



HAL
open science

Linear-quadratic stochastic delayed control and deep learning resolution

William Lefebvre, Enzo Miller

► **To cite this version:**

William Lefebvre, Enzo Miller. Linear-quadratic stochastic delayed control and deep learning resolution. 2021. hal-03145949v2

HAL Id: hal-03145949

<https://hal.science/hal-03145949v2>

Preprint submitted on 22 Feb 2021 (v2), last revised 23 Feb 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Linear-quadratic stochastic delayed control and deep learning resolution *

William Lefebvre[†] Enzo Miller[‡]

February 22, 2021

Abstract

We consider a class of stochastic control problems with a delayed control, both in drift and diffusion, of the type $dX_t = \alpha_{t-d}(bdt + \sigma dW_t)$. We provide a new characterization of the solution in terms of a set of Riccati partial differential equations. Existence and uniqueness are obtained under a sufficient condition expressed directly as a relation between the horizon T and the quantity $d(b/\sigma)^2$. Furthermore, a deep learning scheme¹ is designed and used to illustrate the effect of delay on the Markowitz portfolio allocation problem with execution delay.

Keywords: Linear-quadratic stochastic control; delay; Riccati PDEs; Markowitz portfolio allocation.

MSC Classification: 93E20, 60H10, 34K50.

Contents

1	Introduction	2
2	Formulation of the problem and heuristic approach	4

*We would like to thank Salvatore Federico and Huyên Pham for their useful comments and remarks that helped to improve this article.

[†]BNP Paribas Global Markets, Université de Paris and Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation (LPSM, UMR CNRS 8001), Building Sophie Germain, Avenue de France, 75013 Paris, [wlefevre at lpsm.paris](mailto:wlefevre@lpsm.paris)

[‡]Université de Paris and Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation (LPSM, UMR CNRS 8001), Building Sophie Germain, Avenue de France, 75013 Paris, [enzo.miller at polytechnique.org](mailto:enzo.miller@polytechnique.org)

¹The code is available in a [IPython notebook](#).

3	Verification and existence results	9
4	Deep learning scheme	15
4.1	A quick reminder of PINNs and Deep Galerkin method for PDEs	15
4.2	A tailor-made algorithm for Riccati partial differential equations	15
5	Applications to mean-variance portfolio selection with execution delay	18
5.1	One asset with delay	18
5.2	One asset with delay and one without	22
A	Proof of Proposition 3.3	31
A.1	Slice $t \in [T - d, T]$, initialization	31
A.2	Slice $[T - 2d, T - d]$	32
A.3	From slice $[T - nd, T]$ to $[T - (n + 1)d, T]$	36

1 Introduction

The control of systems whose dynamic contains delays on the state and/or control has attracted the attention of the optimization and engineering communities in the last decades due to its wide variety of applications, allowing to tackle problems where the past of a system influences its present or where an agent controls a system with a latency. As a non-exhaustive list of applications we may cite the following papers, classified by their applications domain: Engineering (Tian and Gao (1999), Huzmezan et al. (2002)); Advertising (Sethi (1974), Pauwels (1977), Gozzi et al. (2005), Gozzi et al. (2009)); Learning by doing with memory effect (d’Albis et al. (2012)); Growth model with lags between investment decision and project completion (Asea and Zak (1999), Hall et al. (1977), Jarlebring and Damm (2007), Bambi (2008), Bambi et al. (2012)); Investment (Tsoukalas (2011), Kydland and Prescott (1982)). More recently, the introduction of delayed control together with mean-field effects was studied (Carmona et al. (2018), Fouque and Zhang (2019)) and new machine learning methods have been designed to numerically solve stochastic control problems with delay (Han and Hu (2021)). We also refer to the monograph Sipahi et al. (2011) to find literature on the various effects of delays on traffic flow modelling, chemical processes, population dynamics, supply chain, etc.

In the optimal control community, two main approaches have emerged: the *structural state method* and the *extended state method*, and we refer to Bensoussan et al. (2007, Part II, Chapter 3) for the study of the latter in the deterministic case and Fabbri and Federico (2014) for the structural state approach in the stochastic case. Let us also mention the paper by Fabbri and Federico (2014) for an overview and exhaustive list of references.

In this paper, we aim at studying the challenging case where there is a delayed control both in the drift and volatility. Except in Fabbri and Federico (2014), this situation is not treated theoretically nor numerically in the references above. The main difficulty

comes from the fact that the natural formulation of a control problem with delayed control involves a *boundary control problem*. Indeed, assume for instance that X denotes a state variable following the simple dynamic $\dot{X}_t = \alpha_{t-d}$, where α denotes the control. For any time t and index $s \in [-d, 0]$, set $u_t(s) = \alpha_{t+s}$, the *memory* of the control α . Then, note that $\partial_t u_t(s) = \partial_s u_t(s)$ and $u_t(0) = \alpha_t$. Thus, the natural infinite dimensional formulation of the controlled system is

$$\begin{aligned} \text{State eq. on } (X, u) & \begin{cases} \dot{X}_t = \mathcal{M}u_t \\ (\partial_t - \partial_s)u_t(s) = 0, \end{cases} \\ \text{Boundary constraint} & \{ \mathcal{B}u_t = \alpha_t, \\ \text{Initial conditions} & \{ u_0(s) = \gamma_s, \quad X_0 = x, \end{cases} \end{aligned}$$

where $t, s \in [0, T] \times [-d, 0]$, $\mathcal{M}u := u(-d)$, $\mathcal{B}u := u(0)$ and γ is the initial value of the control over $[-d, 0]$. Consequently, any delayed controlled problem where the delay appears in the control variable can be recast as a boundary control problem whose geometry is parametrized by the delay d , see Figure 1.

Main contributions. Our goal is to shed some lights on the difficulty related to delayed control on the volatility and to provide a practical and simple tool for designing a numerical scheme practitioners can play with. In this paper, we study the most simple linear-quadratic control problem with delayed control both in drift and volatility. The optimal feedback control and the value function are given in terms of Riccati partial differential equations and the extended state $(x, u) \in H = \mathbb{R} \times L^2([0, T], \mathbb{R})$, where x denotes the position and $s \in [-d, 0] \mapsto u(s)$ the memory of the control. The existence and uniqueness of these latter are proven under a condition, emerging from the delay feature, involving the drift b , the volatility σ , the delay d and the horizon T . Finally, we adopt a deep learning approach in the spirit of the papers by [Raissi et al. \(2019\)](#) (Physics Informed Neural Network) and [Sirignano and Spiliopoulos \(2018\)](#) (Deep Galerkin) to propose a numerical scheme. Our results are illustrated on the celebrated Markowitz portfolio allocation problem where we take into account execution delay. We believe the semi-explicit resolution of infinite dimensional control problem by means of deep learning method will open the door to several interesting applications such as quick simulations of richer models, precise benchmarking of *reinforcement learning* algorithms, etc.

Outline of the paper. The rest of the paper is organized as follows: In Section 2 we formulate the stochastic delayed control problem and derive an heuristic approach through a lifting in an infinite dimensional space, namely the *extended state space* in the spirit of [Ichikawa \(1982\)](#), but without the use of semi-group theory. We state in Section 3 a verification theorem and prove existence and uniqueness results for the Riccati PDEs. A deep learning based numerical scheme with two applications on Markowitz portfolio allocation

is given in Section 4, with a detailed analysis of the effect of the delay feature on the allocation strategy.

Notations.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ satisfying the usual conditions and $a < b$ two real numbers, we denote by

$$L^2([a, b], \mathbb{R}) = \left\{ Y : [a, b] \mapsto \mathbb{R}, \text{ s.t. } \int_a^b |Y_t|^2 dt < \infty \right\},$$

$$L^2_{\mathbb{F}}([a, b], \mathbb{R}) = \left\{ Y : \Omega \times [a, b] \mapsto \mathbb{R}, \mathbb{F} - \text{prog. measurable s.t. } \mathbb{E} \left[\int_a^b |Y_t|^2 dt \right] < \infty \right\}.$$

Here $|\cdot|$ denotes the Euclidean norm on \mathbb{R} or \mathbb{R}^d , and $H = \mathbb{R} \times L^2([0, T], \mathbb{R})$ denotes the *extended state space* endowed with the scalar product $\langle x, y \rangle_H = x_0 y_0 + \int_{-d}^0 x_1(s) y_1(s) ds$. For any $z = (x, u) \in H$, we use the notation $z_0 = x$ and $z_1 = u$.

2 Formulation of the problem and heuristic approach

Let $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t \leq 0}, \mathbb{P})$ be a complete filtered probability space on which a real-valued Brownian motion $(W_t)_{t \leq 0}$ is defined and consider the simple system defined on $[0, T]$ by the following dynamics

$$\begin{cases} dX_t^\alpha = \alpha_{t-d} (bdt + \sigma dW_t), & 0 \leq t \leq T, \\ X_0 = x, \quad \alpha_s = \gamma(s), & s \in [-d, 0], \end{cases} \quad (2.1)$$

endowed with the cost functional

$$J(\alpha) = \mathbb{E} [(X_T^\alpha)^2], \quad (2.2)$$

where $\gamma \in L^2([-d, 0], \mathbb{R})$ and α models the control chosen in the set of admissible strategies \mathcal{A} :

$$\mathcal{A} = \left\{ \alpha \in L^2_{\mathbb{F}}([0, T], \mathbb{R}) \text{ such that (2.1) has a solution satisfying } \mathbb{E} \left[\sup_{t \leq T} |X_t^\alpha|^2 \right] < \infty \right\}.$$

For any $0 \leq a < b \leq T$, we also define the set $\mathcal{A}_{a,b}$ as the restriction of \mathcal{A} to $L^2_{\mathbb{F}}([a, b], \mathbb{R})$.

Remark 2.1. *At this point, we may expect a priori that the optimization problem (2.1)-(2.2) admits an optimizer provided $\sigma \neq 0$, even if the control is not directly penalized. The intuition behind this a priori belief is that, the more α is aggressive in bringing X to 0, the*

more the variance of X increases due to the diffusion term. It is the case in the classical LQ stochastic optimization problem with controlled volatility such as

$$\begin{aligned} dX_t^\alpha &= \alpha_t(bdt + \sigma dW_t), & t \leq T, \\ X_0 &= x, \\ J(\alpha) &= \mathbb{E}[(X_T^\alpha)^2], \end{aligned}$$

where the optimal control reads $\alpha_t^* = -\frac{b}{\sigma^2}X_t^*$ and the value function $V_t = e^{(t-T)\frac{b^2}{\sigma^2}}$. A surprising finding in our paper is the necessity for a more restricting condition on the diffusion coefficient due to the delay feature, see Proposition 3.3.

Remark 2.2. In the rest of the paper we focus on the one dimensional case with delayed control both in drift and volatility which features the main difficulties related to the presence of the delay. Although Proposition 3.3 concerning the existence and uniqueness of a Riccati-PDE system does not directly extend to the multidimensional case, the verification Theorem can easily be adapted to the multidimensional case with delayed state and control.

The first step consists in lifting the dynamics in the infinite dimensional Hilbert space $H = \mathbb{R} \times L^2([0, T], \mathbb{R})$, where the system is naturally Markovian. To do so, denote $u_t(s) = \alpha_{t+s}$ for any $t \leq T$ and $s \in [-d, 0]$, a transport of the control. The dynamics (2.1) then reads

$$\begin{cases} dZ_t^\alpha = AZ_t^\alpha dt + BZ_t^\alpha dW_t + Cd\alpha_t, & 0 \leq t \leq T, \\ Z_0 = (x, \gamma), \end{cases} \quad (2.3)$$

where Z^α is defined as the $H = \mathbb{R} \times L^2([0, T], \mathbb{R})$ -valued random process $Z_t^\alpha = (X_t^\alpha, u_t(\cdot))$ and

$$A = \begin{pmatrix} 0 & b\delta_{-d} \\ 0 & \partial_s \end{pmatrix}, \quad B = \begin{pmatrix} 0 & \sigma\delta_{-d} \\ 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 \\ 1_{0(\cdot)} \end{pmatrix}.$$

Let V be the value function

$$V(t, z) = V(t, (x, u)) = \inf_{\alpha \in \mathcal{A}_{t,T}} \mathbb{E}[(Z_T^\alpha)_0^2] = \inf_{\alpha \in \mathcal{A}_{t,T}} \mathbb{E}[(X_T^\alpha)^2], \quad z \in H,$$

where Z^α denotes the solution to (2.3) starting from $z = (x, u)$ at time t . Then, assuming

$V \in C^{1,2}([0, T] \times L^2([-d, 0]), \mathbb{R})$, the dynamic programming principle reads

$$\begin{aligned}
V(t, z) &= \inf_{\alpha \in \mathcal{A}_{t, t+h}} \mathbb{E}[V(t+h, Z_{t+h}^\alpha)] \\
&= \inf_{\alpha \in \mathcal{A}_{t, t+h}} \mathbb{E} \left[V(t, z) + \int_t^{t+h} \partial_t V(s, Z_s^\alpha) ds + \int_t^{t+h} \partial_z V(s, Z_s^\alpha) dZ_s^\alpha \right. \\
&\quad \left. + \frac{1}{2} \int_t^{t+h} \partial_z^2 V(s, Z_s^\alpha) d\langle Z^\alpha \rangle_s \right]. \\
&= \inf_{\alpha \in \mathcal{A}_{t, t+h}} \mathbb{E} \left[V(t, z) + \int_t^{t+h} \partial_t V(s, Z_s^\alpha) ds + \int_t^{t+h} \partial_z V(s, Z_s^\alpha) (AZ_s^\alpha ds + Cd\alpha_s) \right. \\
&\quad \left. + \frac{1}{2} \int_t^{t+h} \partial_z^2 V(s, Z_s^\alpha) d\langle Z^\alpha \rangle_s \right],
\end{aligned}$$

Note that $1_0(\cdot) = 0_{L^2}$. As a result, simplifying by $V(t, z)$, dividing by h and letting $h \rightarrow 0$ yields (informally) the Hamilton-Jacobi equation

$$\begin{aligned}
\partial_t V + \inf_{\alpha \in \mathbb{R}} \{ \partial_z V A z + \partial_z^2 V (Bz \otimes Bz) \} &= 0, \quad t \leq T, \quad z \in L^2([-d, 0], \mathbb{R}), \\
V(T, z) &= z_0^2.
\end{aligned} \tag{2.4}$$

Recall that in equation (2.4), we have $z_1(0) = u(0) = \alpha$. Let us now assume that the value function V is of the following form

$$V(t, z) = \langle P_t z, z \rangle_H,$$

where $P \in C([0, T], \mathcal{L}(H, H))$ is a self-adjoint bounded positive operator valued function of the form

$$P_t : (x, \gamma(\cdot)) \mapsto \begin{pmatrix} P_{11}(t)x + \int_{-d}^0 P_{12}(t, s)\gamma(s)ds \\ P_{12}(t, \cdot)x + P_{22}(t, \cdot)\gamma(\cdot) + \int_{-d}^0 P_{22}(t, \cdot, s)\gamma(s)ds \end{pmatrix}.$$

Thus, for any $z = (x, u) \in H$ such that $u(0) = \alpha$ and $t \leq T$, equation (2.4) reduces to

$$\langle \dot{P}_t z, z \rangle_H + \inf_{\alpha \in \mathbb{R}} \{ \langle P_t A z, z \rangle_H + \langle P_t z, A z \rangle_H + \langle P_t B z, B z \rangle_H \} = 0. \tag{2.5}$$

Furthermore, using the boundary condition $u(0) = \alpha$ together with integration by part, we

have

$$\begin{aligned}
\langle P_t z, Az \rangle_H &= (P_t z)_0 (Az)_0 + \int_{-d}^0 (P_t z)_1(s) (Az)_1(s) ds \\
&= bu(-d) \left(P_{11}(t)x + \int_{-d}^0 P_{12}(t, s)u(s) ds \right) + \alpha x P_{12}(t, 0) - u(-d)x P_{12}(t, -d) \\
&\quad - x \int_{-d}^0 \partial_s P_{12}(t, s)u(s) ds + \alpha \int_{-d}^0 P_{22}(t, 0, s)u(s) ds \\
&\quad - u(-d) \int_{-d}^0 P_{22}(t, -d, s)u(s) ds - \int_{-d}^0 \int_{-d}^0 \partial_s P_{22}(t, s, r)u(s)u(r) ds dr \\
&\quad + \alpha^2 P_{\hat{2}2}(t, 0) - u(-d)^2 P_{\hat{2}2}(t, -d) - \int_{-d}^0 \partial_s P_{\hat{2}2}(t, s) ds,
\end{aligned} \tag{2.6}$$

and

$$\langle P_t Bz, Bz \rangle_H = \sigma^2 P_{11}(t)u(-d)^2. \tag{2.7}$$

Remark 2.3. In (2.6), along with the integration by part, formulas such as $u\partial_s u = \partial_s u^2$ were (formally) used. However, as it appears in the verification Theorem 3.1, the feedback optimal control obtained is as regular as the controlled process X^α and thus as regular as the Brownian motion W . This is why our approach is only heuristic and justifies the need for the verification Theorem 3.1.

As a consequence, the minimizer of the Hamiltonian in (2.5) reads

$$\alpha^*(t, z) = -\frac{1}{P_{\hat{2}2}(t, 0)} \left(x P_{12}(t, 0) + \int_{-d}^0 P_{22}(t, 0, s)u(s) ds \right). \tag{2.8}$$

Remark 2.4. Note that when $d \rightarrow 0$, then $\alpha^*(t, z) \rightarrow -\frac{b}{\sigma^2}x$ which agrees with the optimal strategy in the undelayed case.

Combining (2.5), (2.6) and (2.7) yields the set of Riccati partial differential equations

$$\begin{aligned}
\dot{P}_{11}(t) &= \frac{P_{12}(t, 0)^2}{P_{\hat{2}2}(t, 0)}, & (\partial_t - \partial_s)(P_{12})(t, s) &= \frac{P_{12}(t, 0)P_{22}(t, s, 0)}{P_{\hat{2}2}(t, 0)}, \\
(\partial_t - \partial_s)(P_{\hat{2}2})(t, s) &= 0, & (\partial_t - \partial_s - \partial_r)(P_{22})(t, s, r) &= \frac{P_{22}(t, s, 0)P_{22}(t, 0, r)}{P_{\hat{2}2}(t, 0)},
\end{aligned} \tag{2.9}$$

accompanied by the boundary conditions

$$\begin{aligned}
P_{12}(t, -d) &= bP_{11}(t), & P_{\hat{2}2}(t, -d) &= \sigma^2 P_{11}(t), \\
P_{22}(t, s, -d) &= bP_{12}(t, s), & P_{22}(t, -d, r) &= bP_{12}(t, r),
\end{aligned} \tag{2.10}$$

and the final conditions

$$P_{11}(T) = 1, \quad P_{12}(T, s) = P_{\hat{2}2}(T, s) = P_{22}(T, s, r) = 0, \quad (2.11)$$

for almost every $s, r \in [-d, 0]$.

Remark 2.5. *Looking at the expression (2.8), we can already guess some effects of the existence of a delay on the optimal strategy. Indeed, from (2.10) one notes that $P_{12} \approx b$, $P_{\hat{2}2} \approx \sigma^2$, $P_{22} \approx b^2$, and we may write*

$$\alpha \approx \frac{-1}{\sigma^2}(bx + db^2\alpha) \approx \frac{-bx}{\sigma^2(1 + d(b/\sigma)^2)}.$$

In Section 4, we illustrate numerically the various effects of the delayed control through two examples of Markowitz portfolio allocation with execution delay.

Note that due to the existence of the delay, the value function is independent of the control chosen after $T - d$, so that $P_{12}(t, s) = P_{\hat{2}2}(t, s) = P_{22}(t, s, r) = 0$ whenever $t + s \geq T - d$ or $t + r \geq T - d$. Similarly, the optimal control defined in (2.8) is ill defined on $[T - d, T]$ so we decide to set to zero the control after time $T - d$ and rewrite

$$\alpha^*(t, z) = -\frac{1_{t \leq T-d}}{P_{\hat{2}2}(t, 0)} \left(xP_{12}(t, 0) + \int_{-d}^0 P_{22}(t, 0, s)u(s)ds \right). \quad (2.12)$$

Thus, to make sense of the set of Ricatti partial differential equations (2.9)-(2.10)-(2.11) and the optimal control (2.12), we adopt the convention $0^2/0 = 0$ and define the concept of solution as follows

Definition 2.6. *A 4-uplets $P = (P_{11}, P_{12}, P_{\hat{2}2}, P_{22})$ is said to be a solution to (2.9)-(2.10)-(2.11) if $P_{11} : [0, T] \mapsto \mathbb{R}$, $P_{22}, P_{\hat{2}2} : [0, T] \times [-d, 0] \mapsto \mathbb{R}$ and $P_{12} : [0, T] \times [-d, 0]^2 \mapsto \mathbb{R}$ are piecewise absolutely continuous functions satisfying (2.9)-(2.10)-(2.11) with $P_{\hat{2}2}(t) > 0$ for any $t < T - d$.*

The reason we chose *piecewise* absolutely continuous functions as our set of functions is because we expect the kernel P to be discontinuous. To illustrate this consideration, cut the domain \mathcal{D} into three pieces $\mathcal{D} = [0, T] \times [-d, 0]^2 = \mathcal{D}_a \cup \mathcal{D}_b \cup \mathcal{D}_c$ as represented in Figure 1, with

$$\begin{aligned} \mathcal{D}_a &= [0, T - d] \times [-d, 0]^2, \\ \mathcal{D}_b &= \{(t, s, r) \in \mathcal{D} \text{ s.t. } t > T - d, \quad t + s \vee r < T - d\}, \\ \mathcal{D}_c &= \{(t, s, r) \in \mathcal{D} \text{ s.t. } t > T - d, \quad t + s \vee r \geq T - d\} \end{aligned}$$

and note that, necessarily, P_{12} , $P_{\hat{2}2}$ and P_{22} are null on \mathcal{D}_c but not on the remaining domain, see also the numerical simulations in Figure 5.

In the next section, we provide a proof of the existence and uniqueness of system (2.9)-(2.10)-(2.11), and a verification theorem yielding rigorously the optimal control and value of (2.1)-(2.2).

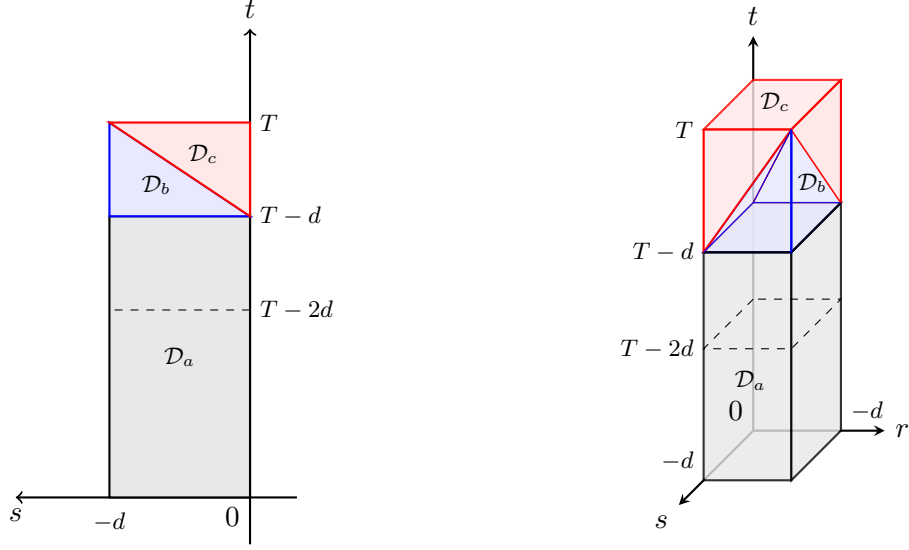


Figure 1: Left: Cross section of \mathcal{D} along $r = 0$. Right: full domain $\mathcal{D} = \mathcal{D}_a \cup \mathcal{D}_b \cup \mathcal{D}_c$.

3 Verification and existence results

In this section, we establish a verification result for the optimization problem (2.1)-(2.2).

Theorem 3.1 (Verification Theorem). *Assume that*

- (i) *There exists a solution P to (2.9)-(2.10)-(2.11) in the sense of Definition 2.6,*
- (ii) *The control strategy defined as*

$$\alpha_t^* = \frac{-1_{t \leq T-d}}{P_{22}(t, 0)} \left\{ X_t^{\alpha^*} P_{11}(t, 0) + \int_{t-d}^t P_{22}(t, 0, s-t) \alpha_s^* ds \right\}. \quad (3.1)$$

where X^{α^*} denotes the controlled state is an admissible control.

Then the optimization problem (2.1)-(2.2) admits (3.1) as an optimal feedback control. Furthermore, for $z = (x, \gamma) \in H$, the value is given by

$$\begin{aligned} V(z) &= P_{11}(0)x^2 + 2x \int_{-d}^0 P_{12}(0, s) \gamma_s ds + \int_{-d}^0 P_{22}(0, s) \gamma_s^2 ds \\ &\quad + \int_{[-d, 0]^2} \gamma_s \gamma_u P_{22}(0, s, r) ds dr \\ &= \langle P_0 z, z \rangle_H. \end{aligned} \quad (3.2)$$

Proof. The proof is a basic application of the martingale optimality principle, see [El Karoui \(1981\)](#). Let $\alpha \in \mathcal{A}$ and define

$$\begin{aligned} V_t^\alpha &= P_{11}(t)(X_t^\alpha)^2 + 2x \int_{t-d}^t P_{12}(t, s-t)\alpha_s ds + \int_{t-d}^t P_{22}(t, s-t)\alpha_s^2 ds \\ &\quad + \int_{[t-d, t]^2} P_{22}(t, s-t, r-t)\alpha_s \alpha_r ds dr \\ &= \langle P_t Z_t^\alpha, Z_t^\alpha \rangle_H. \end{aligned} \tag{3.3}$$

An application of Itô's formula to (3.3) combined with differentiation under the integral symbol, authorized by the assumed boundedness of P and its derivatives, yield

$$\begin{aligned} dV_t^\alpha &= \left\{ \mathbf{1}_t (X_t^\alpha)^2 + 2 \left(\mathbf{2}_t X_t^\alpha \alpha_t + \mathbf{3}_t X_t^\alpha \alpha_{t-d} + X_t^\alpha \int_{t-d}^t \mathbf{4}_t(s) \alpha_s ds \right) \right. \\ &\quad + \mathbf{5}_t \alpha_{t-d}^2 + \alpha_{t-d} \mathbf{6}_t(s) \int_{t-d}^t \alpha_s ds + \alpha_t \int_{t-d}^t \mathbf{7}_t(s) \alpha_s ds \\ &\quad \left. + \int_{t-d}^t \mathbf{8}_t(s) \alpha_s^2 ds + \int_{[t-d, t]^2} \mathbf{9}_t(s, u) \alpha_s \alpha_u ds du \right\} dt + Z_t^\alpha dW_t, \end{aligned}$$

where we have set

$$\begin{aligned} \mathbf{1}_t &= \dot{P}_{11}(t) & \mathbf{2}_t &= P_{12}(t, 0) \\ \mathbf{3}_t &= 2bP_{11}(t) - P_{12}(t, -d) & \mathbf{4}_t(s) &= (\partial_t - \partial_s)(E_2)(t, s-t) \\ \mathbf{5}_t &= \sigma^2 P_{11}(t) - P_{22}(t, -d) & \mathbf{6}_t(s) &= (bP_{12}(t, \cdot) - P_{22}(t, \cdot, -d))(s-t) \\ \mathbf{7}_t(s) &= P_{22}(t, 0, s-t) & \mathbf{8}_t(s) &= (\partial_t - \partial_s)(P_{22})(t, s-t) \\ \mathbf{9}_t(s, r) &= (\partial_t - \partial_s - \partial_r)(P_{22})(t, s-t, r-t), \end{aligned}$$

and

$$Z_t^\alpha = 2\sigma\alpha_{t-d} \left(X_t^\alpha P_{11}(t) + \int_{t-d}^t \alpha_s P_{12}(t, s-t) ds \right).$$

Then, using the set of constraints (2.9)-(2.10)-(2.11) together with (3.1) and a completion of the square in α yield

$$dV_t^\alpha = \left(P_{22}(t, 0) (\alpha_t - \mathcal{T}(\alpha)_t)^2 \right) dt + Z_t^\alpha dW_t,$$

where $\mathcal{T}(\alpha)$ is defined as

$$\mathcal{T}(\alpha)_t = -\frac{1_{t \leq T-d}}{P_{22}(t, 0)} \left\{ X_t^\alpha P_{12}(t, 0) + \int_{t-d}^t \alpha_s P_{22}(t, 0, s-t) ds \right\}, \quad t \leq T.$$

Note that since the kernels P_i 's are bounded and the control α^* is assumed to be admissible, $\alpha^* \in L_{\mathbb{F}}^2([0, T], \mathbb{R})$, X^α is continuous and the stochastic integral $\int_0^\cdot Z_s^\alpha dW_s$ is well posed. Furthermore it is a local martingale. Thus, there exists a localizing increasing sequence of stopping times $\{\tau_k\}_{k \geq 1}$ converging to T such that $\int_0^{\cdot \wedge \tau_k} Z_s^\alpha dW_s$ is a martingale for every $k \geq 1$. Then, for any $k \geq 1$

$$\mathbb{E} [V_{T \wedge \tau_k}^\alpha] = V_0^\alpha + \mathbb{E} \left[\int_0^{T \wedge \tau_k} P_{\hat{2}2}(s, 0) (\alpha_s - \mathcal{T}(\alpha)_s)^2 ds \right].$$

Note that $t \mapsto V_t^\alpha$ is continuous since P is bounded and $\alpha \in L_{\mathbb{F}}^2([-d, T], \mathbb{R})$. Thus, an application of the dominated convergence theorem on the left term (recall that $\mathbb{E} [\sup_{t \leq T} |X_t^\alpha|^2] < \infty$, $\alpha \in L_{\mathbb{F}}^2([-d, T], \mathbb{R})$ as $\alpha \in \mathcal{A}$) combined with the monotone convergence theorem on the right term yields, as $k \rightarrow \infty$

$$\mathbb{E} [V_T^\alpha] = \mathbb{E} [(X_T^\alpha)^2] = V_0^\alpha + \mathbb{E} \left[\int_0^T P_{\hat{2}2}(s, 0) (\alpha_s - \mathcal{T}(\alpha)_s)^2 ds \right].$$

Note that here we used the assumption $P_{\hat{2}2}(t, 0) \geq 0$ on $[0, T]$. Since $P_{\hat{2}2}$ is non-negative, we obtain that the optimal strategy is given by α^* and that the optimal value equals (3.2). \square

Proposition 3.2. *Assume that there exists a bounded 4-uplets P solution to (2.9)-(2.10)-(2.11) in the sense of Definition 2.6. Then (3.1) defines an admissible control.*

Proof. Let $\gamma \in L^2([-d, 0], \mathbb{R})$. To prove the claim, note that it suffices to show that the equation

$$\begin{cases} \alpha_t &= \frac{-1_{t \leq T-d}}{P_{\hat{2}2}(t, 0)} \left\{ P_{12}(t, 0) \left(x + \int_0^t \alpha_{s-d} (b ds + \sigma dW_s) \right) + \int_{t-d}^t \alpha_s P_{22}(t, 0, s-t) ds \right\}, \\ \alpha_s &= \gamma_s, \quad s \in [-d, 0], \end{cases} \quad (3.4)$$

admits a solution in $L_{\mathbb{F}}^2([0, T], \mathbb{R})$ and that the process X , then defined as

$$X_t = x + \int_0^t \alpha_{s-d} (b ds + \sigma dW_s), \quad t \leq T, \quad (3.5)$$

satisfies $\mathbb{E} [\sup_{t \leq T} |X_t|^2] < \infty$. To prove the first point, consider the linear operator ϕ on $L_{\mathbb{F}}^2([0, T], \mathbb{R})$ defined as, for any $a \in L_{\mathbb{F}}^2([0, T], \mathbb{R})$

$$\phi(a)_t = \frac{-1}{P_{\hat{2}2}(t, 0)} \left\{ P_{12}(t, 0) \left(x + \int_0^t \hat{a}_{s-d} (b ds + \sigma dW_s) \right) + \int_{t-d}^t \hat{a}_s P_{22}(t, 0, s-t) ds \right\},$$

where $\hat{a}_t = 1_{t \leq 0} \gamma t + 1_{t > 0} a_t$. For $\lambda \leq 0$, we endow $L_{\mathbb{F}}^2([0, T], \mathbb{R})$ with the norm $\|a\|_{2, \lambda} = \sqrt{\int_0^T e^{-\lambda s} |a_s|^2 ds}$. Then, for any $a, a' \in L_{\mathbb{F}}^2([0, T], \mathbb{R})$, we have

$$\begin{aligned} \|\phi(a) - \phi(a')\|_{2, \lambda}^2 &= \mathbb{E} \left[\int_0^T e^{-\lambda s} |\phi(a)_s - \phi(a')_s|^2 ds \right] \\ &\leq 2(\mathbf{I} + \mathbf{II}). \end{aligned} \quad (3.6)$$

An application of Jensen's inequality on the normalised measure $\frac{dr}{s}$ on $[0, s]$, combined with $\frac{1 - e^{-\lambda(T-r)}}{\lambda} \leq (1 \vee T)(1 \wedge \lambda^{-1})$, and the Burkholder-Davis-Gundy inequality lead to

$$\begin{aligned} \mathbf{I} &\leq \mathbb{E} \left[\int_0^T e^{-\lambda s} \left| \frac{P_{12}(s, 0)}{P_{22}(s, 0)} \int_0^s (\hat{a}_{r-d} - \hat{a}'_{r-d})(bdr + \sigma dW_r) \right|^2 ds \right] \\ &\leq \sup_{s \leq T} \left| \frac{P_{12}(s, 0)}{P_{22}(s, 0)} \right|^2 \left\{ \mathbb{E} \left[\int_0^T e^{-\lambda s} \left(\int_0^s (\hat{a}_{r-d} - \hat{a}'_{r-d}) bdr \right)^2 ds \right] \right. \\ &\quad \left. + \mathbb{E} \left[\int_0^T e^{-\lambda s} \left(\int_0^s (\hat{a}_{r-d} - \hat{a}'_{r-d}) \sigma dW_r \right)^2 ds \right] \right\} \\ &\leq c(1 \wedge \lambda^{-1}) \mathbb{E} \left[\int_0^T e^{-\lambda r} (\hat{a}_{r-d} - \hat{a}'_{r-d})^2 dr \right], \end{aligned}$$

where $c > 0$ depends only on b, σ, T and $\sup_{s \leq T} \left| \frac{P_{12}(s, 0)}{P_{22}(s, 0)} \right|$. Furthermore, we have

$$\begin{aligned} \mathbf{II} &\leq \mathbb{E} \left[\int_0^T e^{-\lambda s} \left| \frac{1}{P_{22}(s, 0)} \int_{s-d}^s (\hat{a}_r - \hat{a}'_r) P_{22}(r, 0, s-r) dr \right|^2 ds \right] \\ &\leq \sup_{\substack{s \leq T \\ r \in [-d, 0]}} \left| \frac{P_{22}(s, 0, r)}{P_{22}(s, 0)} \right|^2 \mathbb{E} \left[\int_0^T e^{-\lambda s} \left| \int_0^s (\hat{a}_r - \hat{a}'_r) dr \right|^2 ds \right] \\ &\leq \hat{c}(1 \wedge \lambda^{-1}) \mathbb{E} \left[\int_0^T e^{-\lambda r} (a_r - \hat{a}'_r)^2 dr \right], \end{aligned}$$

where $\hat{c} > 0$ depends only on T and $\sup_{\substack{s \leq T \\ r \in [-d, 0]}} \left| \frac{P_{22}(s, 0, r)}{P_{22}(s, 0)} \right|$. Consequently, (3.6) reduces to

$$\|\phi(a) - \phi(a')\|_{2, \lambda}^2 \leq (c + \hat{c})(1 \wedge \lambda^{-1}) \|a - a'\|_{2, \lambda}^2.$$

As a result, for λ large enough, ϕ is a contraction on the Banach space $(L_{\mathbb{F}}^2([0, T], \mathbb{R}), \|\cdot\|_{2, \lambda})$, thus proving the existence of $\alpha \in L_{\mathbb{F}}^2([0, T], \mathbb{R})$ solution to (3.4). Finally, an application of Burkholder-Davis-Gundy's inequality to (3.5) yields $\mathbb{E} [\sup_{t \leq T} |X_t|^2]$. The proof is thus complete. \square

Next, we give a sufficient condition for the existence of $P = (P_{11}, P_{12}, P_{22}, P_{22})$ in terms of b, σ, d and T . Let $a = (a_n)_{n \geq 1}$ be the sequence defined as

$$\begin{cases} a_0 &= 1, \\ a_{n+1} &= a_n - \frac{d}{a_n} \left(\frac{b}{\sigma}\right)^2, \quad n \geq 0. \end{cases} \quad (3.7)$$

Let us denote $\mathcal{N} : (d, b, \sigma) \mapsto \inf\{n \geq 1 : a_n > 0 \text{ and } a_{n+1} \leq 0\}$. Clearly, \mathcal{N} is a well defined finite valued function on \mathbb{R}^3 whose image is not restricted to $\{0\}$.

Proposition 3.3. *Assume $\mathcal{N}(d, b, \sigma) \geq 2$ and $T < \mathcal{N}(d, b, \sigma)d$. Then (2.9)-(2.10)-(2.11) has a unique solution in the sense of definition 2.6 on $[0, T]$ with $0 < a_{\mathcal{N}(d, b, \sigma)} \leq P_{11}(0) < 1$.*

Proof. See appendix A. □

Remark 3.4. *Note that when $d = 0$, the sufficient condition above reduces to $\sigma \neq 0$.*

Let us give some intuition as of why the delay feature induces the condition on the coefficients described above to ensure existence. We focus on the first slice $[T - 2d, T - d] \times [-d, 0]$ of the domain, where the solution P is not trivial.

First, note that since $P_{\hat{2}2}$ is a transport of P_{11} which takes the form $P_{\hat{2}2}(t, s) = \sigma^2 P_{11}(t+s+d)1_{t+s+d \leq T} = \sigma^2 1_{t+s+d \leq T}$. On this slice, the kernel P_{11} can be expressed in the following integral form

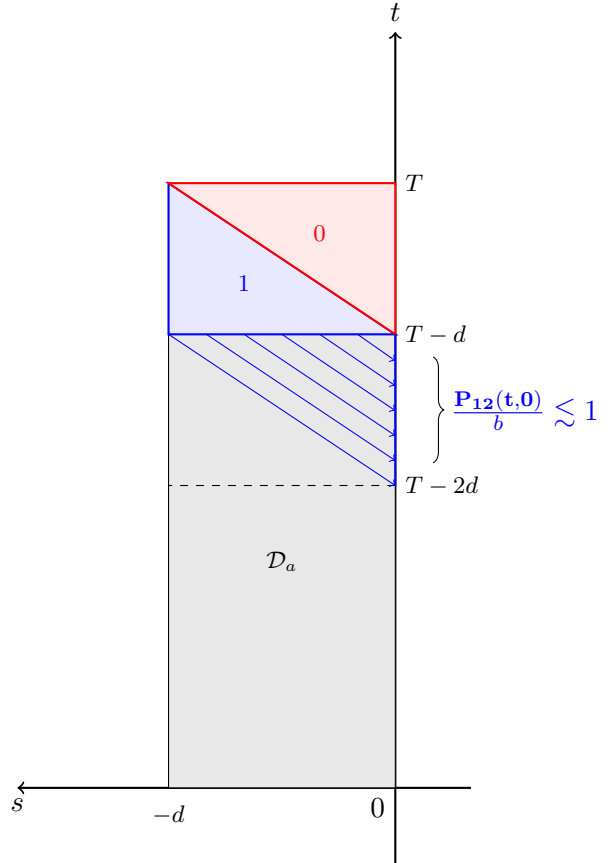
$$P_{11}(t) = 1 - \left(\frac{b}{\sigma}\right)^2 \int_t^{T-d} \left(\frac{\mathbf{P}_{12}(s, 0)}{\mathbf{b}}\right)^2 ds, \quad (3.8)$$

for $t \in [T-2d, T-d]$. Looking at $P_{12}(\cdot, 0)$, we have

$$P_{12}(t, 0) = b - \sigma^{-2} \int_t^{T-d} P_{12}(x, 0) \times \underbrace{P_{22}(x, t-x, 0)}_{\lesssim b^2} dx,$$

see also (A.4). On the right, we represent the value of the normalized kernel P_{12}/b in the different areas of the domain $[T-2d, T] \times [-d, 0]$. If we visualize the evolution of the normalized kernel P_{12}/b in a backward way on the slice $[T-2d, T-d]$, we see that this term is equal to a transport of its value on the boundary $\frac{P_{12}(T-d, s)}{b} = 1$, represented by the blue arrows, minus the integral of a positive source term which is independent of $t \in [T-2d, T-d] \mapsto P_{11}(t)$.

Consequently, the delay $d > 0$ makes the integral term in (3.8) independent of $t \in [T-2d, T-d] \mapsto P_{11}(t)$ and of the order of $d \left(\frac{b}{\sigma}\right)^2$. If this quantity is too large, the kernel P_{11} can then reach negative values, thus making $P_{\hat{2}2}$ negative on the next slice $[T-3d, T-2d]$ and therefore preventing the system (2.9)-(2.10)-(2.11) from having a solution. Repeating this argument from slice to slice of size d in a backward manner induces the aforementioned sufficient condition. Note that these arguments break down when $d = 0$. These arguments are precisely developed in Appendix A.



4 Deep learning scheme

4.1 A quick reminder of PINNs and Deep Galerkin method for PDEs

In order to solve (2.9)-(2.10)-(2.11), we will make use of neural networks in the spirit of the emerging Physics Informed Neural Networks (PINNs) and Deep Galerkin literatures, see Sirignano and Spiliopoulos (2018) and Raissi et al. (2019) to name just a few. We first recall some of the main ideas. Assume we have a nonlinear partial differential equation of the form

$$\begin{aligned} \partial_t u + \mathcal{N}(u) &= 0, & \text{on } \Omega, \\ u &= g, & \text{on } \partial\Omega, \end{aligned} \quad (4.1)$$

where \mathcal{N} is a nonlinear operator, Ω a bounded open subset and g a function on the boundary of the domain. The main idea is to approximate the solution u to (4.1) by a deep neural network. Let us call $t \mapsto u(t, \Theta)$ this network, where Θ and t denote respectively its parameters and a generic element of $\Omega \cup \partial\Omega$. The goal is to find a Θ so that $t \mapsto u(t, \Theta)$ satisfies (4.1). To do so, the idea is to proceed by minimizing the mean square error loss

$$\begin{aligned} \mathcal{L}(\Theta, \mathcal{T}) &= \mathcal{L}_u(\Theta, \mathcal{T}) + \mathcal{L}_f(\Theta, \mathcal{T}), \\ \mathcal{L}_u(\Theta, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} |(\partial_t + \mathcal{N})u(t, \Theta)|^2 1_{t \in \Omega}, \\ \mathcal{L}_f(\Theta, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} |u(t, \Theta) - g(t)|^2 1_{t \in \partial\Omega}, \end{aligned}$$

where \mathcal{L}_f is the loss associated with the initial and boundary constraints on $\partial\Omega$, \mathcal{L}_u the loss associated to the PDE constraint $\partial_t u + \mathcal{N}(u) = 0$ on Ω and \mathcal{T} a random subset of $\partial\Omega \cup \Omega$.

4.2 A tailor-made algorithm for Riccati partial differential equations

Although (2.9)-(2.10)-(2.11) naturally fits the framework of PINNs, we make use of the structure exhibited on the operator P to build a tailor-made algorithm to approximate the system of Riccati transport PDEs (2.9)-(2.10)-(2.11).

Step 1: We define one neural network for each kernel $P_{11}, P_{12}, P_{\hat{2}2}, P_{22}$, as described in Figure 2. Note that usually a unique neural network is used as a surrogate to the function that is to be approximated.

Step 2: We build specific loss functionals for each of our neural networks. To illustrate this, consider for instance the constraint imposed on the derivative of $t \mapsto P_{11}(t)$:

$$\dot{P}_{11}(t) = \frac{P_{12}(t, 0)^2}{P_{\hat{2}2}(t, 0)}, \quad t \in [0, T]. \quad (4.2)$$

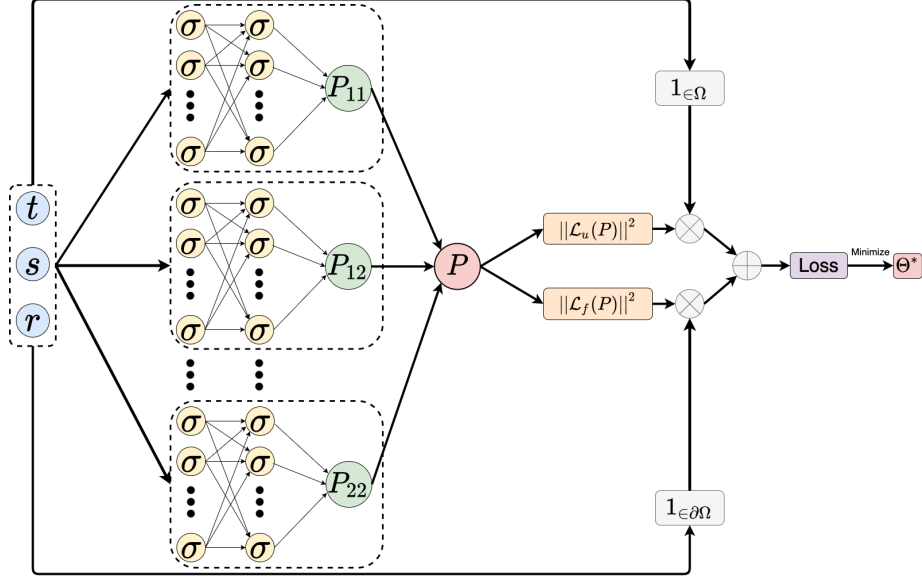


Figure 2: Structure of the model used to solve (2.9)-(2.10)-(2.11).

Note here that, as in the previous sections, we use the convention $0^2/0 = 0$. As a result (4.2) can be rewritten as

$$\begin{aligned} \dot{P}_{11}(t) &= \frac{P_{12}(t, 0)^2}{P_{22}(t, 0)}, & t \leq T - d, \\ \dot{P}_{11}(t) &= 0, & t > T - d. \end{aligned}$$

Thus, a natural contribution to the total loss function to enforce (4.2) would be

$$\mathcal{L}_{11}(\Theta_{11}, \Theta_{12}, \Theta_{22}, \mathcal{T}) = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left(\partial_t P_{11}(\Theta_{11}, t) - \frac{P_{12}(\Theta_{12}, t, 0)^2}{P_{22}(\Theta_{22}, t, 0)} \right)^2, \quad \mathcal{T} \subset [0, T],$$

and the natural gradient descent step associated to the constraint (4.2) would be

$$\Theta_i \leftarrow \Theta_i - \eta \nabla_{\Theta_i} \mathcal{L}_{11}(\Theta_{11}, \Theta_{12}, \Theta_{22}, \mathcal{T}), \quad i \in \{11, 12, \hat{2}2\},$$

Consequently, the constraint (4.2) a priori entails the updating of P_{11}, P_{12} and P_{22} . In particular, it requires to compute the gradient of $(\Theta_{12}, \Theta_{22}) \mapsto \sum_{t \in \mathcal{T}} \frac{P_{12}(\Theta_{12}, t, 0)^2}{P_{22}(\Theta_{22}, t, 0)}$ which is expected to be highly unstable as $t, s \mapsto P_{22}(t, s)$ vanishes for $t + s \geq T - d$. To mitigate this issue, the term $t \mapsto \frac{P_{12}(\Theta_{12}, t, 0)^2}{P_{22}(\Theta_{22}, t, 0)}$ is considered as an exogenous source term for P_{11}

which is fixed when we train P_{11} .

$$\dot{P}_{11}(t) = \frac{P_{12}(t, 0)^2}{\underbrace{P_{\hat{2}2}(t, 0)}} , t \in [0, T].$$

Seen as a fixed exogenous source term when P_{11} is trained

To implement this idea, a second set of neural networks $(\tilde{P}_k(\tilde{\Theta}_k))_{k \in \{11, 12, \hat{2}2, 22\}}$ is initialized with $\tilde{\Theta}_k = \Theta_k$ for $k \in \{11, 12, \hat{2}2, 22\}$ at initialization. These additional networks are then used as surrogates to the right-hand side source terms and will not be used for the computation of the gradients of the losses. They will only be updated at the end of each batch training. Consequently, the gradient descent scheme implemented for each batch \mathcal{T} is the following:

$$\begin{aligned} \text{Step 1} \quad & \begin{cases} \Theta_{11} \leftarrow \Theta_{11} - \eta \nabla_{\Theta_{11}} \left(\mathcal{L}_{11}^r(\Theta_{11}, \tilde{\Theta}_{12}, \tilde{\Theta}_{\hat{2}2}, \mathcal{T}) + \mathcal{L}_1^f(\Theta_{11}, \mathcal{T}) \right) \\ \Theta_{12} \leftarrow \Theta_{12} - \eta \nabla_{\Theta_{12}} \left(\mathcal{L}_{12}^r(\Theta_{12}, \tilde{\Theta}_{\hat{2}2}, \tilde{\Theta}_{22}, \mathcal{T}) + \mathcal{L}_{12}^b(\tilde{\Theta}_{11}, \Theta_{12}, \mathcal{T}) + \mathcal{L}_{12}^f(\Theta_{12}, \mathcal{T}) \right) \\ \Theta_{\hat{2}2} \leftarrow \Theta_{\hat{2}2} - \eta \nabla_{\Theta_{\hat{2}2}} \left(\mathcal{L}_{\hat{2}2}^r(\Theta_{\hat{2}2}, \mathcal{T}) + \mathcal{L}_{\hat{2}2}^b(\tilde{\Theta}_{11}, \Theta_{\hat{2}2}, \mathcal{T}) + \mathcal{L}_{\hat{2}2}^f(\Theta_{\hat{2}2}, \mathcal{T}) \right) \\ \Theta_{22} \leftarrow \Theta_{22} - \eta \nabla_{\Theta_{22}} \left(\mathcal{L}_4^r(\Theta_{\hat{2}2}, \Theta_{22}, \tilde{\Theta}_{22}, \mathcal{T}) + \mathcal{L}_{22}^b(\Theta_{12}, \Theta_{22}, \tilde{\Theta}_{22}, \mathcal{T}) + \mathcal{L}_{22}^f(\Theta_{22}, \mathcal{T}) \right) \end{cases} \\ \text{Step 2} \quad & \left\{ \tilde{\Theta}_k \leftarrow \Theta_k, \quad k \in \{11, 12, \hat{2}2, 22\}, \right. \end{aligned} \tag{4.3}$$

where the \mathcal{L}_i^r 's stand for the *residual loss*, the \mathcal{L}_i^b 's stand for the *boundary loss* and the \mathcal{L}_i^f 's stand for the *final loss*. The precise definitions of the losses are given below.

For each neural network $P_k(\Theta_k)$, $k \in \{11, 12, \hat{2}2, 22\}$, a *follower network* is initialized $\tilde{P}_k(\tilde{\Theta}_k)$, $k \in \{11, 12, \hat{2}2, 22\}$. These follower networks serve as surrogate for the source terms in (2.9)-(2.10)-(2.11). They condition the loss functionals that are used to train the P_k 's and are updated at the end of each batch as described in Algorithm 1.

Losses of P_{11} :

$$\begin{aligned} \mathcal{L}_{11}^r(\Theta_{11}, \tilde{\Theta}_{12}, \tilde{\Theta}_{\hat{2}2}) &= \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left(\partial_t P_{11}(t, \Theta_{11}) - \frac{\tilde{P}_{12}(t, 0, \tilde{\Theta}_{12})^2}{\tilde{P}_{\hat{2}2}(t, 0, \tilde{\Theta}_{\hat{2}2})} \right)^2, \\ \mathcal{L}_{11}^f(\Theta_{11}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left((P_{11}(t, \Theta_{11}) - 1)^2 \mathbf{1}_{t=T} \right). \end{aligned}$$

Losses of P_{12} :

$$\begin{aligned}\mathcal{L}_{12}^r(\Theta_{12}, \tilde{\Theta}_{22}, \tilde{\Theta}_{22}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t,s \in \mathcal{T}} \left((\partial_t - \partial_s)P_{12}(t, s, \Theta_{12}) - \frac{\tilde{P}_{12}(t, 0, \tilde{\Theta}_{12})\tilde{P}_{22}(t, s, 0, \tilde{\Theta}_{22})}{\tilde{P}_{22}(t, 0, \tilde{\Theta}_{22})} \right)^2, \\ \mathcal{L}_{12}^b(\tilde{\Theta}_{11}, \Theta_{12}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t,s \in \mathcal{T}} \left(P_{12}(t, s, \Theta_{12}) - b\tilde{P}_{11}(t, \tilde{\Theta}_{11}) \right)^2 1_{s=0, t \neq 0}, \\ \mathcal{L}_{12}^f(\Theta_{12}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t,s \in \mathcal{T}} (P_{12}(t, s, \Theta_{12})^2 1_{t=T}).\end{aligned}$$

Losses of P_{22} :

$$\begin{aligned}\mathcal{L}_{22}^r(\Theta_{22}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t,s \in \mathcal{T}} ((\partial_t - \partial_s)P_{22}(t, s, \Theta_{22}))^2, \\ \mathcal{L}_{22}^b(\tilde{\Theta}_{11}, \Theta_{22}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t,s \in \mathcal{T}} \left(P_{22}(t, s, \Theta_{22}) - \sigma^2 \tilde{P}_{11}(t, \tilde{\Theta}_{11}) \right)^2 1_{s=0, t \neq T}, \\ \mathcal{L}_{22}^f(\Theta_{22}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t,s \in \mathcal{T}} (P_{22}(t, s, \Theta_{22})^2 1_{t=T}).\end{aligned}$$

Losses of P_{22} :

$$\begin{aligned}\mathcal{L}_{22}^r(\Theta_{22}, \tilde{\Theta}_{22}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t,s,r \in \mathcal{T}} \left((\partial_t - \partial_s - \partial_r)P_4(t, s, r, \Theta_{22}) - \frac{\tilde{P}_{22}(t, 0, r, \tilde{\Theta}_{22})\tilde{P}_{22}(t, s, 0, \tilde{\Theta}_{22})}{\tilde{P}_{22}(t, 0, \tilde{\Theta}_{22})} \right)^2, \\ \mathcal{L}_{22}^b(\tilde{\Theta}_{12}, \Theta_{22}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t,s,r \in \mathcal{T}} \left(P_{22}(t, s, r, \Theta_{22}) - b\tilde{P}_{22}(t, s, \tilde{\Theta}_{12}) \right)^2 1_{r=0, t \neq T} \\ &\quad + \frac{1}{|\mathcal{T}|} \sum_{t,s,r \in \mathcal{T}} \left(P_{22}(t, s, r, \Theta_{22}) - b\tilde{P}_{22}(t, r, \tilde{\Theta}_{12}) \right)^2 1_{s=0, t \neq T}, \\ \mathcal{L}_{22}^f(\Theta_{22}, \mathcal{T}) &= \frac{1}{|\mathcal{T}|} \sum_{t,s,r \in \mathcal{T}} (P_{22}(t, s, r, \Theta_{22})^2 1_{t=T}).\end{aligned}$$

5 Applications to mean-variance portfolio selection with execution delay

5.1 One asset with delay

We now aim at solving the celebrated example of mean-variance portfolio selection, see [Markowitz \(1952\)](#), with execution delay in the spirit of the problem of hedging of European

Algorithm 1: Deep learning scheme to solve (2.9)-(2.10)-(2.11)

Result: A set of optimized parameters $\Theta^* = (\Theta_k^*)_{k \in \{11, 12, \hat{2}2, 22\}}$;
Initialize the learning rate η , the neural networks $P(\Theta) = (P_k(\Theta_k))_k$ and
 $\tilde{P}(\tilde{\Theta}) = (\tilde{P}_k(\tilde{\Theta}_k))_k$;
Copy the weights $\tilde{\Theta}_k \leftarrow \Theta_k$, $k \in \{11, 12, \hat{2}2, 22\}$;
for each batch do
 Randomly sample $\mathcal{T} \subset [0, T] \times [-d, 0]^2$;
 Compute the gradient $\nabla_{\Theta} \mathcal{L}(\Theta, \tilde{\Theta}, \mathcal{T})$ as in (4.3);
 Update $\Theta \leftarrow \Theta - \eta \nabla_{\Theta} \mathcal{L}(\Theta, \tilde{\Theta}, \mathcal{T})$;
 Update $\tilde{\Theta} \leftarrow \Theta$;
end
Return: The set of optimized parameters Θ^* .

options with execution delay presented in [Fabbri and Federico \(2014\)](#). We present here the settings. Let us consider a standard Black-Scholes financial market, composed of a risk-less asset with zero interest rate

$$S_t^0 = 1, \quad t \in [0, T],$$

and a risky asset with dynamics

$$dS_t = S_t \{(\sigma\lambda) dt + \sigma dW_t\}, \quad t \in [0, T],$$

where λ and σ are constants representing respectively the risk premium and the volatility of the risky asset. At every time $t \in [0, T]$ the investor chooses, based on the information \mathcal{F}_t , to allocate the amount of money $\alpha_t \in \mathbb{R}$ into the risky asset. However, due to execution delays this order will be executed at time $t + d$. Set N_t^α (respectively N_t^0) the number of risky (respectively risk-less) shares held at time t . Then, given an investment strategy $\alpha \in \mathcal{A}$, the value $(X_t^\alpha)_{t \in [0, T]}$ of the portfolio, that we suppose self-financing, follows the dynamics

$$\begin{aligned} dX_t^\alpha &= N_t^\alpha dS_t + \underbrace{(dN_t^\alpha)S_t + (dN_t^0)S_t^0}_{=0, \text{ self-financing}} \\ &= \underbrace{N_t^\alpha S_t}_{\alpha_{t-d}} \{(\sigma\lambda)dt + \sigma dW_t\}. \end{aligned}$$

Consequently, the controlled state equation of the portfolio's value is of the form

$$\begin{cases} dX_t^\alpha = \alpha_{t-d} ((\sigma\lambda) dt + \sigma dW_t), & t \in [0, T], \\ X_0 = x_0, \quad \alpha_s = \gamma_s, & \forall s \in [-d, 0], \end{cases}$$

with $x_0 > 0$ and $\gamma \in L^2([-d, 0], \mathbb{R})$. The Mean-Variance portfolio selection problem in continuous-time consists in solving the following constrained problem

$$\begin{cases} \min_{\alpha \in \mathcal{A}} \text{Var}(X_T^\alpha) \\ \text{s.t. } \mathbb{E}[X_T^\alpha] = c. \end{cases} \quad (5.1)$$

It is well-known that problem (5.1) is equivalent to the following **max-min problem**, see Pham (2009, Section 6.6.2)

$$\max_{\eta \in \mathbb{R}} \min_{\alpha \in \mathcal{A}} \mathbb{E} \left[(X_T^\alpha - (c - \eta))^2 \right] - \eta^2. \quad (5.2)$$

Thus, solving problem (5.1) involves two steps. First, the internal minimization problem in terms of the Lagrange multiplier η has to be solved. Second, the optimal value of η for the external maximization problem has to be determined. Thus, with $\xi = c - \eta$, we first define the **Inner optimization problem**:

$$\min_{\alpha \in \mathcal{A}} \mathbb{E} \left[(X_T^\alpha - \xi)^2 \right]. \quad (5.3)$$

Note that, by setting $\tilde{X}^\alpha = X^\alpha - \xi$, the inner problem (5.3) fits into the delayed LQ control problem analysed in Section 3. We first solve the inner optimization problem (5.3) in the following lemma.

Lemma 5.1. *Fix $\eta \in \mathbb{R}$ and $\xi = c - \eta$. Assume $T < d\mathcal{N}(d, (\sigma\lambda), \sigma)$ and define $\alpha^*(\xi)$ as the investment strategy*

$$\alpha_t^*(\xi) = \frac{-1_{t \leq T-d}}{P_{22}(t, 0)} \left\{ (X_t^{\alpha^*} - \xi)P_{12}(t, 0) + \int_{t-d}^t \alpha_s^*(\xi)P_{22}(t, 0, s-t)ds \right\}, \quad (5.4)$$

where P denotes the solution to (2.9)-(2.10)-(2.11) in the sense of Definition 2.6. Then, the inner minimization problem (5.3) admits $\alpha^*(\xi)$ as an admissible optimal feedback strategy and the optimal value is

$$V_0(\xi) = P_{11}(0)(x_0 - \xi)^2 + R(x_0 - \xi, \gamma), \quad (5.5)$$

where $R(\gamma)$ denotes the cost associated to the initial investment strategy γ on $[-d, 0]$

$$R(x, \gamma) = 2x \int_{-d}^0 \gamma_s P_{12}(0, s) ds + \int_{-d}^0 \gamma_s^2 P_{22}(0, s) ds + \int_{[-d, 0]^2} \gamma_s \gamma_u P_{22}(0, s, r) ds dr.$$

Proof. First, note that Proposition 3.3 yields the existence and uniqueness of a solution P to (2.9)-(2.10)-(2.11). Furthermore the admissibility of $\alpha^*(\xi)$ results from Proposition 3.2. For any $\alpha \in \mathcal{A}$, define $\tilde{X}_t^\alpha = X_t^\alpha - \xi$. Then, by Itô's formula we have

$$\begin{cases} d\tilde{X}_t^\alpha = \alpha_{t-d}((\sigma\lambda)dt + \sigma dW_t), & t \in [0, T], \\ \tilde{X}_0 = x_0 - \xi, \quad \alpha_s = \gamma_s, & \forall s \in [-d, 0]. \end{cases}$$

As a result, \tilde{X}^α and X^α have the same dynamics and $\tilde{X}_T^\alpha = X_T^\alpha - \xi$ so that problem (5.3) can be alternatively written as

$$\min_{\alpha \in \mathcal{A}} \mathbb{E} \left[\left(\tilde{X}_T^\alpha \right)^2 \right].$$

Thus, the optimality of $\alpha^*(\xi)$ and the value (5.5) are immediately given by the verification theorem 3.1. \square

Theorem 5.2. *Assume $T < d\mathcal{N}(d, (\sigma\lambda), \sigma)$. Then, the optimal investment strategy for the maximization problem (5.1) is given by $\alpha^*(\xi^*)$ defined in (5.4) with $\xi^* = c - \eta^*$ and*

$$\begin{aligned} \eta^* &= \frac{K(\gamma) + P_{11}(0)(x_0 - c)}{1 - P_{11}(0)}, \\ K(\gamma) &= \int_{-d}^0 \gamma_s P_{12}(0, s) ds. \end{aligned} \tag{5.6}$$

Furthermore, the value of (5.1) is

$$\begin{aligned} \text{Var}(X_T^{\alpha^*}) &= \frac{P_{11}(0)}{1 - P_{11}(0)} (x_0 - c + K(\gamma))^2 \\ &+ \int_{-d}^0 \gamma_s^2 P_{22}(0, s) ds + \int_{[-d, 0]^2} \gamma_s \gamma_u P_{22}(0, s, r) ds dr. \end{aligned} \tag{5.7}$$

Proof. As $T < d\mathcal{N}(d, (\sigma\lambda), \sigma)$, Proposition 3.3 ensures the existence and uniqueness of a solution P to (2.9)-(2.10)-(2.11). From Lemma 5.1 and (5.2), we have that the max-min problem (5.2), which is equivalent to (5.1), reduces to

$$\begin{aligned} &\max_{\eta \in \mathbb{R}} \left\{ V_0(c - \eta) - \eta^2 \right\} \\ &= \max_{\eta \in \mathbb{R}} \left\{ P_{11}(0) (x_0 - (c - \eta))^2 + R(x_0 - (c - \eta), \gamma) - \eta^2 \right\}. \end{aligned}$$

Furthermore, since $T < d\mathcal{N}(d, (\sigma\lambda), \sigma)$, Proposition 3.3 ensures $0 < P_{11}(0) < 1$ so that the maximization problem is strictly concave. Consequently, η^* given by (5.6) is the optimal parameter. Setting $\xi^* = c - \eta^*$ in (5.4) and (5.5) results in the optimality of $\alpha^*(\xi^*)$, and the optimal value (5.7) for the mean-variance problem (5.1). \square

Remark 5.3. *In the absence of pre-investment strategy, $\gamma = 0$, we recover the usual form of the efficient frontier formula*

$$\text{Var}(X_T^{\alpha^*}) = \frac{P_{11}(0)}{1 - P_{11}(0)} (x_0 - c)^2.$$

Our observations from the simulations are the following.

Efficient frontier: In Figure 3, we plot the efficient frontier for different delays d . Note that the greater the delay, the greater is the variance. This could have been foreseen by observing that, when the initial control is set to 0, i.e. $\gamma = 0_{L^2}$, the value function takes the form $V(x, 0_{L^2}) = P_{11}(0)x^2$, see (3.2). As the value function is clearly an increasing function of the delay, the terminal variance of the portfolio $\text{Var}(X_T^{\alpha^*}) = \frac{P_{11}(0)}{1-P_{11}(0)}(x_0 - c)^2$ is also an increasing function of the delay.

Destabilization effect : In Figure 4, we plot different scenarios of portfolio allocation. We observe a destabilization effect and a supplement of volatility induced by the delay feature. We also note the tendency to invest more aggressively for greater values of the delay, as the investor has less time to ensure that the promised yield is achieved. We propose the following interpretation: In the classical setting, where $d = 0$, if Y^* denotes the optimal portfolio value process, the optimal investment strategy is of the form $\alpha_t^* = -\frac{b}{\sigma^2}(Y^{\alpha^*} - \mu^*)$ for a certain constant $\mu^* > c$. It can then easily be shown that $Y^* \leq \mu^*$. Thus, the optimal strategy consists in aiming from below at a fixed target μ^* . When $d > 0$, the optimal control is composed of an additional *inertial term*

$$\alpha_t^*(\xi^*) = \frac{-1_{t \leq T-d}}{P_{22}(t, 0)} \left\{ \underbrace{(X_t^{\alpha^*} - \xi^*)P_{12}(t, 0)}_{\text{Usual mean-reverting term}} + \underbrace{\int_{t-d}^t \alpha_s^*(\xi)P_{22}(t, 0, s-t)ds}_{\text{New inertial term}} \right\},$$

so that, contrary to the case where $d = 0$, the optimal control does not cancel when the target ξ^* is attained. Also, note that at every time t , the agent doesn't have any control on the near future from t to $t + d$.

Kernel P : In Figure 5, we plot the kernels P_{11}, P_{12}, P_{22} , and P_{22} . Note the discontinuity between \mathcal{D}_b and \mathcal{D}_c also described in Figure 1.

5.2 One asset with delay and one without

To further explore the effect of the delay on the control, we now study a toy example where the investor has two investment opportunities, one with a delayed execution and one without. More precisely, consider the following portfolio dynamic

$$\begin{cases} dX_t^{(\alpha, \beta)} = \alpha_t \{(\sigma_1 \lambda_1)dt + \sigma_1 dW_t^1\} + \beta_{t-d} \{(\sigma_2 \lambda_2)dt + \sigma_2 dW_t^2\}, & t \in [0, T], \\ X_0 = x_0, \quad \beta_s = \gamma_s, \quad s \in [-d, 0], \\ \langle W^1, W^2 \rangle_t = \rho t, \end{cases}$$

where $x_0 > 0$ and $\gamma \in L^2([-d, 0], \mathbb{R})$, together with the same optimization objective (5.1) as before. Here, α_t and β_t correspond respectively to the amounts of money the investor

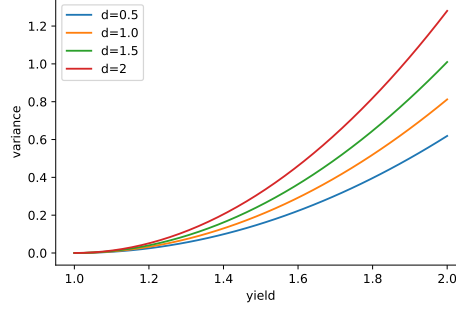


Figure 3: Efficient frontier with $\sigma = 1$, $\lambda = 0.5$, $T = 5$ and $\gamma \equiv 0$.

decides to invest at time t in the undelayed and the delayed risky assets. The constants λ_i and σ_i represent respectively the risk premium and the volatility of the risky asset i . Following the heuristic approach of Section 2, we define the following set of Riccati-PDEs on $[0, T] \times [-d, 0]^2$

$$\begin{aligned} \dot{P}_{11}(t) &= \lambda_1^2 P_{11}(t) + \frac{P_{12}(t, 0)^2}{P_{22}(t, 0)}, & (5.8) \\ (\partial_t - \partial_s)(P_{12})(t, s) &= \lambda_1^2 P_{12}(t, s) + \frac{P_{12}(t, 0)P_{22}(t, s, 0)}{P_{22}(t, 0)}, \\ (\partial_t - \partial_s)(P_{22})(t, s) &= 0, \\ (\partial_t - \partial_s - \partial_r)(P_{22})(t, s, r) &= \lambda_1^2 \frac{P_{12}(t, s)P_{12}(t, r)}{P_{11}(t)} + \frac{P_{22}(t, s, 0)P_{22}(t, 0, r)}{P_{22}(t, 0)}, \end{aligned}$$

accompanied by the boundary conditions, for almost any $t, s \in [0, T] \times [-d, 0]$

$$\begin{aligned} P_{12}(t, -d) &= \lambda_2 \sigma_2 \left(1 - \rho \frac{\lambda_1}{\lambda_2}\right) P_{11}(t), & P_{22}(t, -d) &= \sigma_2^2 (1 - \rho^2) P_{11}(t), & (5.9) \\ P_{22}(t, s, -d) &= \lambda_2 \sigma_2 \left(1 - \rho \frac{\lambda_1}{\lambda_2}\right) P_{12}(t, s), & P_{22}(t, -d, s) &= \lambda_2 \sigma_2 \left(1 - \rho \frac{\lambda_1}{\lambda_2}\right) P_{12}(t, s), \end{aligned}$$

and the terminal constraints

$$P_{11}(T) = 1, \quad P_{12}(T, s) = P_{22}(T, s) = P_{22}(T, s, r) = 0, \quad (5.10)$$

for almost every $s, r \in [-d, 0]$.

As in the previous section, we first solve the inner optimization problem 5.3.

Lemma 5.4. *Fix $\eta \in \mathbb{R}$ and $\xi = c - \eta$. Assume (5.8)-(5.9)-(5.10) admits a piecewise absolutely continuous solution with $P_{22}(t) > 0$ for any $t \leq T - d$, and define the couple*

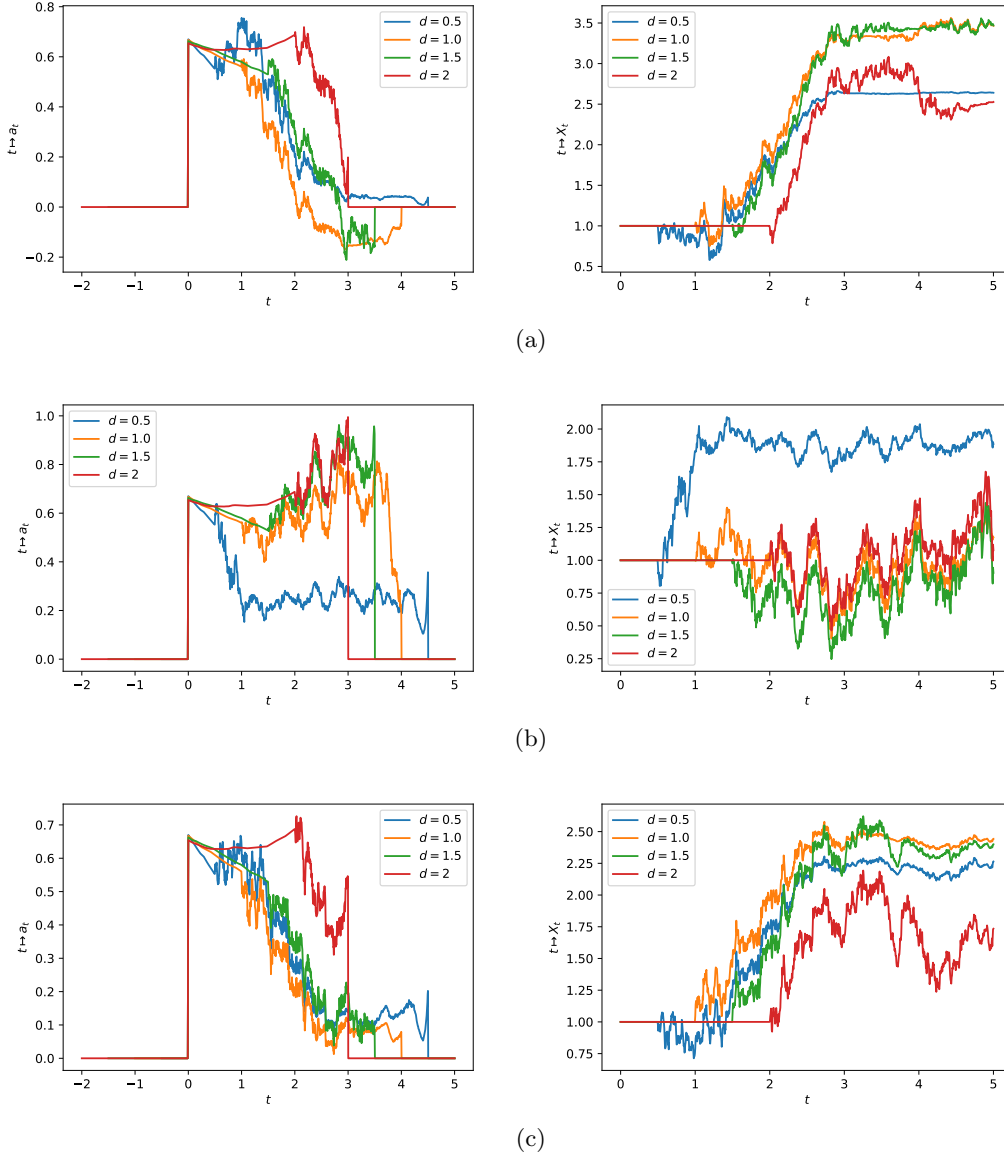
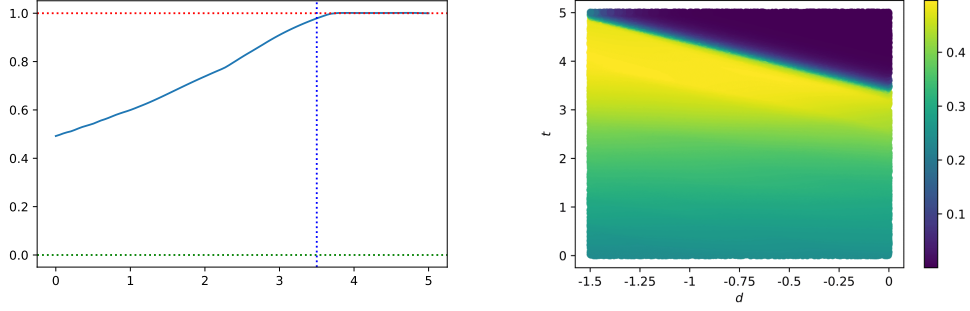
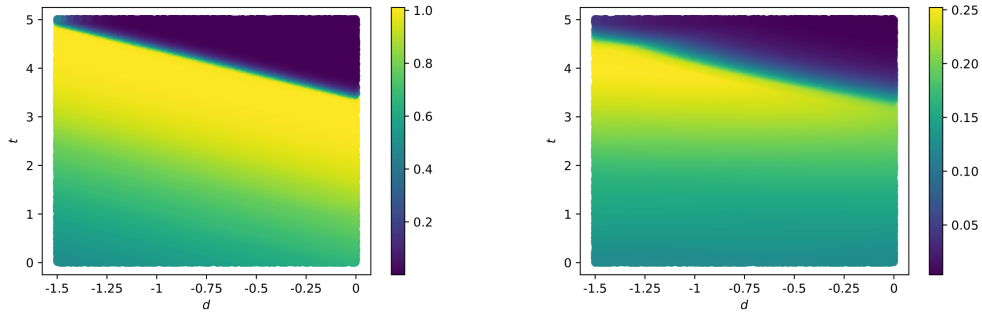


Figure 4: Different scenarios of the optimal portfolio with $c = 1.6$, $\sigma = 1$, $\lambda = 0.5$, and $T = 5$. Left: $t \mapsto \alpha^*$, right: $t \mapsto X_t^*$. Note the destabilization effect and the supplement of volatility induced by the delay feature. Note also the tendency to invest more aggressively the delayed investor has, as she has less time to ensure the promised yield. $\xi^*(d = 0.5) = 2.57$, $\xi^*(d = 1) = 2.68$, $\xi^*(d = 1.5) = 2.80$, $\xi^*(d = 2) = 2.97$.



(a) Right: $t \mapsto P_{11}(t)$. Left: $t, s \mapsto P_{12}(t, s)$.



(b) Right: $t \mapsto P_{22}(t, s)$. Left: $t, s \mapsto P_{22}(t, s, 0)$.

Figure 5: Numerical results of Algorithm 1 with $\sigma = 1$, $\lambda = 0.5$, $d = 1.5$, and $T = 5$.

$(\alpha^*, \beta^*)(\xi)$ as the investment strategies

$$\alpha_t^*(\xi) = - \left\{ \frac{\lambda_1}{\sigma_1} (X_t^* - \xi) + \rho \frac{\sigma_2}{\sigma_1} \beta_{t-d}^* + \frac{\lambda_1}{\sigma_1 P_{11}(t)} \int_{t-d}^t \beta_s^*(\xi) P_{12}(t, s-t) ds \right\},$$

$$\beta_t^*(\xi) = \frac{-1_{t \leq T-d}}{P_{22}(t, 0)} \left\{ P_{12}(t, 0) (X_t^* - \xi) + \int_{t-d}^t \beta_s^*(\xi) P_{22}(t, 0, r-t) dr \right\},$$

where X^* denotes the state process $X^{(\alpha^*, \beta^*)}$. Then, the inner minimization problem (5.3) admits $(\alpha^*(\xi), \beta^*(\xi))$ as an optimal feedback strategy and the optimal value is

$$V_0(\xi) = P_{11}(0)(x_0 - \xi)^2 + R(x_0 - \xi, \gamma),$$

where $R(\gamma)$ denotes the cost associated to the initial investment strategy γ on $[-d, 0]$

$$R(x, \gamma) = 2x \int_{-d}^0 \gamma_s P_{12}(0, s) ds + \int_{-d}^0 \gamma_s^2 P_{22}(0, s) ds + \int_{[-d, 0]^2} \gamma_s \gamma_u P_{22}(0, s, r) ds dr.$$

Proof. The proof is similar to the one of Lemma 5.1. □

Finally, the parameter η^* and efficient frontier $\text{Var}(X_T^*) = f(c)$ are given by the same formulas (5.6) and (5.7) as in the mono-asset case, γ being the pre-investment strategy of the delayed asset.

Remark 5.5. *One surprise that emerges is that the "buy the good stock sell the bad one" criterion is unchanged for the delayed asset. Indeed, the sign of the control for this asset is still given by the sign of $1 - \rho \frac{\lambda_1}{\lambda_2}$, that fixes the sign of the P_{12} and P_{22} , as it would be in the case without delay², see the boundary conditions (5.9). But this threshold disappears in the undelayed asset's control as now only the term $\frac{\lambda_1}{\sigma_1}$ remains in the mean-reverting term.*

Numerical simulations: To exhibit the effect of the correlation ρ , we generate two independent Brownian motions $(W_t^1)_{t \in [0, T]}$ and $(B_t)_{t \in [0, T]}$ and define the Brownian motion $(W_t^2)_{t \in [0, T]}$ as

$$W_t^2 = \rho W_t^1 + \sqrt{1 - \rho^2} B_t, \quad t \in [0, T].$$

We then compare different scenarios with different values of correlation ρ and delay d while fixing W^1 and B . The numerical simulations can be found in Figures 6, 7 and 8. As it could have been expected, we see from (5.9) and Figure 6, that the greater ρ is, the more favored the undelayed asset is.

²When $d = 0$, recall that $\alpha_t^* = \frac{\lambda_1 P_t}{\sigma_1(1-\rho^2)}(1 - \rho \frac{\lambda_2}{\lambda_1})(\xi^* - X_t^*)$ and $\beta^* = \frac{\lambda_2 P_t}{\sigma_2(1-\rho^2)}(1 - \rho \frac{\lambda_1}{\lambda_2})(\xi^* - X_t^*)$ with P being a positive function and $\xi^* \geq X^*$. Thus, in the classical setting, the buy or sell thresholds are $(1 - \rho \frac{\lambda_2}{\lambda_1})$ and $(1 - \rho \frac{\lambda_1}{\lambda_2})$.

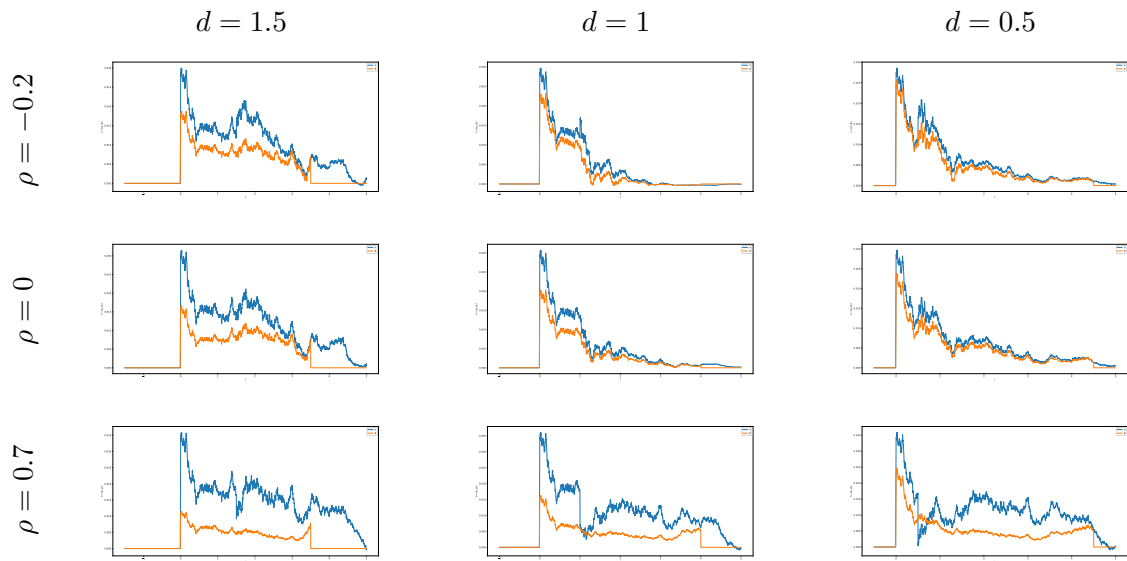


Figure 6: $t \mapsto (\alpha_t^*, \beta_t^*)$, with $\sigma_1 = \sigma_2 = 1$, $\lambda_1 = \lambda_2 = 0.5$ and $T = 5$. Blue : α^* , orange : β^* . The same realizations of W and B were used for all experiments. Note that the more positively correlated the assets are, the more favored the undelayed asset is.

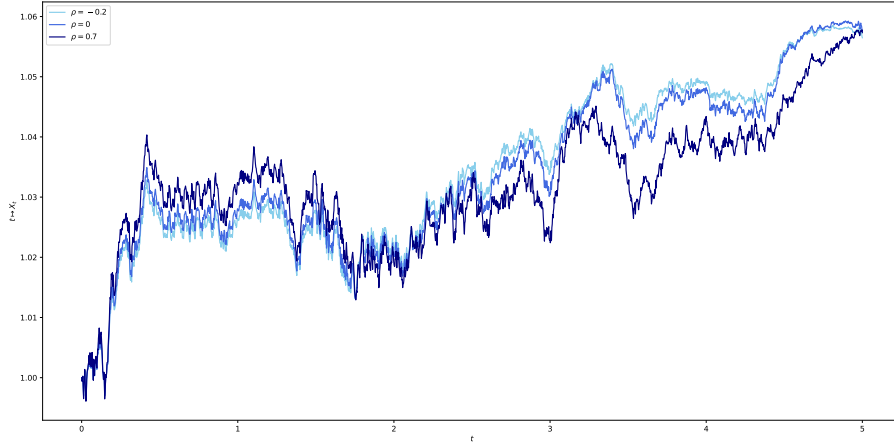


Figure 7: $t \mapsto X_t^*$, with $\sigma_1 = \sigma_2 = 1$, $\lambda_1 = \lambda_2 = 0.5$, $T = 5$ and $d = 1.5$ for $\rho = -0.7$, 0 and 0.7. The same realizations of the Brownian motions W^1 and B was used for all experiments.

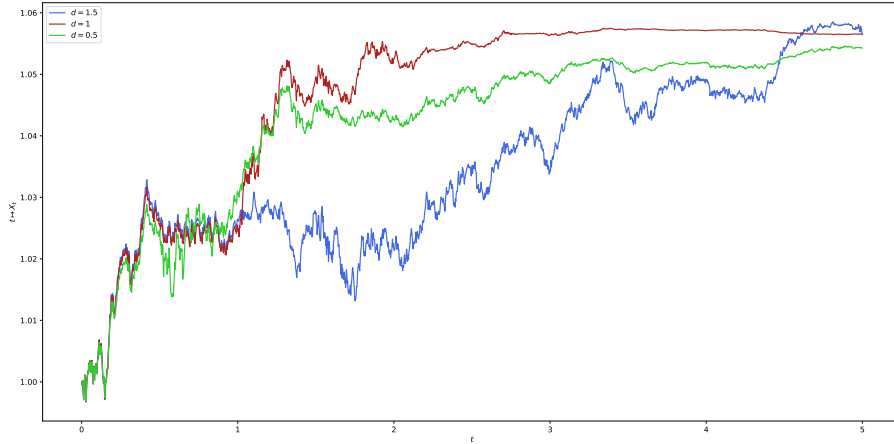


Figure 8: $t \mapsto X_t^*$, with $\sigma_1 = \sigma_2 = 1$, $\lambda_1 = \lambda_2 = 0.5$, $T = 5$ and $\rho = -0.7$ for $d = 1.5$, 1 and 0.5. The same realizations of the Brownian motions W^1 and B was used for all experiments.

References

- Yogish Alekal, Pavol Brunovsky, DH Chyung, and E Lee. The quadratic problem for systems with time delays. *IEEE Transactions on Automatic Control*, 16(6):673–687, 1971.
- Patrick K Asea and Paul J Zak. Time-to-build and cycles. *Journal of economic dynamics and control*, 23(8):1155–1175, 1999.
- Mauro Bambi. Endogenous growth and time-to-build: The ak case. *Journal of Economic Dynamics and Control*, 32(4):1015–1040, 2008.
- Mauro Bambi, Giorgio Fabbri, and Fausto Gozzi. Optimal policy and consumption smoothing effects in the time-to-build ak model. *Economic Theory*, 50(3):635–669, 2012.
- Alain Bensoussan, Giuseppe Da Prato, Michel C Delfour, and Sanjoy K Mitter. *Representation and control of infinite dimensional systems*. Springer Science & Business Media, 2007.
- René Carmona, Jean-Pierre Fouque, Seyyed Mostafa Mousavi, and Li-Hsien Sun. Systemic risk and stochastic games with delay. *Journal of Optimization Theory and Applications*, 179(2):366–399, 2018.
- Hippolyte d’Albis, Emmanuelle Augeraud-Véron, and Alain Venditti. Business cycle fluctuations and learning-by-doing externalities in a one-sector model. *Journal of Mathematical Economics*, 48(5):295–308, 2012.
- Nicole El Karoui. Les aspects probabilistes du contrôle stochastique. In *École d’été de Probabilités de Saint-Flour IX-1979*, pages 73–238. Springer, 1981.
- Giorgio Fabbri and Salvatore Federico. On the infinite-dimensional representation of stochastic controlled systems with delayed control in the diffusion term. *Mathematical Economics Letters*, 2(3-4):33–43, 2014.
- Jean-Pierre Fouque and Zhaoyu Zhang. Deep learning methods for mean field control problems with delay. *arXiv preprint arXiv:1905.00358*, 2019.
- Fausto Gozzi, Sociali di Roma, and Carlo Marinelli. Stochastic optimal control of delay equations arising in advertising models. *Stochastic Partial Differential Equations and Applications- VII*, page 133, 2005.
- Fausto Gozzi, Carlo Marinelli, and Sergei Savin. On controlled linear diffusions with delay in a model of optimal advertising under uncertainty with memory effects. *Journal of optimization theory and applications*, 142(2):291–321, 2009.

- Robert E Hall, Christopher A Sims, Franco Modigliani, and William Brainard. Investment, interest rates, and the effects of stabilization policies. *Brookings papers on economic activity*, 1977(1):61–121, 1977.
- Jiequn Han and Ruimeng Hu. Recurrent neural networks for stochastic control problems with delay. *arXiv preprint arXiv:2101.01385*, 2021.
- Mihai Huzmezan, William A Gough, Guy A Dumont, and Sava Kovac. Time delay integrating systems: a challenge for process control industries. a practical solution. *Control Engineering Practice*, 10(10):1153–1161, 2002.
- Akira Ichikawa. Quadratic control of evolution equations with delays in control. *SIAM Journal on control and optimization*, 20(5):645–668, 1982.
- Elias Jarlebring and Tobias Damm. The lambert w function and the spectrum of some multidimensional time-delay systems. *Automatica*, 43(12):2124–2128, 2007.
- Finn E Kydland and Edward C Prescott. Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pages 1345–1370, 1982.
- Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. doi: 10.1111/j.1540-6261.1952.tb01525.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1952.tb01525.x>.
- Wilfried Pauwels. Optimal dynamic advertising policies in the presence of continuously distributed time lags. *Journal of Optimization Theory and Applications*, 22(1):79–89, 1977.
- Huy en Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media, 2009.
- Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- Suresh P Sethi. Sufficient conditions for the optimal control of a class of systems with continuous lags. *Journal of Optimization Theory and Applications*, 13(5):545–552, 1974.
- Rifat Sipahi, Silviu-Iulian Niculescu, Chaouki T Abdallah, Wim Michiels, and Keqin Gu. Stability and stabilization of systems with time delay. *IEEE Control Systems Magazine*, 31(1):38–65, 2011.
- Justin Sirignano and Konstantinos Spiliopoulos. Dgm: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364, 2018.

Yu-Chu Tian and Furong Gao. Control of integrator processes with dominant time delay. *Industrial & engineering chemistry research*, 38(8):2979–2983, 1999.

John D Tsoukalas. Time to build capital: Revisiting investment-cash-flow sensitivities. *Journal of Economic Dynamics and Control*, 35(7):1000–1016, 2011.

A Proof of Proposition 3.3

Our proof extends Alekal et al. (1971, see Theorem 5) to the case where the volatility is controlled. It consists in slicing the domain \mathcal{D} in slices of size d and proceeding by a backward recursion. More precisely, we show existence, uniqueness, and positiveness of the first coordinates, over a sequence of slices $([T - (n + 1)d, T - nd] \times [-d, 0]^2)_n$. We then concatenate the sequence of absolutely continuous solutions obtained, which yields a piece-wise absolutely continuous solution. In each slice, the proof consists of the following steps

- (1) Show that there exists a unique solution on a small interval;
- (2) Prove that the local solution is Lipschitz;
- (3) As a result extend the solution to the whole slice.

We finally concatenate the sequence of solutions obtained above.

A.1 Slice $t \in [T - d, T]$, initialization

On $\mathcal{D}_b \cup \mathcal{D}_c$, the constraints (2.9)-(2.10)-(2.11) on P_{12} , $P_{\hat{2}2}$ and P_{22} reduce to linear homogeneous transport equations admitting closed form solutions given, for every $(t, s, r) \in \mathcal{D}_b \cup \mathcal{D}_c$, by

$$\begin{aligned} P_{12}(t, s) &= bP_{11}(t + s + d)1_{t+s+d \leq T}, & P_{\hat{2}2}(t, s) &= \sigma^2 P_{11}(t + s + d)1_{t+s+d \leq T}, \\ P_{22}(t, s, r) &= b^2 P_{11}(t + s \vee r + d)1_{t+s \vee r + d \leq T}. \end{aligned}$$

Or, as $P_{11}(t) = 1$ for any $t \geq T - d$, we then have for every $(t, s, r) \in \mathcal{D}_b \cup \mathcal{D}_c$

$$\begin{aligned} P_{11}(t) &= 1, & P_{12}(t, s) &= b1_{t+s+d \leq T}, \\ P_{\hat{2}2}(t, s) &= \sigma^2 1_{t+s+d \leq T}, & P_{22}(t, s, r) &= b^2 1_{t+s \vee r + d \leq T}. \end{aligned}$$

The existence and uniqueness in the sense of Definition 2.6 are thus trivially proved on $[T - d, T]$.

A.2 Slice $[T - 2d, T - d]$

On $[T - 2d, T - d] \times [-d, 0]^2$, we have $P_{\hat{2}2}(t, s) = \sigma^2 P_{11}(t + s + d)$ so that $P_{\hat{2}2}(t, 0) = \sigma^2 P_{11}(t + d) = \sigma^2$. Consequently, the system (2.9)-(2.10)-(2.11) reduces to

$$\begin{aligned} \dot{P}_{11}(t) &= \frac{P_{12}(t, 0)^2}{\sigma^2}, \\ (\partial_t - \partial_s)(P_{12})(t, s) &= \frac{P_{12}(t, 0)P_{22}(t, s, 0)}{\sigma^2}, \\ (\partial_t - \partial_s - \partial_r)(P_{22})(t, s, r) &= \frac{P_{22}(t, s, 0)P_{22}(t, 0, r)}{\sigma^2}, \end{aligned} \quad (\text{A.1})$$

with terminal conditions

$$P_{11}(T - d) = 1, \quad P_{12}(T - d, s) = b, \quad P_{22}(T - d, s, r) = b^2, \quad (\text{A.2})$$

and boundary constraints

$$P_{12}(t, -d) = bP_{11}(t), \quad P_{22}(t, s, -d) = bP_{12}(t, s). \quad (\text{A.3})$$

Thus, for every $(t, s, r) \in [T - 2d, T - d] \times [-d, 0]^2$, the set of equations (A.1) and constraints (A.2)-(A.3) can be rewritten in the following integral form

$$\begin{aligned} P_{11}(t) &= 1 - \sigma^{-2} \int_t^{T-d} P_{12}(x, 0)^2 dx, \\ P_{12}(t, s) &= bP_{11}((T - d) \wedge (t + s + d)) \\ &\quad - \sigma^{-2} \int_t^{(T-d) \wedge (t+s+d)} P_{12}(x, 0)P_{22}(x, t + s - x, 0) dx, \\ P_{22}(t, s, r) &= bP_{12}((T - d) \wedge (t + s \wedge r + d), (s - r) \vee (r - s) - d) \\ &\quad - \sigma^{-2} \int_t^{(T-d) \wedge (t+s \wedge r+d)} P_{22}(x, t + s - x, 0)P_{22}(x, 0, t + r - x) dx. \end{aligned} \quad (\text{A.4})$$

We then make use of the following lemma to prove local existence of a solution.

Lemma A.1. *There exists $\tau \in (0, d]$ such that system (A.4) has a unique absolutely continuous solution on $[T - \tau - d, T - d] \times [-d, 0]^2$.*

Proof. Let $\tau \in (0, d]$ and \mathcal{S}_τ denote the Banach space of absolutely continuous functions $\xi = (\xi_1(\cdot), \xi_2(\cdot, \cdot), \xi_3(\cdot, \cdot, \cdot))$ defined on

$$\mathcal{D}_\tau = \{(t, s, r) \mid T - d - \tau \leq t \leq T - d, -d \leq s, r \leq 0\},$$

endowed with the sup-norm

$$\|\xi\|_\infty = \|\xi_1\|_\infty + \|\xi_2\|_\infty + \|\xi_3\|_\infty,$$

where $\|\xi_1\|_\infty, \|\xi_2\|_\infty$ and $\|\xi_3\|_\infty$ denote, with a slight abuse of notation, the respective sup-norm on $[T-d-\tau, T-d]$, $[T-d-\tau, T-d] \times [-d, 0]$ and $[T-d-\tau, T-d] \times [-d, 0]^2$. Let \mathcal{B}_τ denote the ball in \mathcal{S}_τ

$$\mathcal{B}_\tau = \{(\xi_1, \xi_2, \xi_3) \in \mathcal{S}_\tau : \|\xi_1 - 1\| \leq 1/2, \|\xi_2 - b\| \leq |b|/2, \|\xi_3 - b^2\| \leq b^2/2\},$$

On \mathcal{B}_τ , we denote by $\phi = (\phi_1, \phi_2, \phi_3)$ the operator defined as follows

$$\begin{aligned} (\phi_1 \xi)(t) &= 1 - \sigma^{-2} \int_t^{T-d} \xi_2(x, 0)^2 dx \\ (\phi_2 \xi)(t, s) &= b\phi_1(\xi)((T-d) \wedge (t+s+d)) \\ &\quad - \sigma^{-2} \int_t^{(T-d) \wedge (t+s+d)} \xi_2(r, 0) \xi_3(r, t+s-x, 0) dx \\ (\phi_3 \xi)(t, s, r) &= b\phi_2(\xi)((T-d) \wedge (t+s \wedge r+d), (s-r) \vee (r-s) - d) \\ &\quad - \sigma^{-2} \int_t^{(T-d) \wedge (t+s \wedge r+d)} \xi_3(x, t+s-x, 0) \xi_3(x, 0, t+r-x) dx. \end{aligned}$$

Clearly, there exists $\tilde{\tau} > 0$ such that for any $\tau \leq \tilde{\tau}$, $\phi(\mathcal{B}_\tau) \rightarrow \mathcal{B}_\tau$. We show a contraction property on ϕ . For any $\xi, \xi' \in \mathcal{B}_\tau$, we have the following inequalities

$$\begin{aligned} \|\phi_1(\xi) - \phi_1(\xi')\|_\infty &\leq 4\tau\sigma^{-2}|b|\|\xi_2 - \xi'_2\|_\infty, \\ \|\phi_2(\xi) - \phi_2(\xi')\|_\infty &\leq 4\tau\sigma^{-2}(|b|\|\xi_2 - \xi'_2\|_\infty + |b|^2\|\xi_3 - \xi'_3\|_\infty) \\ &\quad + |b|\|\phi_1(\xi) - \phi_1(\xi')\|_\infty, \\ \|\phi_3(\xi) - \phi_3(\xi')\|_\infty &\leq |b|\|\phi_2(\xi) - \phi_2(\xi')\|_\infty + 4\tau\sigma^{-2}|b|^2\|\xi_3 - \xi'_3\|_\infty. \end{aligned}$$

Consequently, the operator ϕ satisfies

$$\|\phi(\xi) - \phi(\xi')\|_\infty \leq \tau m \|\xi - \xi'\|_\infty,$$

where $m > 0$ depends on b and σ . Therefore, for $\tau < \tilde{\tau} \wedge m^{-1}$, the operator ϕ is a contraction of \mathcal{B}_τ into itself. Thus, ϕ admits a unique fixed point in \mathcal{B}_τ , which is solution to (A.4) on \mathcal{D}_τ . □

Lemma A.2. *Let $\xi = (\xi_1, \xi_2, \xi_3)$ denote the absolutely continuous solution of (A.4) on \mathcal{D}_τ from Lemma A.1. Then ξ is Lipschitz in each variable on \mathcal{D}_τ .*

Proof. As ξ_1, ξ_2 and ξ_3 are continuous on \mathcal{D}_τ , there exists a constant $m > 0$ such that $|\xi_1| \wedge |\xi_2| \wedge |\xi_3| \leq m$ on \mathcal{D}_τ . Thus, ξ_1 is Lipschitz with constant $\kappa = m^2\sigma^{-2}$. Let us now

show that ξ_2 and ξ_3 are Lipschitz in the s -variable. Fix $t \in [T - d - \tau, T - d]$ and $\eta > 0$. Then, for any $s \in [-d, 0]$ such that $s + \eta \in [-d, 0]$, we have

$$\begin{aligned} |\xi_2(t, s) - \xi_2(t, s + \eta)| &\leq \kappa\eta + \sigma^{-2} \left| \int_t^{(T-d)\wedge(t+s+\eta+d)} \xi_2(x, 0)\xi_3(x, t + s + \eta - x, 0)dx \right. \\ &\quad \left. - \int_t^{(T-d)\wedge(t+s+d)} \xi_2(x, 0)\xi_3(x, t + s - x, 0)dx \right| \\ &\leq \kappa\eta + \mathbf{I}(t, s) + \mathbf{II}(t, s), \end{aligned}$$

Since $|\xi_2| \leq m$, it yields

$$\begin{aligned} \mathbf{I}(t, s) &\leq \int_t^{(T-d)\wedge(t+s+d)} |\xi_2(x, 0)| |\xi_3(x, t + s + \eta - x, 0) - \xi_3(x, t + s - x, 0)| dx \\ &\leq m \int_t^{(T-d)\wedge(t+s+d)} \epsilon(x) dx, \end{aligned}$$

where ϵ is defined as

$$\epsilon(x) = \sup_{\substack{s, r \\ \in [-d, 0]^2}} |\xi_3(x, s, r) - \xi_3(x, s + \eta, r)| + \sup_{s \in [-d, 0]} |\xi_2(x, s) - \xi_2(x, s + \eta)|.$$

Furthermore, as $|\xi_2| \wedge |\xi_3| \leq m$ on \mathcal{D}_τ , we have

$$\begin{aligned} \mathbf{II}(t, s) &\leq \int_{(T-d)\wedge(t+s+d)}^{(T-d)\wedge(t+s+\eta+d)} |\xi_2(x, 0)\xi_3(x, t + s + \eta - x, 0)| dx \\ &\leq m^2\eta. \end{aligned}$$

Consequently, for any $t \in [T - d - \tau, T - d]$, we obtain

$$\sup_s |\xi_2(t, s) - \xi_2(t, s + \eta)| \leq m^2\eta + m \int_t^{T-d} \epsilon(r) dr. \quad (\text{A.5})$$

Looking at the equation of ξ_3 in system (A.4), we obtain in a similar manner

$$|\xi_3(t, s, r) - \xi_3(t, s + \eta, r)| \leq |b|\mathbf{I}(t, s, r) + \sigma^{-2}\mathbf{II}(t, s, r). \quad (\text{A.6})$$

An application to the triangle inequality combined with (A.5) and the Lipschitzianity of ξ_1 leads to

$$\begin{aligned} \mathbf{I}(t, s, r) &\leq |\xi_2((T-d) \wedge (t + (s + \eta) \wedge r + d), (s + \eta - r) \vee (r - (s + \eta)) - d) \\ &\quad - \xi_2((T-d) \wedge (t + s \wedge r + d), (s - r) \vee (r - s) - d)| \\ &\leq (\kappa + m^2(1 + \sigma^{-2}))\eta + m \int_{(T-d)\wedge(t+s\wedge r+d)}^{T-d} \epsilon(x) dx \\ &\leq (1 + 2\kappa)\eta + m \int_t^{T-d} \epsilon(x) dx. \end{aligned} \quad (\text{A.7})$$

Furthermore

$$\begin{aligned}
\mathbf{II}(t, s, r) &\leq \left| \int_t^{(T-d)\wedge(t+(s+\eta)\wedge r+d)} \xi_3(x, t+s+\eta-x, 0) \xi_3(x, 0, t+r-x) dx \right. \\
&\quad \left. - \int_t^{(T-d)\wedge(t+s\wedge r+d)} \xi_3(x, t+s-x, 0) \xi_3(x, 0, t+u-x) dx \right| \\
&\leq \int_t^{(T-d)\wedge(t+s\wedge r+d)} |\xi_3(x, 0, t+r-x)| |\xi_3(x, t+s-x, 0) \\
&\quad - \xi_3(x, t+(s+\eta)-x, 0)| dx \\
&\quad + \int_{(T-d)\wedge(t+s\wedge r+d)}^{(T-d)\wedge(t+(s+\eta)\wedge r+d)} |\xi_3(x, t+(s+\eta)-x, 0) \xi_3(r, 0, t+r-x)| dx \\
&\leq m^2 \eta + \int_t^{T-d} \epsilon(r) dr
\end{aligned} \tag{A.8}$$

Thus, inequality (A.7) together with (A.8) and (A.6) yield the existence of a positive constant $c > 0$, independent of η , such that

$$\sup_{\substack{s, r \\ \in [-d, 0]^2}} |\xi_3(t, s, r) - \xi_3(t, s+\eta, r)| \leq c \left(\eta + \int_t^{T-d} \epsilon(r) dr \right),$$

which, combined with (A.5) leads, for any $t \in [T-d-\tau, T-d]$, to

$$\epsilon(t) \leq c \left(\eta + \int_t^{T-d} \epsilon(r) dr \right).$$

Consequently, an application to Gronwall's lemma yields $\epsilon(t) \leq m'\eta$ on $[T-d-\tau, T-d]$, with $m' > 0$. Thus, ξ_2 and ξ_3 are Lipschitz in the s -variable. The arguments for showing that ξ_2 and ξ_3 are Lipschitz in the t -variable and ξ_3 Lipschitz in the r -variable follow the same line. \square

Lemma A.3. *There exists a unique absolutely continuous solution $\xi = (\xi_1, \xi_2, \xi_3)$ of (A.4) on $[T-2d, T-d] \times [-d, 0]^2$ such that $\xi_1 \geq 1 - d \left(\frac{b}{\sigma}\right)^2 > 0$.*

Proof. Let $\theta \in [T-2d, T-d]$ denote the lower limit of all τ 's such that there exists an absolutely continuous solution (ξ_1, ξ_2, ξ_3) to (A.4) on $[\theta, T-d]$. Assume $\theta > T-2d$. From Lemma A.2, ξ_1, ξ_2 and ξ_3 are Lipschitz in each variable and thus admit a limit, when $t \rightarrow \theta$, which is Lipschitz. Therefore, the argument of Lemma (A.1) can be repeated to extend the existence and uniqueness of the solution of system (A.4) on $[\xi, T-d]$ for $T-2d \leq \xi < \theta$. As a result, we necessarily have $\theta = T-2d$. It remains to prove that $0 < \xi_1$. For this, note that since ξ_1 is solution to (A.4), we have

$$\|\xi_1 - 1\|_\infty \leq \frac{d}{\sigma^2} \sup_{t \in [T-2d, T-d]} |\xi_2(t, 0)|^2. \tag{A.9}$$

By injecting the boundary condition (A.2) into the system (A.4), one notes that $t \in [T - 2d, T - d] \mapsto \xi_2(t, 0)$ is solution to

$$\xi_2(t, 0) = b - \sigma^{-2} \int_t^{T-d} \xi_2(x, 0) \xi_3(x, t - x, 0) dx, \quad T - 2d \leq t \leq T - d.$$

Or, for every $t \in [T - 2d, T - d]$, $f_t : x \in [t, T - d] \mapsto f_t(x) := \xi_3(x, t - x, 0)$ takes only positive values as f_t is solution to the system

$$\begin{aligned} f_t(x) &= b^2 - \sigma^{-2} \int_x^{T-d} f_t(u) \xi_3(u, 0, x - u) du, \quad x \in [t, T - d], \\ f_t(T - d) &= b^2, \end{aligned}$$

which can be proven to admit, through a contraction proof in the Banach space $C([t, T - d], \mathbb{R})$, a unique positive solution since ξ and its derivatives are bounded. Similarly, we also have $\xi_2(t, 0) \geq 0$ for any $t \in [T - 2d, T - d]$. As a result, we have $\text{sign}(\xi_2) = \text{sign}(b)$ and

$$\sup_{\substack{t \in \\ [T-2d, T-d]}} |\xi_2(t, 0)| \leq |b|. \quad (\text{A.10})$$

Consequently, (A.9) and (A.10) yield that for any $T - 2d \leq t \leq T - d$, we have $\xi_1 \geq 1 - d \left(\frac{b}{\sigma}\right)^2 = a_2 > 0$ as $\mathcal{N}(d, b, \sigma)$ is assumed to be greater than 2. \square

Finally, by setting $P_{11}(t) = \xi_1(t)$, $P_{12}(t, s) = \xi_2(t, s)$, $P_{22}(t, s, r) = \xi_3(t, s, r)$ and $P_{\hat{2}2}(t, s) = \xi_1(t + s + d)$ for any $(t, s, r) \in [T - 2d, T - d] \times [-d, 0]^2$, Lemma A.3 yields the existence and uniqueness of a solution P to (2.9)-(2.10)-(2.11) in the sense of definition 2.6 on $[T - 2d, T - d]$. The concatenation of the unique solution of (2.9)-(2.10)-(2.11) on $[T - d, T]$ and $[T - 2d, T - d]$ leads to a unique solution on $[T - 2d, T]$.

A.3 From slice $[T - nd, T]$ to $[T - (n + 1)d, T]$

Let n be an integer such that $2 \leq n < \mathcal{N}(d, b, \sigma)$. Assume that there exists a solution P to (2.9)-(2.10)-(2.11) in the sense of Definition 2.6 on $[T - nd, T]$ such that $0 < a_n \leq P_{11}(t) \leq 1$, for any $t \geq T - nd$. Recall the Definition (3.7) of $(a_n)_{n \geq 0}$. Consider the following system on $[T - (n + 1)d, T - nd] \times [-d, 0]^2$

$$\begin{aligned} P_{11}(t) &= P_{11}(T - nd) - \int_t^{T-nd} \frac{P_{12}(x, 0)^2}{\sigma^2 P_{11}(x + d)} dx, \\ P_{12}(t, s) &= b P_{11}((T - nd) \wedge (t + s + d)) - \int_t^{(T-nd) \wedge (t+s+d)} \frac{P_{12}(x, 0) P_{22}(x, t + s - x, 0)}{\sigma^2 P_{11}(x + d)} dx, \\ P_{22}(t, s, r) &= b P_{12}((T - nd) \wedge (t + s \wedge r + d), (s - r) \vee (r - s) - d) \\ &\quad - \int_t^{(T-nd) \wedge (t+s \wedge r+d)} \frac{P_{22}(x, t + s - x, 0) P_{22}(x, 0, t + r - x)}{\sigma^2 P_{11}(x + d)} dx. \end{aligned} \quad (\text{A.11})$$

Note that this system is the same as (A.4), the only difference being the term $x \in [T - (n + 1)d, T - nd] \mapsto P_{11}(x + d)$ which comes from the previous slice $[T - nd, T - (n - 1)d]$. Therefore, it can be considered as a positive continuous coefficient by induction hypothesis. As result, existence and uniqueness on $[T - (n + 1)d, T - nd]$ can be proven in the same fashion as in Lemmas A.1-A.2-A.3. It remains to prove that $P_{11}(t) \geq a_{n+1}$ for any $t \in [T - (n + 1)d, T - nd]$. As in Lemma A.3, and by using the induction hypothesis, we have

$$|P_{12}(t, -d)| \leq |bP_{11}(T - nd)| \leq |b|, \quad t \in [T - (n + 1)d, T - nd]. \quad (\text{A.12})$$

Furthermore, P_{11} satisfies (A.11) on $[T - (n + 1)d, T - nd]$, which, combined with $P_{11} \geq a_n$ on $[T - nd, T - (n - 1)d]$ and (A.12) yields

$$\begin{aligned} P_{11}(t) &\geq P_{11}(T - nd) - \frac{d}{a_n} \left(\frac{b}{\sigma} \right)^2 \\ &\geq a_n - \frac{d}{a_n} \left(\frac{b}{\sigma} \right)^2 = a_{n+1} > 0, \end{aligned}$$

for any $t \in [T - (n + 1)d, T - nd]$, which ends the proof.