proof then follows the proof of Proposition 5.5. $\square$

6.1. Symmetric Gaussian mixtures

We end the article with a study of the case when the prior is given by a symmetric Gaussian mixture. More precisely, let μ be given by

$$\mu(du) = \frac{1}{2\sigma\sqrt{2\pi}} \left(\exp \left\{ -\frac{(u-m)^2}{2\sigma^2} \right\} + \exp \left\{ -\frac{(u+m)^2}{2\sigma^2} \right\} \right) du$$

with $m \in (0, \infty)$ and $\sigma > 0$, i.e., a mixture of the two Gaussians $N(m, \sigma)$ and $N(-m, \sigma)$. Then

$$F(t, y) = \frac{1}{\sqrt{1+\sigma^2 t}} \left(\exp \left\{ -\frac{1}{2\sigma^2} \left(\frac{(m+\sigma^2 y)^2}{1+\sigma^2 t} - m^2 \right) \right\} \right)$$

$$+ \exp \left\{ -\frac{1}{2\sigma^2} \left(\frac{(-m + \sigma^2 y)^2}{1 + \sigma^2 t} - m^2 \right) \right\},$$

and straightforward calculations yield

$$H(t, y) = \frac{\sigma^2}{1 + \sigma^2 t} + \frac{4m^2}{(1 + \sigma^2 t)^2} \left(\exp \left\{ \frac{my}{1 + \sigma^2 t} \right\} + \exp \left\{ \frac{-my}{1 + \sigma^2 t} \right\} \right)^{-2}.$$

It follows that $\Psi(t, \cdot)$ is decreasing on $[0, \infty)$ and satisfies

$$\Psi(t, 0) = \frac{\sigma^2}{1 + \sigma^2 t} + \frac{m^2}{(1 + \sigma^2 t)^2}, \quad \Psi(t, \infty) = \frac{\sigma^2}{1 + \sigma^2 t}.$$

Consequently, by (ii) of [Proposition 6.1](#), there exists a non-increasing boundary $b : [0, \infty) \rightarrow [0, \infty]$ such that

$$\mathcal{C} = \{(t, x) \in [0, \infty) \times \mathcal{I}_\mu : |x| < b(t)\}.$$

Furthermore, $b(t) = \infty$ for $t \in [0, (c^{-1/2} - \sigma^{-2})^+)$ and $b(t) = 0$ for

$$t \geq \frac{1}{2\sqrt{c}} \left(1 - 2\sigma^{-2}\sqrt{c} + \sqrt{1 + 4m^2\sigma^{-4}c^{1/2}} \right)^+.$$

Dedication

The second author is grateful to his teacher, Dr. Václav E. Beneš, for suggesting this problem to him more than thirteen years ago, and for urging him to make progress on it ever since.

We dedicate this paper to Dr. Beneš on the occasion of his 90th birthday, with affection and respect.

Acknowledgments

We are heavily indebted to Lane Chun Yeung for the interest he took in this work, the many suggestions, comments and corrections he made, and the additional examples he provided.

Many thanks go also to Miguel-Ángel Garrido, Daniel Lacker, Martin Larsson, Johannes Ruf and Walter Schachermayer for the many critical comments to this work that helped us sharpen our thinking; and to the two referees, for spotting errors and helping us clarify and sharpen the exposition of the paper.

References

- [1] A. Bain, D. Crisan, *Fundamentals of Stochastic Filtering*. Stochastic Modeling and Applied Probability, vol. 60, Springer, New York, 2009.
- [2] U. Cetin, A. Novikov, A.N. Shiryaev, Bayesian sequential estimation of a drift of fractional Brownian motion, *Sequential Anal.* 32 (2013) 288–296.
- [3] E. Ekström, J. Vaicenavicius, Optimal liquidation of an asset under drift uncertainty, *SIAM J. Financial Math.* 7 (2016) 357–381.
- [4] N. El Karoui, Les Aspects Probabilistes Du Contrôle Stochastique, in: *Lecture Notes in Mathematics*, vol. 876, 1981, pp. 73–238.
- [5] A.G. Fakeev, Optimal stopping of a Markov process, *Theory Probab. Appl.* 15 (1971) 324–331.
- [6] S. Janson, J. Tysk, Volatility time and properties of option prices, *Ann. Appl. Probab.* 13 (2003) 890–913.
- [7] I. Karatzas, S.E. Shreve, *Brownian Motion and Stochastic Calculus*, second ed., in: *Graduate Texts in Mathematics*, vol. 113, Springer-Verlag, New York, 1991.
- [8] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, sixth ed., Springer-Verlag, New York, 2007.

- [9] D. Revuz, M. Yor, Continuous Martingales and Brownian Motion, third ed., in: Grundlehren der Mathematischen Wissenschaften, vol. 293, Springer-Verlag, Berlin, 1999.
- [10] H. Robbins, D. Siegmund, Statistical tests of power one, and the integral representation of solutions of certain parabolic differential equations, Bull. Inst. Math. Acad. Sin. 1 (1973) 93–120, (in Taipei).
- [11] M.R. Sampford, Some inequalities on Mill's ratio and related functions, Ann. Math. Statist. 24 (1953) 130–132.
- [12] A.N. Shiryaev, Optimal Stopping Rules, Springer-Verlag, New York, 1978.
- [13] P. van Moerbeke, Optimal stopping and free boundary problems, Arch. Ration. Mech. Anal. 60 (1976) 101–148.
- [14] D.V. Widder, Positive temperatures on an infinite rod, Trans. Amer. Math. Soc. 75 (1944) 510–525.