A Class of Finite-Dimensional Numerically Solvable McKean-Vlasov Control Problems

Alessandro Balata, Côme Huré, Mathieu Laurière, Huyên Pham, Isaque Pimentel

A CLASS OF FINITE-DIMENSIONAL NUMERICALLY SOLVABLE MCKEAN-VLASOV CONTROL PROBLEMS *

Alessandro Balata[1], Côme Huré[2], Mathieu Laurière[3], Huyên Pham[4] and Isaque Pimentel[5]

Abstract. We address a class of McKean-Vlasov (MKV) control problems with common noise, called polynomial conditional MKV, and extending the known class of linear quadratic stochastic MKV control problems. We show how this polynomial class can be reduced by suitable Markov embedding to finite-dimensional stochastic control problems, and provide a discussion and comparison of three probabilistic numerical methods for solving the reduced control problem: quantization, regression by control randomization, and regress later methods. Our numerical results are illustrated on various examples from portfolio selection and liquidation under drift uncertainty, and a model of interbank systemic risk with partial observation.

Keywords: McKean-Vlasov control, polynomial class, quantization, regress later, control randomization.

*

[1] University of Leeds, Leeds, United Kingdom
e-mail: A.Balata@leeds.ac.uk
[2] Univ. Paris Diderot - LPMA, Paris, France;
e-mail: hure@lpsm.paris
[3] ECNU-NYU Institute of Mathematical Sciences, NYU Shanghai, Shanghai, China;
e-mail: mathieu.lauriere@nyu.edu
[4] Univ. Paris Diderot - LPMA, Paris, France;
e-mail: pham@math.univ-paris-diderot.fr
[5] CMAP, Ecole Polytechnique, Palaiseau, France;
e-mail: isaque.santa-brigida-pimentel@polytechnique.edu
, 2018

1. Introduction

The optimal control of McKean-Vlasov (also called mean-field) dynamics is a rather new topic in the area of stochastic control and applied probability, which has been knowing a surge of interest with the emergence of the mean-field game theory. It is motivated on one hand by the asymptotic formulation of cooperative equilibrium for a large population of particles (players) in mean-field interaction, and on the other hand from control problems with cost functional involving nonlinear functional of the law of the state process (e.g. the mean-variance portfolio selection problem or risk measure in finance).

In this paper, we are interested in the context of McKean-Vlasov control (MKV) problem under partial observation and common noise, whose formulation is described as follows. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with two independent Brownian motions B and W^0, let us consider the controlled stochastic McKean-Vlasov dynamics in $\mathbb{R}^n$:

$$dX_s = b(X_s, \mathbb{P}^{W^0}_{X_s}, \alpha_s)ds + \sigma(X_s, \mathbb{P}^{W^0}_{X_s}, \alpha_s)dB_s + \sigma_0(X_s, \mathbb{P}^{W^0}_{X_s}, \alpha_s)dW^0_s \tag{1}$$

where $\mathbb{P}^{W^0}_{X_s}$ denotes the conditional distribution of X_s given W^0 (or equivalently given $\mathcal{F}^0_s$ where $\mathbb{F}^0 = (\mathcal{F}^0_t)_t$ is the natural filtration generated by W^0), and the control α is $\mathbb{F}^0$-progressive valued in some Polish space A. The cost functional over a finite horizon T associated to the stochastic McKean-Vlasov equation (1) (sometimes called conditional McKean-Vlasov equation) for a control process α, is

$$J(\alpha) = \mathbb{E}\Big[\int_0^T f(X_t, \mathbb{P}^{W^0}_{X_t}, \alpha_t)dt + g(X_T, \mathbb{P}^{W^0}_{X_T})\Big],$$

and the objective is to minimize over an admissible set $\mathcal{A}$ of control processes the cost functional:

$$V_0 = \inf_{\alpha\in\mathcal{A}} J(\alpha). \tag{2}$$

Notice that classical partial observation control problem (without McKean-Vlasov dependence on the coefficients) arises as a particular case of (1)-(2). We refer to the introduction in [19] for the details.

Let us recall from [19] the dynamic programming equation associated to the conditional McKean-Vlasov (MKV) control problem (2). We start by defining a suitable dynamic version of this problem. Let us consider $\mathcal{F}_0$ a sub σ-algebra of $\mathcal{F}$ independent of B, W^0, and denote by $\mathcal{P}_2(\mathbb{R}^n)$ the set of all probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ with a finite second-order moment, endowed with the 2-Wasserstein metric $\mathcal{W}_2$. It is assumed w.l.o.g. that $\mathcal{F}_0$ is rich enough in the sense that $\mathcal{P}_2(\mathbb{R}^n) = \{\mathcal{L}(\xi) : \xi \in L^2(\mathcal{F}_0; \mathbb{R}^n)\}$, where $\mathcal{L}(\xi)$ denotes the law of ξ. Given a control $\alpha \in \mathcal{A}$, we consider the dynamic version of (1) starting from $\xi \in L^2(\mathcal{F}_0; \mathbb{R}^n)$ at time $t \in [0, T]$, and written as:

$$\begin{aligned} X^{t,\xi,\alpha}_s &= \xi + \int_t^s b(X^{t,\xi,\alpha}_u, \mathbb{P}^{W^0}_{X^{t,\xi,\alpha}_u}, \alpha_u)du + \int_t^s \sigma(X^{t,\xi,\alpha}_u, \mathbb{P}^{W^0}_{X^{t,\xi,\alpha}_u}, \alpha_u)dB_u \\ &\quad + \int_t^s \sigma_0(X^{t,\xi,\alpha}_u, \mathbb{P}^{W^0}_{X^{t,\xi,\alpha}_u}, \alpha_u)dW^0_u \quad t \leqslant s \leqslant T. \end{aligned}$$

Let us then define the dynamic cost functional:

$$J(t, \xi, \alpha) = \mathbb{E}\Big[\int_t^T f(X^{t,\xi,\alpha}_s, \mathbb{P}^{W^0}_{X^{t,\xi,\alpha}_s}, \alpha_s)ds + g(X^{t,\xi,\alpha}_T, \mathbb{P}^{W^0}_{X^{t,\xi,\alpha}_T})\Big],$$

for $(t, \xi) \in [0, T] \times L^2(\mathcal{F}_0; \mathbb{R}^n), \alpha \in \mathcal{A}$, and notice by the law of conditional expectations, and as α is $\mathbb{F}^0$-progressive that

$$J(t, \xi, \alpha) = \mathbb{E}\Big[\int_t^T \hat{f}(\mathbb{P}^{W^0}_{X^{t,\xi,\alpha}_s}, \alpha_s)ds + \hat{g}(\mathbb{P}^{W^0}_{X^{t,\xi,\alpha}_s})\Big],$$

where $\hat{f} : \mathcal{P}_2(\mathbb{R}^n) \times A \to \mathbb{R}$, $\hat{g} : \mathcal{P}_2(\mathbb{R}^n) \to \mathbb{R}$ are defined by

$$\hat{f}(\mu, a) := \mu(f(., \mu, a)) := \int_{\mathbb{R}^n} f(x, \mu, a)\mu(dx), \tag{3}$$

$$\hat{g}(\mu) := \mu(g(., \mu)) := \int_{\mathbb{R}^n} g(x, \mu)\mu(dx). \tag{4}$$

Moreover, notice that the conditional law of $X_s^{t,\xi,\alpha}$ given W^0 depends on ξ only through its law $\mathcal{L}(\xi)$, and we can then define for $\alpha \in \mathcal{A}$:

$$\rho_s^{t,\mu,\alpha} := \mathbb{P}^{W^0}_{X_s^{t,\xi,\alpha}}, \quad \text{for } t \leqslant s,\ \mu = \mathcal{L}(\xi) \in \mathcal{P}_2(\mathbb{R}^n).$$

Therefore, the dynamic cost functional $J(t, \xi, \alpha)$ depends on $\xi \in L^2(\mathcal{F}_0; \mathbb{R}^n)$ only through its law $\mathcal{L}(\xi)$, and by misuse of notation, we write $J(t, \mu, \alpha) = J(t, \xi, \alpha)$ when $\mu = \mathcal{L}(\xi)$. We then consider the value function for the conditional McKean-Vlasov control problem (2), defined on $[0, T] \times \mathcal{P}_2(\mathbb{R}^n)$ by

$$v(t, \mu) = \inf_{\alpha \in \mathcal{A}} J(t, \mu, \alpha) = \inf_{\alpha \in \mathcal{A}} \mathbb{E}\Big[\int_t^T \hat{f}(\rho_s^{t,\mu,\alpha}, \alpha_s)ds + \hat{g}(\rho_T^{t,\mu,\alpha})\Big], \tag{5}$$

and notice that at time $t = 0$, when $\xi = x_0$ is a constant, then $V_0 = v(0, \delta_{x_0})$.

It is shown in [19] that dynamic programming principle (DPP) for the conditional McKean-Vlasov control problem (5) holds: for $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^n)$,

$$v(t, \mu) = \inf_{\alpha \in \mathcal{A}} \mathbb{E}\Big[\int_t^\theta \hat{f}(\rho_s^{t,\mu,\alpha}, \alpha_s)ds + v(\theta, \rho_\theta^{t,\mu,\alpha})\Big],$$

for any $\mathbb{F}^0$-stopping time θ valued in $[t, T]$. Next, by relying on the notion of differentiability with respect to probability measures introduced by P.L. Lions [11] (see also the lecture notes [4]) and the chain rule (Itô's formula) along flow of probability measures (see [3], [7]), we derive the Hamilton-Jacobi-Bellman equation for v:

$$\begin{cases} \partial_t v + \inf_{a \in A} \Big[\hat{f}(\mu, a) + \mu\big(\mathbb{L}^a v(t, \mu)\big) + \mu \otimes \mu\big(\mathbb{M}^a v(t, \mu)\big)\Big] = 0, \quad (t, \mu) \in [0, T) \times \mathcal{P}_2(\mathbb{R}^n), \\ v(T, \mu) = \hat{g}(\mu), \quad \mu \in \mathcal{P}_2(\mathbb{R}^n), \end{cases} \tag{6}$$

where for $\phi \in \mathcal{C}_b^2(\mathcal{P}_2(\mathbb{R}^n))$, $a \in A$, and $\mu \in \mathcal{P}_2(\mathbb{R}^n)$, $\mathbb{L}^a\phi(\mu)$ is the function $\mathbb{R}^n \to \mathbb{R}$ defined by

$$\mathbb{L}^a\phi(\mu)(x) := \partial_\mu\phi(\mu)(x).b(x, \mu, a) + \frac{1}{2}\mathrm{tr}\big(\partial_x\partial_\mu\phi(\mu)(x)(\sigma\sigma^\intercal + \sigma_0\sigma_0^\intercal)(x, \mu, a)\big), \tag{7}$$

and $\mathbb{M}^a\phi(\mu)$ is the function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$\mathbb{M}^a\phi(\mu)(x, x') := \frac{1}{2}\mathrm{tr}\big(\partial_\mu^2\phi(\mu)(x, x')\sigma_0(x, \mu, a)\sigma_0^\intercal(x', \mu, a)\big). \tag{8}$$

The Hamilton-Jacobi-Bellman (HJB) equation (6) is a fully nonlinear PDE in the infinite dimensional Wasserstein space, and do not have in general explicit solution except the notable important class of linear-quadratic McKean-Vlasov control problem. Numerical resolution for MKV control problem or equivalently for the associated HJB equation is a challenging problem due to the nonlinearity of the optimization problem and the infinite dimensional feature of the Wasserstein space. In this work, our purpose is to investigate some classes of MKV control problems, which can be reduced to finite dimensional problems in view of numerical resolution.

2. Polynomial McKean-Vlasov control problem

2.1. Main assumptions

We make two kinds of assumptions on the coefficients of the model: one on the dependence on x and the other on the dependence on μ.

Assumptions: dependence on x: we consider a class of models where the coefficients of the MKV equation are linear w.r.t. the state variable X, i.e., in the form

$$\left\{\begin{array}{rcl} b(x,\mu,a) & = & b_0(\mu,a)+b_1(\mu,a)x, \\ \sigma(x,\mu,a) & = & \vartheta_0(\mu,a)+\vartheta_1(\mu,a)x, \\ \sigma_0(x,\mu,a) & = & \gamma_0(\mu,a)+\gamma_1(\mu,a)x, \end{array}\right. \tag{1}$$

while the running and terminal cost functions are polynomial in the state variable in the sense that they are in the form

$$\begin{array}{rcl} f(x,\mu,a) & = & f_0(\mu,a)+f_1(\mu,a)x+\sum_{k=2}^{p} f_k(\mu,a)|x|^k, \\ g(x,\mu) & = & g_0(\mu)+g_1(\mu)x+\sum_{k=2}^{p} g_k(\mu)|x|^k, \end{array}$$

for some integer $p \geqslant 2$.

Assumptions: dependence on μ: we assume that all the coefficients depend on μ through its first p moments, i.e., in the form

$$\left\{\begin{array}{llll} b_0(\mu,a) = \bar{b}_0(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a), & & b_1(\mu,a) = \bar{b}_1(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a) \\ \vartheta_0(\mu,a) = \bar{\vartheta}_0(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a), & & \vartheta_1(\mu,a) = \bar{\vartheta}_1(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a) \\ \gamma_0(\mu,a) = \bar{\gamma}_0(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a), & & \gamma_1(\mu,a) = \bar{\gamma}_1(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a) \\ f_k(\mu,a) = \bar{f}_k(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a), & g_k(\mu) = \bar{g}_k(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p), & k=0,\ldots,p, \end{array}\right. \tag{2}$$

where, given $\mu \in \mathcal{P}_p(\mathbb{R}^n)$, we denote by

$$\bar{\mu} := \int_{\mathbb{R}^n} x\mu(dx), \qquad \bar{\mu}_k := \int_{\mathbb{R}^n} |x|^k \mu(dx), \quad k=2,\ldots,p.$$

Notice that in this case, the functions $\hat{f}$ and $\hat{g}$ defined in (3)-(4) are given by

$$\begin{array}{rcl} \hat{f}(\mu,a) & = & \bar{f}_0(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a)+\bar{f}_1(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a)\bar{\mu}+\sum_{k=2}^{p}\bar{f}_k(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a)\bar{\mu}_k \\ & =: & \bar{f}(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p,a) \\ \hat{g}(\mu) & = & \bar{g}_0(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p)+\bar{g}_1(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p)\bar{\mu}++\sum_{k=2}^{p}\bar{g}_k(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p) \\ & =: & \bar{g}(\bar{\mu},\bar{\mu}_2,\ldots,\bar{\mu}_p). \end{array}$$

Remark 0.1. A more general class of running and terminal cost functions, would be to consider multi-polynomial of degree p functions f and g in the form

$$f(x,\mu,a) = \sum_{|\mathbf{k}|=0}^{p} f_{\mathbf{k}}(\mu,a)x^{\mathbf{k}}, \qquad g(x,\mu) = \sum_{|\mathbf{k}|=0}^{p} g_{\mathbf{k}}(\mu)x^{\mathbf{k}},$$

where we use multi-index notations $\mathbf{k} = (k_1, \ldots, k_n) \in \mathbb{N}^n$, $|\mathbf{k}| = k_1 + \ldots + k_n$, and $x^{\mathbf{k}} = x_1^{k_1} \ldots x_n^{k_n}$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Given $\mu \in \mathcal{P}_p(\mathbb{R}^n)$, we denote by

$$\mu^{\mathbf{k}} \;=\; \int_{\mathbb{R}^n} x^{\mathbf{k}} \mu(dx),$$

and we assume that all the coefficients would depend on μ through $\mu^{\mathbf{k}}$, $1 \leqslant |k| \leqslant p$. □

2.2. **Markovian embedding**

Given the controlled process $X = X^\alpha$ solution to the stochastic McKean-Vlasov dynamics (1), denote by

$$\bar{X}_t \;=\; \mathbb{E}[X_t | W^0], \qquad Y_t^k \;=\; \mathbb{E}[|X_t|^k | W^0], \quad k = 2, \ldots, p.$$

To alleviate the notations, let us assume that $n = 1$ (otherwise multi-indices should be used). From the linear/polynomial assumptions (1)-(2), by Itô's formula and taking conditional expections, we can derive the dynamics of $(\bar{X}, Y^2, \ldots, Y^p)$ as

$$\left\{ \begin{array}{rcl} d\bar{X}_t &=& \bar{B}(\bar{X}_t, Y_t^2, \ldots, Y_t^p, \alpha_t)dt + \bar{\Sigma}(\bar{X}_t, Y_t^2, \ldots, Y_t^p, \alpha_t)dW_t^0, \\ dY_t^k &=& B_k(\bar{X}_t, Y_t^2, \ldots, Y_t^p, \alpha_t)dt + \Sigma_k(\bar{X}_t, Y_t^2, \ldots, Y_t^p, \alpha_t)dW_t^0, \quad k = 2, \ldots, p, \end{array} \right. \tag{3}$$

where

$$\begin{array}{rcl} \bar{B}(\bar{x}, y^2, \ldots, y^p, a) &=& \bar{b}_0(\bar{x}, y^2, \ldots, y^p, a) + \bar{b}_1(\bar{x}, y^2, \ldots, y^p, a)\bar{x} \\ \bar{\Sigma}(\bar{x}, y^2, \ldots, y^p, a) &=& \bar{\gamma}_0(\bar{x}, y^2, \ldots, y^p, a) + \bar{\gamma}_1(\bar{x}, y^2, \ldots, y^p, a)\bar{x}, \\ B_k(\bar{x}, y^2, \ldots, y^p, a) &=& k\bar{b}_0(y^2, \ldots, y^p, a)y^{k-1} + k\bar{b}_1(y^2, \ldots, y^p, a)y^k \\ && + \dfrac{k(k-1)}{2}(\bar{\vartheta}_0(y^2, \ldots, y^p, a))^2 y^{k-2} + \dfrac{k(k-1)}{2}(\bar{\vartheta}_1(y^2, \ldots, y^p, a))^2 y^k \\ && + k(k-1)\bar{\vartheta}_0(y^2, \ldots, y^p, a)\bar{\vartheta}_1(y^2, \ldots, y^p, a)y^{k-1} \\ && + \dfrac{k(k-1)}{2}(\bar{\gamma}_0(y^2, \ldots, y^p, a))^2 y^{k-2} + \dfrac{k(k-1)}{2}(\bar{\gamma}_1(y^2, \ldots, y^p, a))^2 y^k \\ && + k(k-1)\bar{\gamma}_0(y^2, \ldots, y^p, a)\bar{\gamma}_1(y^2, \ldots, y^p, a)y^{k-1} \\ \Sigma_k(\bar{x}, y^2, \ldots, y^p, a) &=& k\left(\bar{\gamma}_0(y^2, \ldots, y^p, a)y^{k-1} + \bar{\gamma}_1(y^2, \ldots, y^p, a)y^k\right) \end{array}$$

while the cost functional is written as

$$J(\alpha) \;=\; \mathbb{E}\Big[\int_0^T \bar{f}(\bar{X}_t, Y_t^2, \ldots, Y_t^p, \alpha_t)dt + \bar{g}(\bar{X}_T, Y_T^2, \ldots, Y_T^p) \Big]. \tag{4}$$

The McKean-Vlasov control problem is then reduced in this polynomial framework into a finite-dimensional control problem with $\mathbb{F}^0$-adapted controlled variables $(\bar{X}, Y^2, \ldots, Y^p)$. In the next section, we describe three probabilistic numerical methods for solving finite-dimensional stochastic control problems, and will apply in the last section each of these methods to three applications arising from polynomial MKV control problems under partial observation and common noise.

3. Numerical methods

To explain our numerical methods for the resolution of (3)-(4), we choose the following setting.

Let us introduce the process Z controlled by an adapted process α taking values in A, solution to

$$dZ_t^\alpha = b(Z_t^\alpha, \alpha_t)dt + \sigma_0(Z_t^\alpha, \alpha_t)dW_t^0$$

and

$$J(t, z, \alpha) = \mathbb{E}\left[\int_t^T f(Z_t^\alpha, \alpha_t)dt + f(Z_T^\alpha) \right].$$

Introducing now a time discretisation $0 = t_0, t_1, \ldots, t_N = T$ we can write the Euler approximation of the SDE governing the process Z_t^α:

$$Z_{t_{n+1}}^\alpha = Z_{t_n}^\alpha + b(Z_{t_n}^\alpha, \alpha_{t_n})\Delta t + \sigma_0(Z_{t_n}^\alpha, \alpha_{t_n})\Delta W_{t_n}^0 \tag{1}$$

and the discrete equivalent of $J(t, z, \alpha)$:

$$J(t_n, z, \alpha) = \mathbb{E}\left[\sum_{s=n}^{N} f(Z_{t_s}^\alpha, \alpha_{t_s})\Delta t + f(Z_{t_N}^\alpha)\right].$$

We can now give an alternative representation of the value function $V(t_n, z) = \sup_{(\alpha_{t_s})_{s=n}^N \in \mathcal{A}} \{J(t_n, z, \alpha)\}$ through the dynamic programming equation; given the known terminal condition $g(z)$:

$$\begin{cases} V(T_N, z) = g(z) \\ V(t_n, z) = \sup_\alpha \left\{ f(Z_{t_n}^\alpha, \alpha)\Delta t + \mathbb{E}_\alpha\big[V(t_{n+1}, Z_{t_{n+1}})\big|Z_{t_n} = z\big] \right\} \end{cases} \tag{2}$$

The dynamic programming equation 2 inspires a numerical methods that approximate the value function iteratively backward in time, starting from the terminal condition. The main difficulty in implementing such approach lies in the estimation of conditional expectations $\mathbb{E}_\alpha\big[V(t_{n+1}, Z_{t_{n+1}}^\alpha)\big|Z_{t_n} = z\big]$. In the present section we will briefly introduce three numerical methods that we ought to test in the task of solving CMKV problems. Two of these methods belong to the class of Regression Monte Carlo techniques, a family of algorithms whose effectiveness highly relies on the choice of the basis functions used to project future time value functions; the third algorithm, quantization, approximate the controlled process $Z_{t_n}^\alpha$ with a particular finite state Markov chain for which expectations can be approximated quickly.

3.1. **Regression Monte Carlo**

As introduced above, the family of Regression Monte Carlo algorithms is based on the idea of approximating the conditional expectation $\mathbb{E}_\alpha\big[V(t_{n+1}, Z_{t_{n+1}}^\alpha)\big|Z_{t_n} = z\big]$ by projecting the next step value function $V(t_{n+1}, Z_{t_{n+1}}^\alpha)$ onto a finite collection of functions $\{\phi_k\}$ in the basis of $L^2(Z_{t_n})$.

In the simpler, uncontrolled case, the method works as follow: starting from the known terminal condition, we approximate $\mathbb{E}\big[g(Z_{t_N})\big|Z_{t_{N-1}} = z\big] \sim \sum_{k=1}^K \beta_k \phi_k(Z_{t_{N-1}})$, where the regression coefficients $\beta^N = \{\beta_k\}_{k=1}^K$ are computed as follow:

$$\beta^N = \underset{\beta}{\operatorname{argmin}} \left\{ \mathbb{E}\Big[\big(\mathbb{E}\big[g(Z_{t_N})\big|Z_{t_{N-1}} = z\big] - \sum_{k=1}^K \beta_k \phi_k(Z_{t_{N-1}})\big)^2 \Big| Z_{t_{N-1}} = z \Big] \right\}. \tag{3}$$

The procedure is iterated backward in time substituting $V(t_{n+1}, Z_{t_{n+1}}^\alpha)$, which is known at time n, with the terminal condition g in the equation (3). In practical implementations the outer expectation in (3) is approximated via Monte Carlo implementations, from a simulated set of M samples $\{Z_{t_s}\}_{n=0,m=1}^{N,M}$ obtained from the Euler scheme (1) from a given initial condition.

This approach, also known as regress now, unfortunately can not be directly applied to the controlled case; intuitively this is the case because, since the control is initially unknown, it is impossible to simulate the set of samples to be used to approximate conditional expectations and, in turn, compute the optimal control.

Performance iteration. The performance iteration approach, was first introduced in [12] under the name of policy iteration and gained most of the popularity Regression Monte Carlo algorithms enjoy today. This technique builds on the value iteration method, substituting the iteration over the value function (2),

with an iteration over pathwise performances. In practice we solve the following problem:

$$\begin{cases} \mathcal{V}(T_N, z) = g(z) \\ \mathcal{V}(t_n, z) = \sum_{i=n}^{N} f(Z_{t_i}^{\alpha}, \hat{\alpha}_{t_i})\Delta t \end{cases}$$
$$\text{where } \hat{\alpha}_{t_i} = arg\sup_{\alpha}\Big\{f(Z_{t_i}^{\alpha}, \alpha)\Delta t + \mathbb{E}_{\alpha}\big[V(t_{n+1}, Z_{t_{n+1}})\big|Z_{t_n} = z\big]\Big\},$$

where $\mathbb{E}_{\alpha}\big[V(t_{n+1}, Z_{t_{n+1}})\big|Z_{t_n} = z\big] = \mathbb{E}_{\alpha}\big[\mathcal{V}(t_{n+1}, Z_{t_{n+1}})\big|Z_{t_n} = z\big] \approx \sum_{k=1}^{K} \beta_k \phi_k(Z_{t_n})$ with

$$\beta^n = \underset{\beta}{\operatorname{argmin}}\Big\{\mathbb{E}\Big[\big(\mathbb{E}\big[V(t_{n+1}, Z_{t_{n+1}})\big|Z_{t_n} = z\big] - \sum_{k=1}^{K} \beta_k \phi_k(Z_{t_n})\big)^2\Big|Z_{t_n} = z\Big]\Big\}.$$

We retrieve the value function via Monte Carlo average: $V(t_n, z) = \mathbb{E}[\mathcal{V}(t_0, z)|Z_{t_0} = z] \approx \frac{1}{M}\sum_{m=1}^{M} \mathcal{V}^m(t_0, z)$. Performance iteration, albeit slower and subject to higher variance in the regression step, is able to limit the backward propagation of the error improving the quality of the estimations. For further details see [12].

3.1.1. *Regress Later*

In this section we present a regress-later idea in which conditional expectation with respect to (Z_{t_n}) are computed in two stages. First, a conditional expectation with respect to $(Z_{t_{n+1}})$ is approximated in a regression step by a linear combination of basis functions of $(Z_{t_{n+1}})$. Then, analytical formulas are applied to condition this linear combination of functions of future values on present values (Z_{t_n}). For further details see [8], [13] or [1].

Unlike the traditional regress-now method for approximating conditional expectations, the regress-later approach imposes conditions on basis functions:

Assumption 1. *For each basis function ϕ_k, $k = 1, \ldots, K$, the conditional expectation*

$$\hat{\phi}_k^n(z, a) = \mathbb{E}[\phi_k(Z_{t_{n+1}})|Z_{t_n} = z, \alpha_n = a]$$

can be computed analytically.

We will now present regress-later solution to value iteration procedure. Notice that a completely analogous approach can be used in the case of performance iteration. Assume that at time $n+1$ the value function $V(t_{n+1}, \cdot)$ has been computed for a set of training points $\{Z_{t_{n+1}}^m\}_{m=1}^M$. We perform a regression to approximate $V(t_{n+1}, \cdot)$ with a linear combination of basis functions:

$$V(n+1, x) \approx \sum_{k=1}^{K} \beta_k^{n+1} \phi_k(x),$$

where

$$\beta^{n+1} = \underset{\beta \in \mathbb{R}^K}{\operatorname{argmin}}\Big\{\sum_{m=1}^{M}\Big[V(t_{n+1}, Z_{t_{n+1}}^m) - \sum_{k=1}^{K} \beta_k^{n+1}\phi_k(Z_{n+1}^m)\Big]^2\Big\}. \tag{4}$$

Moving now to time t_n we would like to compute an optimal control $\alpha_n(Z_{t_{n+1}}^m)$ which, further, determines the value function. We select a set of training points $\{Z_{t_{n+1}}^m\}_{m=1}^M$, which can be generated independently from $\{Z_{n+1}^m\}_{m=1}^M$, therefore, removing the limitation of the regress-now approach for which points at different times must be linked through their true dynamics.

Using the regress-later approximation of the conditional expectation and recalling Assumption 1 we obtain the optimal control α_n^m corresponding to the point (Z_n^m),

$$\alpha_n^m = \underset{a \in \mathcal{U}}{\operatorname{argmax}}\Big\{f(n, Z_{t_n}^m, a) + \sum_{k=1}^{K} \beta_k^{n+1}\hat{\phi}_k^n\big(Z_{t_n}^m, a\big)\Big\}.$$

Notice that we are able to exploit the linearity of conditional expectations because β is a constant with respect to $\mathcal{F}_{t_n}$. An approximation of the value function at time n is the computed as

$$V(n, Z_n^m) = f(n, Z_n^m, \alpha_n^m) + \sum_{k=1}^{K} \beta_k^{n+1} \hat{\phi}_k^n \big(Z_{t_n}^m, \alpha_n^m\big).$$

Algorithm 1 Regress-later Monte Carlo algorithm (RLMC) - Value iteration

Inputs:

- M: number of training points,
- μ: distribution of training points,
- K: number of basis functions,
- $\{\phi_k\}_{k=1}^K$: family of basis functions,
- x: input of the value function

Pre-compute the inverse of the covariance matrix $\mathcal{A}$
Generate i.i.d. training points $\{\tilde{Z}_N^m\}_{m=1}^M$ accordingly to the distribution μ.
Initialise the value function $\tilde{V}(N, \tilde{Z}_N^m) = g(\tilde{Z}_N^m), \quad \forall m = 1, \dots, M$
for $n = N-1$ to 1 **do**
    $\hat{\beta}_{n+1} = \mathcal{A}^{-1} \frac{1}{M} \sum_{m=1}^{M} \Big[\hat{V}(n+1, \tilde{Z}_{n+1}^m)\phi(\tilde{Z}_{n+1}^m)\Big]$
    Generate a new layer of i.i.d. training points $\{\tilde{Z}_n^m\}_{m=1}^M$ accordingly to the distribution μ.
    For all m do

$$\tilde{V}(N, \tilde{Z}_N^m) = \sup_{a \in A} \Big\{ f(n, \tilde{Z}_n^m, a) + \sum_{k=1}^{K} \beta_k^{n+1} \hat{\phi}_k(\tilde{Z}_n^m, a) \Big\}$$

Evaluate the policy to obtain $\hat{V}$

Outputs: $\{\hat{\beta}_n^k\}_{n,k=1}^{N,K}$, $\hat{V}(0, x)$

3.1.2. *Control randomization*

Control randomisation was introduced in [10]; it differs from the previous methods in that the control becomes a state variable and it is simulated along trajectories of $(Z_t)_{t=1}^N$. We denote the initial random control by $\{\tilde{\alpha}_n^m\}_{n,m=1}^{N,M}$. In the case of value iteration, $\{V(n+1, Z_{n+1}^m)\}_{m=1}^M$ is regressed against basis functions evaluated at the points $\{Z_n^m, \tilde{\alpha}_n^m\}_{m=1}^M$,i.e. $\mathbb{E}_{z,\alpha}\big[V(n+1, Z_{n+1})\big] \sim \sum_{k=1}^K \beta_k^n \phi(z, \alpha)$, where

$$\beta^n = \underset{\beta \in \mathbb{R}^K}{\operatorname{argmin}} \Big\{ \sum_{m=1}^{M} \Big[V(n+1, Z_{n+1}^m) - \sum_{k=1}^{K} \beta_k \phi_k(Z_n^m, \tilde{\alpha}_n^m)\Big]^2 \Big\}.$$

These regression basis functions are dependent now on the random control $\tilde{\alpha}_n$, in addition to Z_n so that the estimated continuation value will depend on the choice of the control (which is different on each sample trajectory).

An optimal control at time n given $Z_n = Z_n^m$ is approximated by the expression

$$\alpha_n^m = \underset{a \in \mathcal{A}}{\operatorname{argmax}} \Big\{ f(n, Z_n^m, a) + \hat{\mathbb{E}}_{n, Z_n^m, a}\big[V(n+1, Z_{n+1}))\big] \Big\}, \tag{5}$$

where, with a slight abuse of notation, we included in the approximate conditional expectation $\hat{\mathbb{E}}$ the dependence on the control α. In general multiple runs of the method could be needed to obtain precise estimates because the initial choice of the dummy control could drive the training points far from where the optimal control would have driven them. In practice, after having computed an approximated policy backward in time, such policy is used to drive M simulations of the process Z^α forward in time, which in turn produce control paths that can be fed as random control in a new backward procedure, leading to more accurate results.

Algorithm 2 Control randomization algorithm (CR) - Value iteration

Inputs:

- M: number of training points,
- μ: initial distribution of training points,
- K: number of basis functions,
- $\{\phi_k\}_{k=1}^K$: family of basis functions,
- x: input of the value function

1: Generate m trajectories, $\{\tilde{Z}_n^m\}_{n,m=1}^{N,M}$, starting with i.i.d. initial distribution μ.

2: Initialise the value function $\tilde{V}(N, \tilde{Z}_N^m) = g(\tilde{Z}_N^m), \quad m = 1, \dots, M$

3: **for** $n = N-1$ to 1 **do**

4: Approximate $\mathbb{E}[V(n+1, \tilde{Z}_{n+1}) \mid \tilde{Z}_n, \tilde{\alpha}_n]$ by regressing $\{V(n+1, \tilde{Z}_{n+1}^m)\}_m$ against $\{\tilde{Z}_n^m, \tilde{\alpha}_n^m\}_m$. Denote by $\{\hat{\beta}_{n+1}^k\}_{k=1}^K$ the family of regression coefficients.

5: Compute $\tilde{\alpha}(n, \tilde{Z}_n^m) := \text{argmax}_{a \in A} \left\{ f(n, \tilde{Z}_n^m, a)\, \delta t + \sum_{k=1}^K \hat{\beta}_{n+1}^k \phi_k(\tilde{Z}_n^m, a) \right\}$.

6: **Evaluate the policy to obtain** $\hat{V}$

Outputs: $\{\hat{\beta}_n^k\}_{n,k=1}^{N,K}$, $\hat{V}(0, x)$

3.2. **Quantization**

In this section we present a method that relies on Markovian quantization to solve control problems. Markovian quantization methods has been proven to be very efficient for solving control problems associated with high-dimensional processes (see e.g. [15]). We first recall the general idea of the quantization.

In the spirit of the Markov chain approximation method, we approximate the Euler scheme Z_{t_k} at every date $k \in \{0, ..., t_N\}$ by a process $\widehat{Z}_{t_k}$ taking finitely many states. Lets fix a control $(\alpha_{t_k})_{k \in \{0,..,N\}}$. At each discrete time t_k, we consider a grid $\Gamma_k = \left\{z_k^1, ..., z_k^{N_k}\right\}$ on the state space $\mathbb{R}^d$ of $Z_{t_k}^\alpha$. We denote by π_k the projection on the grid Γ_k, and define the quantized controlled process $\left(Z_{t_k}^\alpha\right)_k$ as follows:

$$\begin{cases} \widehat{Z}_0^\alpha = \bar{Z}_0^\alpha (= Z_0 = z_0) \\ \widehat{Z}_{k+1}^\alpha = \pi_{k+1}\Big(G_{\Delta t}\big(t_k, \widehat{Z}_k^\alpha, \alpha_k, \epsilon_{k+1}\big)\Big), k = 0, ..., N-1 \end{cases} \tag{6}$$

where $G_h(t, z, a, \epsilon) := z + b(z, a)\Delta t + \sigma_0(z, a)\sqrt{\Delta t}\varepsilon$ (see the notations in (1)). $\left(\widehat{Z}_k^\alpha\right)_k$ is called the quantized process associated to $\left(Z_{t_k}^\alpha\right)_k$. We now consider the stochastic control problem in discrete time:

$$\widehat{V}(n, z) = \inf_\alpha \mathbb{E}_{n,z}^\alpha \left[\sum_{k=n}^{N-1} \Delta t f\big(k, \hat{Z}_{t_k}, \alpha_k\big) + g\big(\hat{Z}_{t_N}\big) \right]. \tag{7}$$

One can find results in the literature that show that the quantized value function $\hat{v}$ converges to the value function v under reasonable conditions on the drift and volatility of the controlled process. (see e.g. [15]).

The value function associated to the quantized process $\left(\widehat{Z}_k^\alpha\right)_k$ can be recursively computed using the dynamic programming principle:

$$\begin{cases} \widehat{V}(N, z) = g(z) \text{ for } z \in \Gamma_N \\ \widehat{V}(n, z) = \inf\limits_{a \in A} \left[\Delta t f(n, z, a) + \mathbb{E}_{n,z}^a \big[\widehat{V}\big(n+1, \widehat{Z}_{n+1}\big) \big] \right]. \end{cases} \tag{8}$$

Notice that the conditional expectation in (8) is a finite sum. More precisely

$$\mathbb{E}_{n,z}^a \big[\widehat{V}\big(n+1, \widehat{Z}_{n+1}\big) \big] = \sum_{z' \in \Gamma_{n+1}} p_{zz'}(a) \widehat{V}(n+1, z'),$$

where we denoted by $p_{zz'}(a)$ the transition probabilities: $p_{zz'}(a) := \mathbb{P}_a\big(\hat{Z}_{k+1} = z_{k+1}^j \big| \hat{Z}_k = z\big)$. However, in general, there is no closed-form formulas to compute these conditional expectations. One needs to efficiently approximate these expectations to solve (8).
A natural way to approximate the conditional expectations is to quantize the noise ε. Doing so, the quantization algorithm becomes a two steps quantization algorithm: the first step is the quantization of (Z_t), and the second is the quantization of the noise.

In all the applications that we consider in the next section, the noise is a 1D-Brownian Motion W. So we just need to have an optimal grid Γ^1 of quantization for $\mathcal{N}(0,1)$ to quantize the noise. Indeed, if $\varepsilon \sim \sqrt{\Delta t}\mathcal{N}(0,1)$ then $\sqrt{\Delta t}\Gamma^1 =: \Gamma^\varepsilon$ is an optimal grid of quantization of ε, and the sequence $(\Gamma_i^t)_{i\leqslant N}$ is an optimal grid for $(W_{t_i})_{0\leqslant i\leqslant N}$. Note that this sequence is actually not optimal for the markovian quantized Brownian motion. The reader can find a procedure in [15] to build such optimal grids.
Doing a markovian quantization of the Brownian motion gives the following approximation for the conditional expectation:

$$\mathbb{E}_{n,z}^a\big[\widehat{V}\big(n+1, \widehat{Z}_{n+1}\big)\big] \approx \sum_{e\in\Gamma^\varepsilon} \mathbb{P}(\widehat{\varepsilon} = e)\widehat{V}\Big(n+1, \pi_{n+1}\big(G(n,z,a,e)\big)\Big)$$

This approximation is fast to compute. Indeed, when considering a 1D Brownian Motion, one just needs to take no more than 50 terms in Γ^ε to get a good quantization of the noise. So the conditional expectation can be approximated by a sum of only 50 terms. First, note that there are procedures in the literature to compute the optimal grids for the noise $\mathcal{N}_d(0,1)$ and the weights of the Voronoi cells. (see e.g. the CLVQ algorithm section in [16], or [14]). Secondly, note that this approximation is not continuous with respect to a. In some situations, it may be useful to have continuity with respect to the control. One can proceed as explained in the remarks for the Q algorithm designed for the portfolio liquidation problem at subsection 4.1.2.

Algorithm 3 Quantization algorithm (Q) - Value iteration

Inputs:

- Γ_k : a N_k-quantizer of $\bar{Z}_k$
- $\hat{Z}_k$: the quantization of $\bar{Z}_k$ on the grid Γ_k,
- Γ^ε : a quantizer for the noise.

1: Initialise the value function $\tilde{V}(N,z) = g(z), \quad \forall z \in \Gamma_N$
2: **for** $n = N-1$ to 1 **do**
3: Update the value function with:

$$\widehat{V}(n,z) = \inf_{a\in A}\left[f(n,z,a)\,\Delta t + \sum_{e\in\Gamma^\varepsilon} \mathbb{P}(\widehat{\varepsilon} = e)\widehat{V}\Big(n+1, \pi_{n+1}\big(G(n,z,a,e)\big)\Big)\right] \quad \forall z \in \Gamma_n$$

4: Define the optimal strategy $a^*(n,.z)_{z\in\Gamma_n}$ as the minimizers of (3):

$$\forall z \in \Gamma_n, \quad a^*(n,z) = \operatorname*{argmin}_{a\in A}\left[f(n,z,a)\,\delta t + \sum_{e\in\Gamma^\varepsilon} \mathbb{P}(\widehat{\varepsilon} = e)\widehat{V}\Big(n+1, \pi_{n+1}\big(G(n,z,a,e)\big)\Big)\right]$$

Outputs: $\big(a^*(n,z)_{z\in\Gamma_n}\big)_{1\leqslant n\leqslant N-1}$, $\hat{V}(0,x)$

4. Applications and numerical results

4.1. **Portfolio optimization under drift uncertainty**

4.1.1. *The model*

We consider a financial market model with one risk-free asset, assumed to be equal to one, and d risky assets of price process $S = (S^1, \ldots, S^d)$ governed

$$dS_t \quad = \quad \mathrm{diag}(S_t)(\beta_t dt + \sigma dB_t^0),$$

where B^0 is a d-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}^0)$, σ is the $d \times d$ invertible matrix volatility coefficient, assumed to be a known constant. However, the drift (β_t) of the asset (which is typically a diffusion process governed by another independent Brownian motion B) is unknown and unobservable like the Brownian motion B^0. The agent can actually only observe the stock prices S, and we denote by $\mathbb{F}^S$ the available information filtration $\mathbb{F}^S$, i.e., the filtration generated by S.

In this context, we shall consider two important classes of optimization problems in finance:

(1) *Portfolio liquidation.* We consider the problem of an agent (trader) who has to liquidate a large number y_0 of shares in some asset (we consider one stock $d = 1$) within a finite time T, and faces execution costs and market price impact. In contrast with frictionless Merton problem, we do not consider mark-to-market value of the portfolio and instead consider separately the amount on the cash account and the inventory Y, i.e., the position or number of shares held at any time. The strategy of the agent is then described by a real-valued $\mathbb{F}^S$-adapted process α, representing the velocity at which she buys ($\alpha_t > 0$) or sells ($\alpha_t < 0$) the asset, and the inventory is thus given by

$$Y_t \quad = \quad y_0 + \int_0^t \alpha_u du, \quad 0 \leqslant t \leqslant T.$$

The objective of the trader is to minimize over α the total liquidation cost

$$J_2(\alpha) \quad = \quad \mathbb{E}^0\Big[\int_0^T \alpha_t \big(S_t + f(\alpha_t)\big) dt + \ell(Y_T)\Big]$$

where $f(.)$ is an increasing function, $f(0) = 0$, representing a temporary price impact, and $\ell(.)$ is a loss function, i.e. a convex function with $\ell(0) = 0$, penalising the trader when she does not succeed to liquidate all her shares.

(2) *Portfolio selection.* The set $\mathcal{A}$ of portfolio strategies, representing the amount invested in the assets, consists in all $\mathbb{F}^S$-adapted processes α valued in some set A of $\mathbb{R}^d$, and satisfying $\int_0^T |\alpha_t|^2 dt < \infty$. The dynamics of wealth process $X = X^\alpha$ associated to a portfolio strategy α is then governed by

$$dX_t \quad = \quad \alpha_t.\beta_t dt + \alpha_t^\intercal \sigma dB_t^0, \quad X_0 \; = \; x_0 \in \mathbb{R},$$

and as in Merton portfolio selection problem, the objective of the agent is to maximise over portfolio strategies the utility of terminal wealth

$$J_1(\alpha) \quad = \quad \mathbb{E}^0[U(X_T)],$$

where U is an utility function on $\mathbb{R}$, e.g. CARA function $U(x) = -\exp(-px)$, $p > 0$.

Let us show how one can reformulate the above problems into a McKean-Vlasov type problem under partial observation and common noise as described in Section 1. We first introduce the so-called probability reference $\mathbb{P}$, which makes the observation price process a martingale. Let us then define the process

$$Z_t \quad = \quad \exp\big(- \int_0^t \sigma^{-1}\beta_u dB_u^0 - \frac{1}{2}\int_0^t |\sigma^{-1}\beta_u|^2 du\big), \quad 0 \leqslant t \leqslant T,$$

which is a $(\mathbb{P}^0, \mathbb{F})$-martingale (under suitable integrability conditions on β), and defines a probability measure $\mathbb{P} \sim \mathbb{P}^0$ through its density: $\frac{d\mathbb{P}}{d\mathbb{P}^0}\Big|_{\mathcal{F}_t} = Z_t$, and under which the process

$$W_t^0 \quad := \quad B_t^0 + \int_0^t \sigma^{-1}\beta_u du, \quad 0 \leqslant t \leqslant T,$$

is a $(\mathbb{P}, \mathbb{F})$-Brownian motion by Girsanov's theorem, and the dynamics of S is

$$dS_t \quad = \quad \mathrm{diag}(S_t)\sigma dW_t^0.$$

Notice that $\mathbb{F}^S = \mathbb{F}^0$ the filtration generated by W^0. We also denote by $L_t = 1/Z_t$, which is $(\mathbb{P}, \mathbb{F})$-martingale, governed by

$$dL_t \quad = \quad L_t\sigma^{-1}\beta_t.dW_t^0.$$

Next, we use Bayes formula and rewrite the gain (resp. cost) functionals of our two portfolio optimization problems as

$$\begin{aligned}
J_1(\alpha) &= \mathbb{E}\big[L_T U(X_T)\big] \;=\; \mathbb{E}\big[\bar{L}_T^0 U(X_T)\big] \;=\; \mathbb{E}\big[\bar{L}_T^0 U(\bar{X}_T^0)\big] \\
J_2(\alpha) &= \mathbb{E}\Big[\int_0^T L_t\alpha_t(S_t + \gamma\alpha_t)dt + \eta L_T Y_T^2\Big] \\
&= \mathbb{E}\Big[\int_0^T \bar{L}_t^0\alpha_t(S_t + f(\alpha_t))dt + \bar{L}_T^0\ell(Y_T)\Big] \;=\; \mathbb{E}\Big[\int_0^T \bar{L}_t^0\alpha_t(\bar{S}_t^0 + f(\alpha_t))dt + \bar{L}_T^0\ell(\bar{Y}_T^0)\Big]
\end{aligned}$$

where $\bar{L}_t^0 := \mathbb{E}[L_t|W^0] = \int \ell \mathbb{P}_{L_t}^{W^0}(d\ell)$, $\bar{X}_t^0 := \mathbb{E}[X_t|W^0] = \int x\mathbb{P}_{X_t}^{W^0}(dx) = X_t$, $\bar{Y}_t^0 := \mathbb{E}[Y_t|W^0] = \int y\mathbb{P}_{Y_t}^{W^0}(dy)$ $= Y_t$ $\bar{S}_t^0 := \mathbb{E}[S_t|W^0] = \int s\mathbb{P}_{S_t}^{W^0}(ds) = S_t$, and we used the law of conditional expectations and the fact that S, X and Y are $\mathbb{F}^0$-adapted. This formulation of the functional J_1 (resp. J_2) fits into the MKV framework of Section 1 with state variables (X, L, β) (resp. (Y, S, L, β))

We now consider the special case when β is an $\mathcal{F}_0$-measurable random variable distributed according to some probability distribution $\nu(db)$: this corresponds to a Bayesian point of view when the agent's belief about the drift is modeled by a prior distribution. In this case, let us show how our partial observation problem can be embedded into a finite-dimensional full observation Markov control problem. Indeed, by noting that β is independent of the Brownian motion W^0 under $\mathbb{P}$, we have

$$\bar{L}_t^0 \quad = \quad \mathbb{E}\big[\exp\big(\sigma^{-1}\beta.W_t^0 - \frac{1}{2}|\sigma^{-1}\beta|^2 t\big)\big|W^0\big] \;=\; F(t, W_t^0),$$

where

$$F(t,w) \quad := \quad \int \exp\big(\sigma^{-1}b.w - \frac{1}{2}|\sigma^{-1}b|^2 t\big)\nu(db).$$

Hence, the functionals J_1 and J_2 can be written as

$$J_1(\alpha) \quad = \quad \mathbb{E}\big[F(T, W_T^0)U(X_T)\big] \tag{1}$$

$$J_2(\alpha) \quad = \quad \mathbb{E}\Big[\int_0^T F(t, W_t^0)\alpha_t(S_t + f(\alpha_t))dt + F(T, W_T^0)\ell(Y_T)\Big]. \tag{2}$$

We are then reduced to a $(\mathbb{P}, \mathbb{F}^0)$-control problem with state variables (W^0, X) for problem (1) and (W^0, S, Y) for problem (2) with dynamics under $\mathbb{P}$:

$$\begin{aligned}
dS_t &= \mathrm{diag}(S_t)\sigma dW_t^0 \\
dX_t &= \alpha_t^\intercal \sigma dW_t^0 \\
dY_t &= \alpha_t dt.
\end{aligned}$$

4.1.2. *Numerical results*

Let us now illustrate numerically the impact of uncertain Bayesian drift on the portfolio liquidation problem and the portfolio selection problem, by considering a Gaussian prior distribution $\beta \rightsquigarrow \nu = \mathcal{N}(b_0, \gamma_0^2)$. In this case, F is explicitly given by:

$$F(t,w) = \frac{\sigma\gamma_0}{\sqrt{\sigma^2+\gamma_0^2 t}} \exp\Big(\frac{1}{2(\sigma^2+\gamma_0^2 t)}(-b_0^2 t + 2b_0\sigma w + \gamma_0^2 w^2)\Big).$$

1. Portfolio liquidation. Let us first consider the portfolio liquidation problem (2) with a linear price impact function $f(a) = \gamma a$, $\gamma > 0$, and a quadratic loss function $\ell(y) = \eta y^2$, $\eta > 0$. The optimal trading rate is given by (see [18])

$$\alpha_t^* \quad = \quad -\frac{Y_t^*}{T-t+\gamma/\eta} + \frac{1}{2\gamma}\Big(\frac{1}{T-t+\gamma/\eta}\int_t^T \mathbb{E}^0[S_u|\mathcal{F}_t^S]du - S_t\Big)$$

where Y^* is the associated inventory with feedback control α^*: $dY_t^* = \alpha_t^* dt$, $Y_0^* = y_0$. Since we consider a Gaussian prior $\mathcal{N}(b_0, \gamma_0^2)$ for β, the optimal trading rate is explicitly given by

$$\alpha_t^* \quad = \quad -\frac{1}{T-t+\gamma/\eta}\left\{Y_t^* + \frac{1}{2\gamma}\left[-\frac{1}{\gamma_0}\sqrt{\frac{\pi}{2}}e^{-\frac{b_0^2}{2\gamma_0^2}}\left(\mathrm{erfi}\left(\frac{b_0+\gamma_0^2(T-t)}{\sqrt{2}\gamma_0}\right) - \mathrm{erfi}\left(\frac{b_0}{\sqrt{2}\gamma_0}\right)\right) + (T-t+\frac{\gamma}{\eta})\right]S_t\right\},$$

where `erfi` is the imaginary error function, defined as:

$$\mathrm{erfi}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{t^2}dt.$$

Remark. In particular, when the price process is a martingale, i.e. $b_0 = 0$, and in the limiting case when the penalty parameter η goes to infinity, corresponding to the final constraint $Y_T = 0$, we see that α_t^* converges to $-Y_t^*/(T-t)$, hence independent of the price process, and leading to an explicit optimal inventory: $Y_t^* = y_0\frac{T-t}{T}$ with constant trading rate $\alpha_t^* = -y_0/T$. We retrieve the well-known VWAP strategy obtained in Almgren-Criss.

We solve the problem numerically, taking $N = 100$ for the time discretization, and fixing the other parameters as follows: γ=5, S_0=6, Y_0=1, η=100 and σ=0.4. We run two sets of forward Monte Carlo simulations for $\beta_0 = 0.1$, $T = 1$ and $\beta_0 = -0.1$, $T = 0.5$ changing the value of γ_0. We tested the Regress Later Monte Carlo algorithm (RLMC), the Control Randomization algorithm (CR) and the quantization algorithm (Q). In particular we wanted to compare the performance of these algorithms with the optimal strategy (Opt). We also tested a benchmark strategy (Bench) which consists in liquidating the inventory at a constant rate $-y_0/T$. The test consisted in computing a forward Monte Carlo with 500000 samples, following these different strategies, to estimate the value functions at time 0.
We display the results obtained by the different algorithms in table 1. Plots of the tables are available in figure 1. One can observe from figure 1 that the algorithms perform better than the optimal strategy (Opt) in some cases. This is due to the fact that discretizing in time the initial problem, and Opt is not optimal for the time-discretized portfolio liquidation problem. Also, it can be observed from the same table that the quality of the policy estimated by Regress Later decreases with increasing uncertainty on the drift γ_0 at a faster pace than quantization, showing that the latter is better suited for highly unstable situations. The policy estimated by Control Randomisation deteriorates likewise when γ_0 is increased.
Figure 2 shows a sample of the inventory $(Y_t)_{t\in[0,T]}$ when the agent follows the optimal strategy and the quantization algorithm. On can see that the strategies differ a little bit when the drift is high. Also, one can notice that given the penalization parameters that we took, it is optimal to short some stocks at terminal time, when the drift is high. Finally, notice that the concaveness of the curves comes from the fact that the running cost does not penalize the inventory. In the latter case, we expect the curves of the inventory w.r.t time to be convex.
Remarks on the RL and CR algorithms The implementation of Regression Monte Carlo algorithms has required intense tuning and the use of the performance iteration technique introduced in Section 3.1

TABLE 1. Portfolio Liquidation results. Value functions at time 0 when the agent follows the different strategies.

	$\beta_0 = 0.1$, $T = 1$					$\beta_0 = -0.1$, $T = 1/2$				
γ_0	Opt	RLMC	CR	Q	Bench	Opt	RLMC	CR	Q	Bench
0.1	-1.347	-1.356	-1.278	-1.368	-1.318	3.689	3.687	3.995	3.686	4.144
0.2	-1.385	-1.390	-1.283	-1.401	-1.348	3.682	3.682	3.847	3.679	4.138
0.3	-1.445	-1.446	-1.314	-1.460	-1.402	3.670	3.674	4.034	3.667	4.126
0.4	-1.523	-1.524	-1.323	-1.556	-1.485	3.655	3.674	4.128	3.650	4.108
0.5	-1.642	-1.637	-1.348	-1.673	-1.585	3.636	3.664	4.243	3.630	4.088
0.6	-1.783	-1.777	-1.425	-1.826	-1.711	3.611	3.640	4.386	3.607	4.064
0.7	-1.973	-1.927	-1.513	-2.018	-1.870	3.581	3.613	4.783	3.572	4.029
0.8	-2.213	-2.003	-1.637	-2.243	-2.057	3.545	3.575	5.142	3.537	3.992
0.9	-2.526	-2.457	-1.819	-2.516	-2.288	3.5	3.530	5.345	3.498	3.952
1	-2.918	-2.801	-1.806	-2.829	-2.56	3.453	3.513	6.765	3.452	3.903

in order to obtain satisfactory results. Paramount is, in addition, the distribution chosen for the training points in Regress Later and for the initial control in Control Randomisation. The problem of finding the best set of data to provide to the backward procedure is similar in the two Regression Monte Carlo algorithms however little study is available in literature; for more details on this problem in the Regress Later setting see [13] and [1]. Finally note that we observed very high volatility in the quality of the policy estimated by control randomisation, for this reason we estimated the policy 50 times, and report in table 1 the results provided by the best performing one; increasing the number of training points further affects the variability only marginally.

Remarks on the Q algorithm The quantization algorithm had to be modified a little bit to perform well in these simulations. We decided to smooth the previous approximations of the conditional expectations with respect to the control. The previous approximation was as follows:

$$\mathbb{E}^a_{n,w,y}\big[\widehat{V}\big(n+1,\widehat{W}_{n+1},\widehat{Y}_{n+1}\big)\big] \approx \sum_{e\in\Gamma^\varepsilon} \mathbb{P}(\widehat{\varepsilon}=e)\widehat{V}\Big(n+1,\pi^W_{n+1}\big(G(n,w,y,a,e)\big),\pi^Y_{n+1}\big(G(n,w,y,a,e)\big)\Big).$$

One can notice that we quantized each dimension of the process (W_t, Y_t) separately. This multidimensional quantization method as already been studied before (see e.g. [17]). Denote by $G_w(n,w,y,a,e)$ and $G_y(n,w,y,a,e)$ the noise component and the portfolio components of $G(n,w,y,a,e)$, ie: $G(n,w,y,a,e) = \big(G_w(n,w,y,a,e), G_y(n,w,y,a,e)\big)$. See (6) for the definition of G.

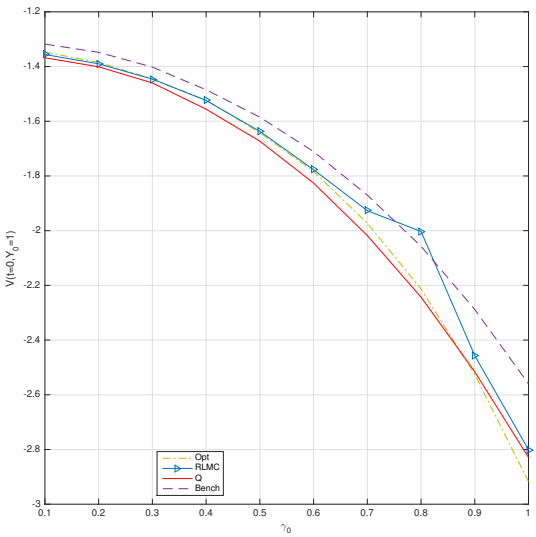
The improved approximation is as follows:

$$\mathbb{E}^a_{n,w,y}\big[\widehat{V}\big(n+1,\widehat{W}_{n+1},\widehat{Y}_{n+1}\big)\big] \approx \sum_{e\in\Gamma^\varepsilon} \mathbb{P}(\widehat{\varepsilon}=e)\Big[\lambda^{e,\hat{w},y}\widehat{V}\big(n+1,y_+\big) + (1-\lambda^{e,\hat{w},y})\widehat{V}\big(n+1,\hat{w},y_-\big)\Big],$$

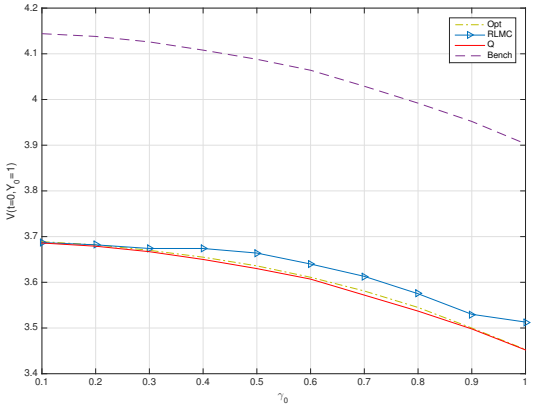
where y_- and y_+ are the two closest states in Γ^Y_{n+1} from $G_y(n,w,y,a,e)$, such that $y_- < G_y(n,w,y,a,e) < y_+$; and $\lambda^{e,\hat{w},y} := \frac{G_y(n,w,y,a,e)-y_-}{y_+-y_-}$. This approximation is continuous with respect to the control a. In particular, this feature will be useful when using usual algorithms to solve the minimization problems that come up from the Bellman equation.

2. Portfolio selection. Consider the portfolio selection problem (1) with one risky asset. We choose a CARA utility function $U(x) = -\exp(-px)$, with $p > 0$. It has been shown in [9] (see their Corollary 1) that the optimal portfolio strategy is explicitly given by

$$\alpha^*_t \quad = \quad \frac{\sigma^2+\gamma_0^2 t}{\sigma^2+\gamma_0^2 T}\frac{\hat{\beta}_t}{p\sigma^2}$$



$\beta_0 = 0.1$ and $T = 1$.



$\beta_0 = -0.1$ and $T = 0.5$.

FIGURE 1. Results for the portfolio liquidation problem. Value functions at time 0 when the agent follows different strategies w.r.t γ_0. We took γ=5, S_0=6, Y_0=1, η=100 and σ=0,4.

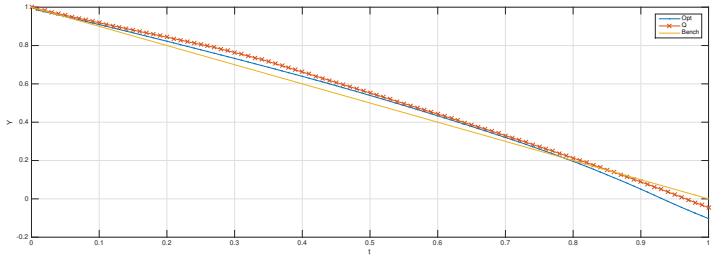


FIGURE 2. Simulation of $(Y_t)_{t\in[0,T]}$ using the optimal strategy (Opt), the quantization algorithm (Q), and the Benchmark strategy (Bench) to solve the portfolio liquidation problem. We took $T = 1$, $\sigma = 0.4$, $\gamma_0 = 1$, $b_0 = 0.1$, $S_0 = 6$, $Y_0 = 1$, $N = 100$, $\gamma = 5$, $\eta = 100$. We notice that when the drift is high, the inventory at terminal time Y_T is negative if the agent follows both the optimal strategy and the quantization strategy.

where

$$\hat{\beta}_t \quad = \quad \mathbb{E}^0[\beta|\mathcal{F}_t^S] \quad = \quad \frac{\sigma^2}{\sigma^2+\gamma_0^2 t}b_0 + \frac{\gamma_0^2}{\sigma^2+\gamma_0^2 t}\Big(\ln\frac{S_t}{S_0} + \frac{1}{2}\sigma^2 t\Big),$$

is the posterior mean of the drift (Bayesian learning on the drift), and the optimal performance by

$$J_1(\alpha^*) \quad = \quad -\exp\Big[-p\Big(x_0 + \frac{1}{2p}\big(\ln\big(\frac{\sigma^2+\gamma_0^2 T}{\sigma^2}\big) - \frac{\gamma_0^2 T}{\sigma^2+\gamma_0^2 T}\big) + \frac{b_0^2}{2p\sigma^2}\frac{\sigma^2 T}{\sigma^2+\gamma_0^2 T}\Big)\Big].$$

The Portfolio Selection problem, even though in many aspects similar to the Portfolio Liquidation problem, it is interesting in his own merit because the control acts only on the variance of the controlled wealth process. We tested the Regress Later Monte Carlo algorithm (RLMC), the Control Randomization algorithm (CR), the quantization algorithm (Q) on the portfolio selection problem. Similarly to what has been done for Portfolio Liquidation problem, we discretised time choosing $N = 100$ and solved the discrete time problem

TABLE 2. Portfolio Selection results. Value functions at time 0 when the agent follows Opt and Q strategies.

	$\beta_0 = 0.1$, $T = 1$				$\beta_0 = -0.1$, $T = 1/2$			
γ_0	Opt	RLMC	CR	Q	Opt	RLMC	CR	Q
0.1	-0.98522	-	-	-0.98522	-0.99239	-	-	-0.99239
0.2	-0.98272	-	-	-0.98272	-0.99138	-	-	-0.99138
0.3	-0.97302	-	-	-0.97301	-0.98847	-	-	-0.98847
0.4	-0.95401	-	-	-0.95398	-0.98115	-	-	-0.98115
0.5	-0.92764	-	-	-0.92757	-0.96911	-	-	-0.96911
0.6	-0.89678	-	-	-0.89666	-0.95219	-	-	-0.95218
0.7	-0.86370	-	-	-0.86350	-0.93219	-	-	-0.93217
0.8	-0.83035	-	-	-0.83007	-0.91080	-	-	-0.91076
0.9	-0.79795	-	-	-0.79758	-0.88655	-	-	-0.88650
1	-0.76703	-	-	-0.76658	-0.86311	-	-	-0.86304

associated. We considered two set of experiments, $\beta_0 = 0.1$, $T = 1$ and $\beta_0 = -0.1$, $T = 1/2$, for values of $\gamma_0 \in [0, \sigma]$, $P = 1$, $\sigma = 0.4$. Given all these different parameters, we compared the performance of these algorithms with the one of the optimal strategy (Opt). The general test consists in computing a forward Monte Carlo with 500000 samples, following these different strategies, to estimate the value function at time 0. We present the results of our numerical experiments in table 2. One can see that the quantization algorithm is doing slightly better than the theoretical optimal strategy (Opt) for the continuous time problem. This is due to the fact that Opt is not optimal for the time-discretized portfolio selection problem. We also present figure 3 which shows a sample of the wealth of the agent following the optimal strategy and the quantization algorithm. One can see that the strategies slightly differ when the drift is high, and remain the same when the drift is low.

Remarks on the RL and CR algorithms When implementing Regression Monte Carlo algorithms, and choosing basis functions, the control on variance implies that low order polynomial can not be used alone, as they can easily cause the control to be bang bang between the boundaries of its domain. Similarly piecewise approximations are not very effective, as the dependence on the control is very weak requiring an high number of local supports, making the computational complexity overwhelming. We tested both value and performance iteration and tried to employ different kinds of basis functions and training points, unfortunately both Regress Later and Control Randomisation do not cope well with controlling the dynamics of a process through the variance only. A tailor made implementation of Regression Monte Carlo to deal with these sort of problems is outside the scope of this paper and further investigation will follow in future work.

Remarks on the Q algorithm We designed the same quantization algorithm as the one built to solve the portfolio liquidation problem. We nevertheless had to take more points in the grids to avoid problems in the borders of the grids.

4.2. **A model of interbank systemic risk with partial observation**

4.2.1. *The model*

We consider a model of systemic risk inspired by the model in [6]. The monetary reserves of N banks lending to and borrowing from each other are governed by the system

$$dX_t^i = \frac{\kappa}{N}\sum_{j=1}^{N}(X_t^j - X_t^i)dt + \sigma X_t^i(\sqrt{1-\rho^2}dW_t^i + \rho dW_t^0), \quad i = 1, \dots, N,$$

where W^i, $i = 1, \dots, N$, are independent Brownian motions, representing the idiosyncratic risk of each bank, W^0 is a common noise (systematic risk) independent of W^i, and $\rho \in [-1, 1]$. The mean-reversion coefficient $\kappa > 0$ models the strength of interaction between the banks where bank i can lend to and borrow from banks j with an amount proportional to the difference of their reserves. In the asymptotic regime when $N \to \infty$,

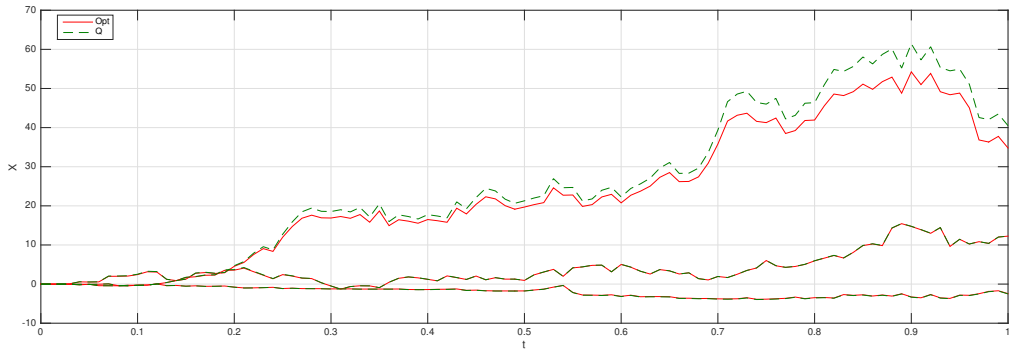


FIGURE 3. 3 simulations of the agent's wealth $(X_t)_{t\in[0,T]}$ when the latter follows the optimal strategy (Opt) and the quantized strategy (Q) to solve the portfolio selection problem. We took σ=0.4, T=1, P=0.1, γ_0=5, b_0=0.1. One can see that the two strategies are the same when the drift is low, but Q performs slightly better than Opt when the drift is high.

the theory of propagation of chaos implies that the reserve state X^i of individual banks become independent and identically distributed conditionally on the common noise W^0, with a state governed by

$$dX_t \quad = \quad \kappa(\mathbb{E}[X_t|W^0] - X_t)dt + \sigma X_t(\sqrt{1-\rho^2}dB_t + \rho dW_t^0),$$

for some Brownian motion B independent of W^0.

Let us now consider a central bank, viewed as a social planner, who only observes the common noise and not the reserves of each bank, and can influence the strength of the interaction between the individual banks, through an $\mathbb{F}^0$-adapted control process α_t. The reserve of the representative bank in the asymptotic regime is then driven by

$$dX_t \quad = \quad (\kappa+\alpha_t)(\mathbb{E}[X_t|W^0] - X_t)dt + \sigma X_t(\sqrt{1-\rho^2}dB_t + \rho dW_t^0),$$

and the objective of the central bank is to minimize

$$J(\alpha) \quad = \quad \mathbb{E}\Big[\int_0^T \frac{1}{2}\alpha_t^2 + \frac{\eta}{2}(X_t - \mathbb{E}[X_t|W^0])^2 dt + \frac{c}{2}(X_T - \mathbb{E}[X_T|W^0])^2\Big],$$

where $\eta > 0$ and $c > 0$ penalize the departure of the reserve from the average. This is a McKean-Vlasov control problem under partial observation, but notice that it does not belong to the class of LQ MKV problems due to the control α which appears in a multiplicative form with the state. However, it fits into our class of polynomial MKV problem, and can be embedded into standard control problem as follows: We set $\bar{X}_t = \mathbb{E}[X_t|W^0]$ and $Y_t = \mathbb{E}[(X_t - \bar{X}_t)^2|W^0]$. The cost functional is then written as

$$J(\alpha) \quad = \quad \mathbb{E}\Big[\int_0^T \frac{1}{2}\alpha_t^2 + \frac{\eta}{2}Y_t dt + \frac{c}{2}Y_T\Big]$$

where (after some straightforward calculations) the dynamics of $\bar{X}$ and Y are governed by

$$\begin{aligned} d\bar{X}_t &= \sigma\rho\bar{X}_t dW_t^0 \\ dY_t &= \big[(\sigma^2 - 2(\kappa+\alpha_t))Y_t + \sigma^2(1-\rho^2)\bar{X}_t^2\big]dt + 2\rho\sigma Y_t dW_t^0. \end{aligned}$$

We have then reduced our problem to a $(\mathbb{P},\mathbb{F}^0)$-control problem in dimension two with state variables $(\bar{X},Y)$, which is neither LQ, but can be solved numerically.

4.2.2. *Numerical results*

For this problem, no analytical solution is available, so we decided to compare the policies estimated by our algorithms with a benchmark (Bench) obtained by solving the corresponding 2-dimensional Hamilton-Jacobi-Bellman equation by deterministic methods. To compare the different algorithms, we computed the value functions at time 0 following each strategy. We run a forward Monte Carlo with 500 000 samples, using the following parameters $T = 1$, $\sigma = 0.1$, $\kappa = 0.5$ and $X_0 = 10$ to estimate the different value functions. In table 3 we display the results of our numerical experiments for two situations $\eta = 10$, $c = 100$ and $\eta = 100$, $\rho = 0.5$ varying the value of ρ in the first case, and the value of c in the second. Plots of the two tables are available in figure 4. One can see that Q performs often better than Bench. This is due to the fact that Bench is not optimal for the time-discretized systemic risk controlled problem. Also, notice that Regression Monte Carlo algorithms perform well and in particular RLMC obtains very fast and stable estimations of the optimal strategy.
Figure 5 shows two examples of paths $(X_t)_{t\in[0,T]}$ controlled by RLMC (RLMC), $(X_t)_{t\in[0,T]}$ naively controlled by $\alpha = 0$ (uncontrolled), and the conditional expectation of X $(\bar{X}_t)_{t\in[0,T]}$ $(E(X|W))$. One can see in these two examples that the optimal control given by RLMC is to:

- do nothing when the terminal time is far, i.e. take $\alpha = 0$, to not to pay any running cost.
- catch $\bar{X}$ when the terminal time is getting close, to minimize the terminal cost.

Finally we present a sample of paths $(Y_t)_{t\in[0,T]}$ controlled by the decisions given by Q in figure 6. One can see that the optimal Q-strategy minimizes the running cost first by letting Y grow; and deals with the terminal cost later by making Y small when the terminal time is approaching.

Remarks on the RL and CR algorithms

- RLMC has been designed using only linear and quadratic functions as basis functions. That is why the latter is able to provide fast results.
- A requirement for the convergence of the scheme is a careful generation of the training points for the process Y which should be relatively concentrated around zero.

Remarks on the Q algorithm Given the dynamic of $(Y_t)_{t\in[0,T]}$, it is straightforward that $Y > 0$ on $(0, T]$. However, the Euler scheme used to approximate the dynamic of Y does not prevent the associated process $(Y_{t_i})_{0<i\leqslant N}$ to be non-positive. When implementing the Q algorithm for the systemic risk problem, we forced $(\hat{Y}_{t_i})_{0<i\leqslant N}$ to remain positive by choosing positive points for the grids Γ_i^Y that quantize the states of Y_{t_i}, $\quad 0 < i \leqslant N$.
Also, given the expression of the instantaneous and terminal reward, one can expect Y to stay close to 0, but we do not have any idea of how small Y should stay for the strategy to be optimal (see figure 6 to see a posteriori where Y lies). To deal with this situation, we decided to use a method of boostrapping: first, we chose some random grids with lot of points near 0, and computed the optimal strategy on these grids. Then, we run forward Monte Carlo simulations and generated an empirical distribution of the quantized Y. Secondly, we build new grids of quantization for Y by generating new points according to the empirical distribution that we got from in the previous step. Finally, we computed the optimal strategy on the new grids and compute the corresponding optimal strategy. The Q strategy performed better after Bootstrapping, but not significantly since our first naive guess for the grids (i.e. before bootstrapping) was already good enough.

5. Conclusion

In this work we have investigated how to use probabilistic numerical methods for some classes of mean field control problem via Markovian embedding. We focused on two types of Regression Monte Carlo methods (namely, Regress Later and Control Randomization) and Quantization. We have then presented three different examples of applications.

We found that the Regression Monte Carlo algorithms perform well in problems of control of the drift. In such problems they are much faster than Quantization for similar precision. In particular we noticed that Regress Later is usually more reliable than Control Randomisation; often the choice of an uniform distribution of the training points on an appropriate interval is sufficient to obtain high quality estimations. On the other hand Control Randomisation is very sensitive to the choice of the distribution of the randomised

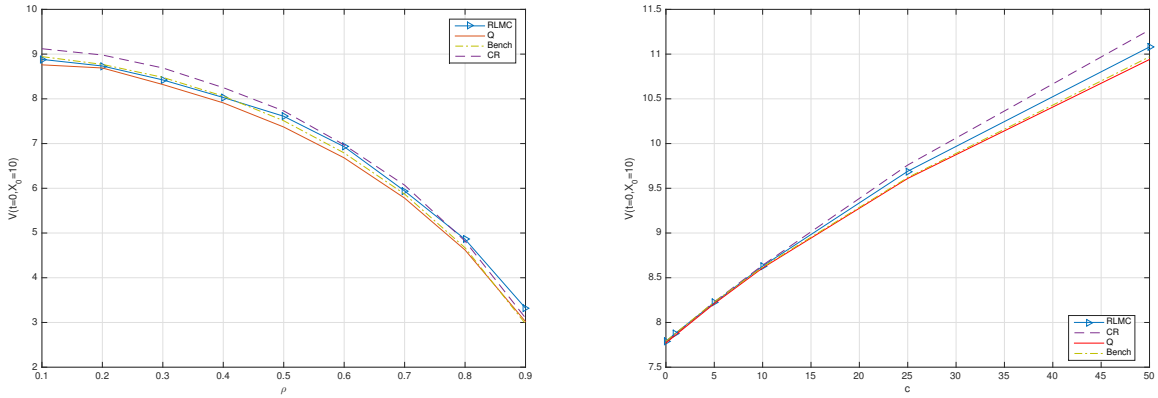
ρ	RLMC	CR	Q	Bench
0.1	8.88	9.12	8.76	8.94
0.2	8.73	8.98	8.69	8.77
0.3	8.42	8.69	8.32	8.48
0.4	8.02	8.25	7.91	8.06
0.5	7.61	7.73	7.37	7.51
0.6	6.93	6.97	6.68	6.79
0.7	5.94	6.07	5.78	5.87
0.8	4.86	4.82	4.62	4.67
0.9	3.32	3.10	3.02	2.97

$c = 100$ and $\eta = 10$.

c	RLMC	CR	Q	Bench
0	7.79	7.78	7.77	7.79
1	7.88	7.87	7.86	7.88
5	8.22	8.23	8.21	8.23
10	8.63	8.64	8.61	8.62
25	9.69	9.76	9.61	9.62
50	11.08	11.27	10.94	10.97

$\rho = 0.5$ and $\eta = 100$.

TABLE 3. Results for the systemic risk problem. Value functions at time 0 when the agent follows different strategies. We took $T = 1$, $N = 100$, $\sigma = 0.1$, $\kappa = 0.5$, $X_0 = 10$.



$c = 100$ and $\eta = 10$. $\rho = 0.5$ and $\eta = 100$.

FIGURE 4. Results for the systemic risk problem. Value function at time 0 when the agent follows different strategies with respect to ρ and c. We took T=1, N=100, σ=0.1, κ=0.5, X_0=10.

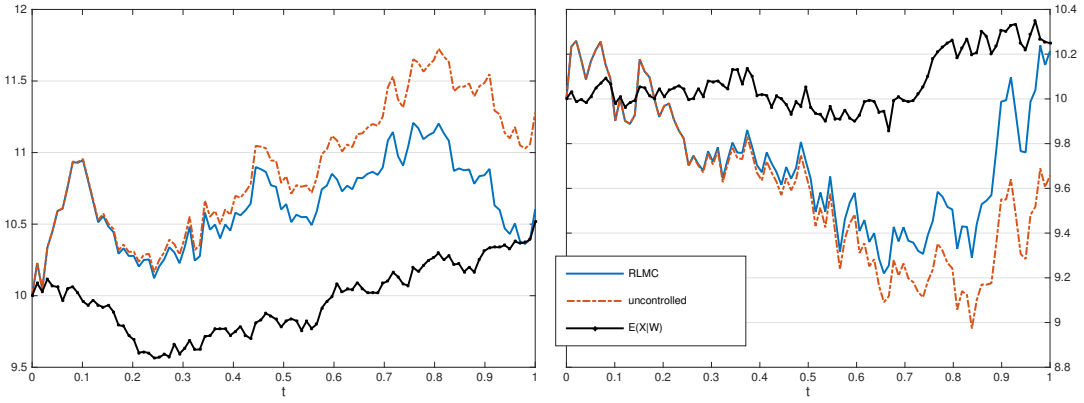


FIGURE 5. Sample of $(X_t)_{t\in[0,T]}$ controlled by RLMC (RLMC), $(X_t)_{t\in[0,T]}$ naively controlled taken $\alpha = 0$ (uncontrolled), and $\bar{X}$ $(E(X|W))$. The optimal control for the systemic risk problem (computed by RLMC) is to do nothing at first, and catch $\bar{X}$ when the terminal time is getting close.

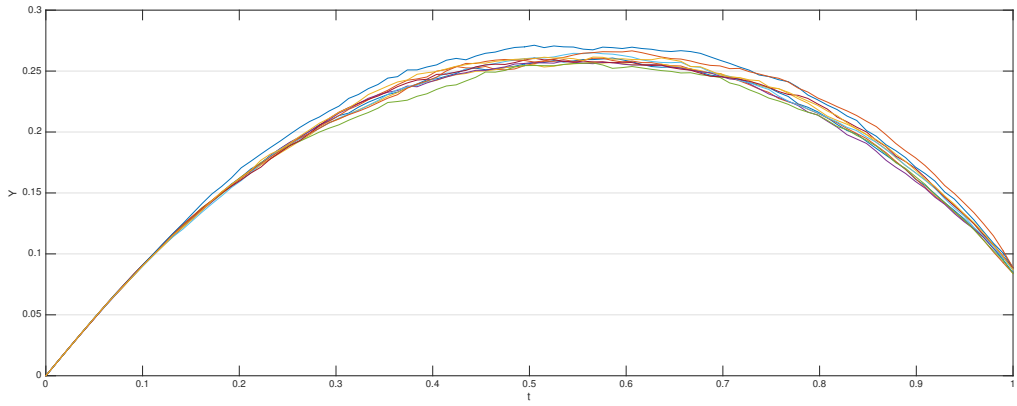


FIGURE 6. Sample of $(Y_t)_{t\in[0,T]}$ controlled by Q. The optimal control for the systemic risk problem is to let Y get large at first, and make it small when terminal time is approaching.

control, and often few repetition are necessary before finding a good control distribution. We have also tried to use the performance iteration, or path recomputation method, but on the examples we considered it was very time consuming and did not help much in terms of accuracy. Despite the success of Regression Monte Carlo methods in problems with control on the drift, the example of Portfolio Selection highlighted a possible weakness of these algorithms. When the control acts on the variance only we found difficult to make the numerical scheme converge to sensible results within the computational resources available. We realised that the study of these problems and the solution via Regression Monte Carlo methods is outside the scope of this paper. This is probably related to another limitation of this family of methods: the choice of the basis functions for the regression. Indeed, for some problems the basis might be very large or might require several steps of trials and errors.

Quantization, on the other hand, provided the most stable and accurate results for the three different kinds of control problems that has been considered. An interesting feature of the quantization methods is that one has to choose the grids to quantize the controlled process. It is possible to exploit this feature in the cases where one has, a priori, a rough idea of where the controlled process should be driven by the optimal strategy (see e.g. the liquidation problem). In this case, one should build grids with many points located where the process is supposed to go. In the case where one has no guess of where the optimal process goes, it is always possible to use bootstrapping methods to build better grids iteratively, starting from a random guess for the grid (see e.g. systemic risk). However, note that this second alternative require more computation-time. In both cases, one has to be particularly attentive to the borders of the grids that have been built. Indeed, the decisions computed by Q at the borders might easily be wrong if the grids do not have a good shape at the borders. Except in very special cases, it seems not possible to avoid the use of deterministic algorithms (such as gradient descent methods or extensive search) to find the optimal action at each point of the grid. A smooth expression of the conditional expectations of the quantized processes is necessary for the deterministic algorithms to converge properly. Use of parallel computing can alleviate the time consuming task of searching for the optimal control.

REFERENCES

[1] Balata A. and Palczewski J. (2018): Regress-Later Monte Carlo for optimal control of Markov processes, arXiv:1712.09705

[2] Broadie M. and Glasserman P. (2004): A stochastic mesh method for pricing high-dimensional American options, *Journal of Computational Finance*, **7** (35), 35-72.

[3] Buckdahn R., Li J., Peng S. and Rainer C. (2015): Mean-field stochastic differential equations and associated PDEs, arXiv: 1407.1215, to appear in the *Annals of Probability*.

[4] Cardaliaguet P. Notes on mean field games, Notes from P.L. Lions lectures at Collège de France (2013).

[5] Carmona R. and Delarue F. (2014): The Master equation for large population equilibriums, D. Crisan et al. (eds.), Stochastic Analysis and Applications 2014, Springer Proceedings in Mathematics & Statistics 100.

[6] Carmona R., Fouque J.P. and Sun L. (2015): Mean Field Games and Systemic Risk, *Communications in Mathematical Sciences*,**13**(4), 911-933.
[7] Chassagneux J.F., Crisan D. and Delarue F. (2015): A probabilistic approach to classical solutions of the master equation for large population equilibria, arXiv: 1411.3009.
[8] Glasserman P. and Yu B. (2002): Simulation for American Options: Regression Now or Regression Later?,*Monte Carlo and Quasi-Monte Carlo Methods*, chapter , pages 213–226. Springer Berlin Heidelberg.
[9] Guéant 0. and Pu J. (2016): Portfolio choice, portfolio liquidation, and portfolio transition under drift uncertainty, arXiv:1611.07843
[10] Kharroubi I., Langrené N. and Pham H. (2014) A numerical algorithm for fully nonlinear HJB equations: an approach by control randomization. *Monte Carlo Methods and Applications*, **20** (2), 145–165.
[11] Lions P.L. *Cours au Collège de France: Théorie des jeux à champ moyens*, audio conference 2006-2012.
[12] Longstaff F. and Schwartz E. (2001): Valuing American Options by Simulation: A Simple Least-Squares Approach, *The Review of Financial Studies*, **14**(1), 113-147.
[13] Nadarajah S., Margot F. and Secomandi N. (2017). Comparison of least squares Monte Carlo methods with applications to energy real options, *European Journal of Operational Research*, **256**, 196?204.
[14] Pagès G. (1997): A space vector quantization method for numerical integration *Journal of Applied and Computational Mathematics, 89, 1-38*
[15] Pagès, G., Pham, H. and Printems, J. (2004) An Optimal Markovian Quantization Algorithm for Multidimensional Stochastic Control Problems. *Handbook on Numerical Methods in Finance*, chapter 7, pages 253–298. Birkhauser.
[16] Pagès, G. and Printems, J. (2003): Optimal quadratic quantization for numerics: the Gaussian case *Monte Carlo Methods & Applications Journal, 9(2), 135-166*
[17] Pagès, G. and Sagna, A. (2015): Markovian and product quantization of an Rd-valued Euler scheme of a diffusion process with applications to finance *pré-pub LPMA 1670*
[18] Pham H. (2016): Linear quadratic optimal control of conditional McKean-Vlasov equation with random coefficients and applications, *Probability, Uncertainty and Quantitative Risk*, 1-7.
[19] Pham H. and Wei X. (2017): Dynamic programming for optimal control of stochastic McKean-Vlasov dynamics, *SIAM Journal on Control and Optimization*, **55**(2),1069-1101.