

Algorithmic trading in a microstructural limit order book model

Frédéric Abergel, Côme Huré, Huyên Pham

▶ **To cite this version:**

Frédéric Abergel, Côme Huré, Huyên Pham. Algorithmic trading in a microstructural limit order book model. Quantitative Finance, 2020, 20 (8), pp.1263-1283. 10.1080/14697688.2020.1729396 . hal-01514987v3

HAL Id: hal-01514987
https://hal.science/hal-01514987v3

Submitted on 19 Feb 2020

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.

Algorithmic trading in a microstructural limit order book model

Prof. Frédéric Abergel† Côme Huré*‡ Prof. Huyên Pham‡‡

† MICS Laboratory - CentraleSupelec, frederic.abergel@ecp.fr
‡LPSM - Paris 7 Diderot University, hure@lpsm.paris
‡‡ LPSM - Paris 7 Diderot University and CREST - ENSAE, pham@lpsm.paris

We propose a microstructural modeling framework for studying optimal market-making policies in a FIFO (first in first out) limit order book (order book). In this context, the limit orders, market orders, and cancel orders arrivals in the order book are modeled as point processes with intensities that only depend on the state of the order book. These are high-dimensional models which are realistic from a micro-structure point of view and have been recently developed in the literature. In this context, we consider a market maker who stands ready to buy and sell stock on a regular and continuous basis at a publicly quoted price, and identifies the strategies that maximize their P&L penalized by their inventory. An extension of the methodology is proposed to solve market-making problems where the orders arrivals are modeled using Hawkes processes with exponential kernel.

We apply the theory of Markov Decision Processes and dynamic programming method to characterize analytically the solutions to our optimal market-making problem. The second part of the paper deals with the numerical aspect of the high-dimensional trading problem. We use a control randomization method combined with quantization method to compute the optimal strategies. Several computational tests are performed on simulated data to illustrate the efficiency of the computed optimal strategy. In particular, we simulated an order book with constant/ symmetric/ asymmetrical/ state dependent intensities, and compared the computed optimal strategy with naive strategies. Some codes are available on `https://github.com/comeh`.

Keywords: Limit order book, pure-jump controlled process, high-frequency trading, high-dimensional stochastic control, Markov Decision Process, quantization, local regression

1. Introduction

Most of the markets use a limit order book (order book) mechanism to facilitate trade. Any market participant can interact with the order book by sending either market orders or limit orders. In such type of markets, the market makers play a fundamental role by providing liquidity to other market participants, typically to impatient agents who are willing to cross the bid-ask spread. The profit made by a market-making strategy comes from the alternation of buy and sell orders.

From the mathematical modeling point of view, the market-making problem corresponds to the choice of an optimal strategy for the placement of orders in the order book. Such a strategy should maximize the expected utility function of the wealth of the market maker up to a penalization of their inventory. In the recent litterature, several works focused on the problem of market-making through stochastic control methods. The seminal paper by Avellaneda and Stoikov (2007) inspired

*corresponding author

by the work of Ho and Stoll (1979) proposes a framework for trading in an order driven market. They modeled a reference price for the stock as a Wiener process, and the arrival of a buy or sell liquidity-consuming order at a distance δ from the reference price is described by a point process with an intensity in an exponential form decreasing with δ. They characterized the optimal market-making strategies that maximize an exponential utility function of terminal wealth. Since this paper, other authors have worked on related market-making problems. Guéant *et al.* (2012) generalized the market-making problem of Avellaneda and Stoikov (2007) by dealing with the inventory risk. Cartea and Jaimungal (2013) also designed algorithms that manage inventory risk. Fodra and Pham (2015b) and Fodra and Pham (2015a) considered a model designed to be a good compromise between accuracy and tractability, where the stock price is driven by a Markov Renewal Process, and solved the market-making problem. Guilbaud and Pham (2013) also considered a model for the mid-price, modeled the spread as a discrete Markov chain that jumps according to a stochastic clock, and studied the performance of the market-making strategy both theoretically and numerically. Cartea and Jaimungal (2010) employed a hidden Markov model to examine the intra-day changes of dynamics of the order book. Very recently, Cartea *et al.* (2015) and Guéant (2016) published monographs in which they developped models for algorithmic trading in different contexts. El Aoud and Abergel (2015) extended the framework of Avellaneda and Stoikov to the options market-making. A common feature of all these works is that a model for the price or/and the spread is considered, and the order book is then built from these quantities. This approach leads to models that predict well the long-term behavior of the order book. The reason for this choice is that it is generally easier to solve the market-making problem when the controlled process is low-dimensional. Yet, some recent works have introduced accurate and sophisticated micro-structural order book models. These models reproduce accurately the short-term behavior of the market data. The focus is on conditional probabilities of events, given the state of the order book and the positions of the market maker. Abergel *et al.* (2016) proposed models of order book where the arrivals of orders in the order book are driven by Poisson processes or Hawkes processes. Cont *et al.* (2007) also modeled the orders arrivals with Poisson processes. Huang *et al.* (2015) proposed a queue-reactive model for the order book. In this model the arrivals of orders are driven by Cox point processes with intensities that only depend on the state of the order book (they are not time dependent). Other tractable dynamic models of order-driven market are available (see e.g. Cont *et al.* (2007), Rosu (2008), Cartea *et al.* (2014)).

In this paper we adopt the micro-structural model of order book in Abergel *et al.* (2016), and solve the associated trading problem. The problem is formulated in the general framework of Piecewise Deterministic Markov Decision Process (PDMDP), see Bäuerle and Rieder (2011). Given the model of order book, the PDMDP formulation is natural. Indeed, between two jumps, the order book remains constant, so one can see the modeled order book as a point process where the time becomes a component of the state space. As for the control, the market maker fixes their strategy as a deterministic function of the time right after each jump time. We prove that the value function of the market-making problem is equal to the value function of an associated non-finite horizon Markov decision process (MDP). This provides a characterization of the value function in terms of a fixed point dynamic programming equation. Jacquier and Liu (2018) recently followed a similar idea to solve an optimal liquidation problem, while Baradel *et al.* (2018) and Lehalle *et al.* (2018) also tackled this problem of reward functional maximization in a micro-structure model of order book framework.

The second part of the paper deals with the numerical simulation of the value functions. The computation is challenging because the micro-structural model used to model the order book leads to a high-dimensional pure jump controlled process, so evaluating the value function is computationally intensive. We rely on control randomization and Markovian quantization methods to compute the value functions. Markovian quantization has been proved to be very efficient for solving control problems associated with high-dimensional Markov processes. We first quantize the jump times and then quantize the state space of the order book. See Pagès *et al.* (2004) for a general description of

quantization applied to controlled processes. The projections are time-consuming in the algorithm, but Fast approximate nearest neighbors algorithms (see e.g. Muja and Lowe (2009)) can be implemented to alleviate the procedure. We borrow the values of intensities of the arrivals of orders for the order book simulations from Huang *et al.* (2015) in order to test our optimal trading strategies.

The paper is organized as follows. The model setup is introduced in Section 2: we present the micro-structural model for the order book, and show how the market maker interacts with the market. In Section 3, we prove the existence and provide a characterization of the value function and optimal trading strategies. In Section 4, we introduce a quantization-based algorithm to numerically solve a general class of discrete-time control problem with finite horizon, and then apply it on our trading problem. We then present some results of numerical tests on simulated order book. Section 5 presents an extension of our model when order arrivals are driven by Hawkes processes, and finally the appendix collects some results used in the paper.

2. Model setup

2.1. *Order book representation*

We consider a model of the order book inspired by the one introduced in chapter 6 of Abergel *et al.* (2016).

Let us fix $K \geq 0$. An order book is supposed to be fully described by K *limits*[1] on the bid side and K limits on the ask side. Denote by pa_t the *best ask* at time t, which is the cheapest price a participant in the market is willing to sell a stock at time t, and by pb_t the *best bid* at time t, which is the highest price a participant in the market is willing to buy a stock at time t. We use the pair of vectors $(\underline{a}_t, \underline{b}_t) = (a_t^1, ..., a_t^K, b_t^1, \ldots, b_t^K)$ where

- a_t^i is the number of shares available i ticks away from pb_t,
- -b_t^i is the number of shares available i ticks away from pa_t,

to describe the order book. The vectors $\underline{a}_t$ and $\underline{b}_t$ describe respectively the ask and the bid sides at time t. The quantities a_t^i, $1 \leq i \leq K$, live in the discrete space $q\mathbb{N}$ where $q \in \mathbb{R}^*$ is the minimum order size on each specific market (*lot size*). The quantities b_t^i, $1 \leq i \leq K$, live in the discrete space $-q\mathbb{N}$. By convention, the a^i are non-negative, and the b^i are non-positive for $0 \leq i \leq K$. The tick size ϵ represents the smallest intervall between different price levels. We assume in the sequel that the orders arrivals have the same size $q = 1$, and set the tick size to $\epsilon = 1$ for simplicity.

Constant boundary conditions are imposed outside the moving frame of size $2K$ in order to guarantee that both sides of the LOB are never empty: we assume that all the limits up to the K-th ones are equal to a_∞ in the ask side, and equal to b_∞ in the bid side, with $a_\infty, -b_\infty \in \mathbb{N}$. The order book can receive at any time three different kinds of orders from general market participants: market orders, limit orders and cancel orders. The orders arrivals are modeled by the following point processes:

- M^+ stands for the buy market orders flow, and we denote by λ^{M^+} its intensity,
- M^- stands for the sell market orders flow, and we denote by λ^{M^-} its intensity,
- L_i^+, for $i \in \{1, ...K\}$, stands for the sell orders flow at the i^{th} limit on the ask side, and we denote by $\lambda_i^{L^+}$ its intensity,

[1] limit is also referred to as *quote* or *price level* in the literature.

- L_i^-, for $i \in \{1, ...K\}$, stands for the buy orders flow at the i^{th} limit on the bid side, and we denote by $\lambda_i^{L^-}$ its intensity,
- C_i^+, for $i \in \{1, ...K\}$, stands for the cancel orders flow at the i^{th} limit on the ask side, and we denote by $\lambda_i^{C^+}$ its intensity,
- C_i^-, for $i \in \{1, ...K\}$, stands for the cancel orders flow at the i^{th} limit on the bid side, and we denote by $\lambda_i^{C^-}$ its intensity.

We assume in the sequel that

(Harrivals) The orders arrivals from general market participants (market orders, limit orders and cancel orders) occur according to Markov jump processes which intensities only depends on the couple $\big(\underline{a}, \underline{b}\big)$. Moreover, we assume that the all the intensities are at most linear w.r.t. the couple $\big(\underline{a}, \underline{b}\big)$ and are constant between two events.

Under **(Harrivals)**, Let λ^L, λ^C, λ^M be positive real constants such that

$$\sum_{i=1}^{K} \lambda_i^{L^+}(\underline{a}, \underline{b}) + \sum_{i=1}^{K} \lambda_i^{L^-}(\underline{a}, \underline{b}) \leq \lambda^L \big(|\underline{a}| + |\underline{b}|\big),$$
$$\sum_{i=1}^{K} \lambda_i^{C^+}(\underline{a}, \underline{b}) + \sum_{i=1}^{K} \lambda_i^{C^-}(\underline{a}, \underline{b}) \leq \lambda^C \big(|\underline{a}| + |\underline{b}|\big),$$
$$\lambda^{M^+}(\underline{a}, \underline{b}) + \lambda^{M^-}(\underline{a}, \underline{b}) \leq \lambda^M \big(|\underline{a}| + |\underline{b}|\big),$$

for all state $(\underline{a}, \underline{b})$ of the LOB, where $|a| := \sum_{k=1}^{K} -a^k$ and $|b| := \sum_{k=1}^{K} b^k$.

REMARK 2.1 *The linear conditions on the intensities are required to prove that the control problem is well-posed.*

REMARK 2.2 *We assume the intensity to be constant between jumps in* ***(Harrivals)*** *for simplicity. All the results proposed in Section 3 can be extended to the case where the intensities of the jump processes are deterministic between jumps. Such an extension is considered in Section 5, where the arrivals are modeled using Hawkes processes with exponential kernel.*

REMARK 2.3 *Some information can be integrated to the order book model by adding new processes. For example, some exogenous processes that send orders to the best-ask and best-bid limits can be added to model the predictions of the mid-price and its volatility that an agent may have. Doing so is critical to manage the risk-reward tradeoffs.*

2.2. *Market maker strategies*

We assume that the order book matching is done on price/time priority, which means that each limit of the order book is a queue where the first order in the queue is the first one to be executed[1].

We consider a market maker who stands ready to send buy and sell limit orders on a regular and continuous basis at quoted prices. A usual assumption in stochastic control theory to characterize the value function as solution to a HJB equation is to constrain the control space to be compact. In this spirit, we shall make the following assumption:

(Hcontrol) Assume that at any time, the total number of limit orders placed by the marker maker does not exceed a fixed (possibly large) integer $\bar{M}$.

[1] such an order book is sometimes referred to in the literature as an order book governed by a *FIFO* (*First In First Out*) rule.

2.2.1. **Control of the market maker.** The market maker can choose at any time to keep, cancel or take positions in the order book (as long as she does not hold more than $\bar{M}$ positions in the order book). Her positions are fully described by the following $\bar{M}-$dimensional vectors $\underline{ra}_t$, $\underline{rb}_t$, $\underline{na}_t$, $\underline{nb}_t$ where $\underline{ra}$ (resp. $\underline{rb}$) records the limits in which the market maker's sell (resp. buy) orders are located; and $\underline{na}$ (resp. $\underline{nb}$) records the ranks in the queues of each market maker's sell (resp. buy) orders. In order to guarantee that the strategy of the market maker is predictable w.r.t. the natural filtration generated by the orders arrivals processes, we shall make the following assumption.

(Harrivals2) The intensities do not depend on the control. Moreover, the market maker does not cross the spread.

To simplify the theoritical analysis, we also make the following assumption: **(Harrivals3)** Assume that the market maker does not change their strategy between two orders arrivals of the order book. In other words, the market maker makes a decision right after one of the order arrivals processes $L^{\pm}, C^{\pm}, M^{\pm}$ jumps, and keep it until the next the jump of an order arrival.

REMARK 2.4 *Assumption* ***(Harrivals3)*** *is mild if the order book jumps frequently, since the market maker can change their decisions frequently in such a case. It can also easily be relaxed by considering piecewise constant controls between jumps (which seems well-adapted to most of the time-discretized control problems met in the industry) or any other parametric family of functions. The results and proofs can then be extended, relying mainly on the PDMDP*[1].

We provide in Figure 1 a graphical representation of the controlled LOB. Notice that the market maker interacts with the order book by placing orders at some limits. The latter have ranks that evolve after each orders arrivals.

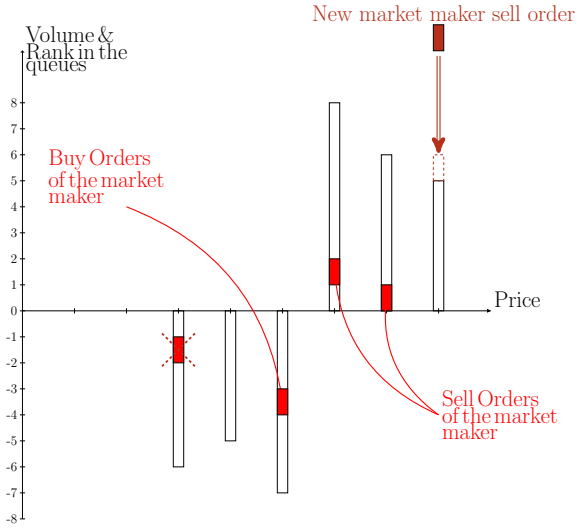


Figure 1.: Example of market maker's placements and decisions she might make. In this example: the ask-side of the order book is described by $\underline{a} = (7, 5, 4)$; and the bid-side by $\underline{b} = (-6, -4, -5)$. The market maker positions are described by $\underline{ra} = (0, 1, -1)$ and $\underline{rb} = (0, 2, -1...)$; and the associated ranks vectors are $\underline{na} = (2, 1, -1\dots)$ and $\underline{nb} = (4, 2, -1)$. We have $i0 = 0$.
After each order arrival, she can send new limit orders (see action on the top right), cancel some positions (see dashed cross on the bottom left), or just keep their orders unchanged.

Denote by $(T_n)_{n\in\mathbb{N}}$ the sequence of jump times of the order book. We denote by $\mathbb{A}$ the set of

[1]PDMDP stands for Piecewise Deterministic Markov Decision Process, and refers to the the control processes which have deterministic dynamics between (random) jumps.

the admissible strategies, defined as the predictable processes $\big(\underline{ra}_t, \underline{rb}_t\big)_{t\leq T}$ such that the control is constant between two consecutive arrivals of orders from the participants, and such that the order of the market maker do not cross the spread. These conditions reads:

- for all $n \in \mathbb{N}$, $\big(\underline{ra}_t, \underline{rb}_t\big) \in \{1, ..., K\}^{\bar{M}} \times \{1, ..., K\}^{\bar{M}}$ are constant on $\big(T_n, T_{n+1}\big]$
- $ra_*, rb_* \geq i0$

where, for every vector $\underline{a}$: $a_* = \min\limits_{1\leq i\leq K}\{a_i \ s.t. \ \underline{a}_i \neq -1\}$; and: $i0 = \operatorname{argmin}\limits_{1\leq i\leq K}\big(\underline{a}_i \ s.t. \ a_i > 0\big)$. The control is the double vector of the positions of the $\bar{M}$ market maker's orders in the order book.

By convention, we set in the sequel $ra_i(t) = -1$ if the ith market maker's order is not placed in the order book.

2.2.2. **Controlled order book.** The order book, controlled by the market maker, is fully described by the following state process Z:

$$Z_t := \big(X_t, Y_t, \underline{a}_t, \underline{b}_t, \underline{na}_t, \underline{nb}_t, pa_t, pb_t, \underline{ra}_t, \underline{rb}_t\big),$$

where, at time t:

- X_t is the cash held by the market maker on a zero interest account.
- Y_t is the inventory of the market maker, i.e. it is the (signed) number of shares held by the market maker.
- pa_t is the ask price, i.e. the cheapest price a general market participant is willing to sell stock.
- pb_t is the bid price, i.e. the highest price a general market participant is willing to buy stock.
- $\underline{a}_t = (a_1(t), \ldots, a_K(t))$ (resp. $\underline{b}_t = (b_1(t), \ldots, b_K(t))$) describes the ask (resp. bid) side: $i \in \{1, \ldots, K\}$, $a_i(t)$ is the sum of all the general market participants' sell orders which are i ticks away from the bid (resp. ask) price.
- $\underline{ra}_t$ (resp. $\underline{ra}_t$) describes the market maker's orders in the ask (resp. bid) side: for $i \in \{1, ..., \bar{M}\}$, $\underline{ra}_t(i)$ is the number of ticks between the i-th market maker's sell (resp. bid) order and the bid (resp. ask) price. By convention, we set $\underline{ra}_t(i) = -1$ (resp. $\underline{rb}_t(i) = -1$) if the i-th sell (resp. buy) order of the market maker is not placed in the order book. As a result $\underline{ra}_t(i), \underline{rb}_t(i) \in \{1, \ldots, K\} \cup \{-1\}$.
- $\underline{na}_t$ (resp. $\underline{nb}_t$) describes the ranks of the market maker's orders in the ask (resp. bid) side. For $i \in \{1, ..., \bar{M}\}$, $\underline{na}_t(i) \in \big\{-1, ..., |a| + \bar{M}\big\}$ (resp. $\underline{nb}_t(i) \in \big\{-1, ..., |b| + \bar{M}\big\}$) is the rank of the i-th sell (resp. buy) orders of the market maker in the queue. By convention, we assume that $\underline{na}_t(i) = -1$ (resp. $\underline{nb}_t(i) = -1$) if the i-th sell (resp. buy) order of the market maker is not placed in the order book.

3. Presentation of the market-making problem. Theoretical resolution.

3.1. *Definition of the market-making problem and well-posedness of the value function*

We denote by V the value function for the following market-making problem:

$$V(t,z) = \sup_{\alpha\in\mathbb{A}} \ \mathbb{E}^{\alpha}_{t,z}\left[\int_t^T f\big(\alpha_s, Z_s\big)\,\mathrm{d}s + g\big(Z_T\big)\right], \quad (t,z) \in [0,T] \times E, \tag{3.1}$$

where:

- $\mathbb{A}$ is the set of the admissible strategies, defined in Section 2.2.1.
- f and g are respectively the instantaneous and terminal reward functions.
- $\mathbb{E}^{\alpha}_{t,z}$ stands for the expectation conditioned by $Z_t = z$ and when strategy $\alpha = (\alpha_s)_{t \leq s < T}$ is followed on $[t, T]$.

EXAMPLE 3.1 *The terminal reward g can be defined as the sum of the market maker's terminal wealth function and an inventory penalization term, i.e. $g : z \mapsto x + L(y) - \eta y^2$ where L is the amount earned from the immediate liquidation of the inventory*[1]*. We remind that z stands for a state of order book; ε is the tick size of the LOB; x is the value of the risk-free account of the market maker; η is the penalization parameter of the market maker; and where we remind that y stands for the (signed) market maker's inventory.*

The running reward f can stand for a penalization of inventory term: $f(z) := -\gamma y^2$, with $\gamma > 0$.

We shall assume the following conditions on the rewards to insure the well-posedness of the market-making problem.

(Hrewards) The expectation of the integrated running reward is uniformly upper-bounded w.r.t. the strategies in $\mathbb{A}$, i.e.

$$\sup_{\alpha \in \mathbb{A}} \mathbb{E}^{\alpha}_{t,z} \left[\int_t^T f^+(Z_s, \alpha_s) \, \mathrm{d}s \right] < +\infty$$

holds; where for all state z and action a, we denote $f^+(z, a) := \max(f(z, a), 0)$. Moreover, the terminal reward $g(Z_T)$ is a.s. at most linear with respect to the number of events up to time T, denoted by N_T in the sequel, i.e. there exists a constant $c_1 > 0$ such as $g(Z_T) \leq c_1 N_T$, a.s..

REMARK 3.1 *Under Assumption **(Hcontrol)**, Assumption **(Hrewards)** holds when g is defined as the wealth of the market maker plus an inventory penalization. In particular, we have $g(Z_T) \leq N_T \bar{M}$, where $\bar{M}$ is the maximal number of orders that can be sent by the market maker, which holds a.s. since the best profit the market maker can make is when their buy (resp. sell) limit orders are all executed, and then the price keeps going to the right (resp. left) direction. Hence the second condition of **(Hrewards)** holds with $c_1 = \bar{M}$.*

The following Lemma 3.1 tackles the well-posedness of the control problem.

LEMMA 3.1 *Under **(Hrewards)** and **(Hcontrol)**, the value function is well-defined, i.e.*

$$\sup_{\alpha \in \mathbb{A}} \mathbb{E}^{\alpha}_{t,z} \left[g(Z_T) + \int_t^T f\big(\alpha_s, Z_s\big) \, \mathrm{d}s \right] < +\infty,$$

where, as defined previously, $\mathbb{E}^{\alpha}_{t,z}[.]$ stands for the expectation conditioned by the event $\{Z_t = z\}$, assuming that strategy $\alpha \in \mathbb{A}$ is followed in $[t, T]$.

---

[1] L is defined as follows:

$$L(z) = \begin{cases} \sum_{k=1}^{\jmath-1} \big[a_k(pa + k\epsilon)\big] + (y - a_0 - \ldots - a_{\jmath-1})(pa + \jmath\epsilon) & \text{if } y < 0 \\ -\sum_{k=1}^{\jmath-1} \big[b_k(pb - k\epsilon)\big] + (y + b_0 + \ldots + b_{\jmath-1})(pb - \jmath\epsilon) & \text{if } y > 0 \\ 0 & \text{if } y = 0, \end{cases}$$

for all $z = \big(x, y, \underline{a}, \underline{b}, \underline{na}, \underline{nb}, pa, pb, \underline{ra}, \underline{rb}\big)$, and where we define:

$$\jmath := \begin{cases} \min\big\{j \big| \sum_{i=1}^{j} a_i > -y\big\} & \text{if } y < 0 \\ \min\big\{j \big| \sum_{i=1}^{j} |b_i| > y\big\} & \text{if } y > 0. \end{cases}$$

Proof. Denote by $(N_t)_t$ the sum of all the arrivals of orders up to time t. Under **(Hrewards)**, we can bound $\mathbb{E}^{\alpha}_{t,z}\left[\int_t^T f\big(\alpha_s, Z_s\big)\,\mathrm{d}s + g(Z_T)\right]$, the reward functional at time t associated to a strategy $\alpha \in \mathbb{A}$, as follows:

$$\begin{aligned}\mathbb{E}^{\alpha}_{t,z}\left[\int_t^T f\left(\alpha_s, Z_s\right)\mathrm{d}s + g(Z_T)\right] &\leq \sup_{\alpha\in\mathbb{A}} \mathbb{E}^{\alpha}_{t,z}\left[g(Z_T)\right] + \sup_{\alpha\in\mathbb{A}} \mathbb{E}^{\alpha}\left[\int_t^T f^{+}(Z_s,\alpha_s)\,\mathrm{d}s\right] \\ &\leq c_1 \sup_{\alpha\in\mathbb{A}} \mathbb{E}^{\alpha}_{t,0}\left[N_T\right] + \sup_{\alpha\in\mathbb{A}} \mathbb{E}^{\alpha}_{t,z}\left[\int_t^T f^{+}(Z_s,\alpha_s)\,\mathrm{d}s\right], \qquad (3.2)\end{aligned}$$

where once again, for all general process M and all $m \in E$, $\mathbb{E}^{\alpha}_{t,m}[M_T]$ stands for the expectation of M_T conditioned by $M_t = m$ and assuming that the market maker follows strategy $\alpha \in \mathbb{A}$ in $[t, T]$. Let us show that the first term in the r.h.s. of (3.2) is bounded. On one hand, we have:

$$\mathbb{E}^{\alpha}_{t,0}\left[N_T\right] \leq \|\lambda\|_{\infty} \int_0^T \mathbb{E}\left(|a|_t + |b|_t\right)\mathrm{d}t, \qquad (3.3)$$

where $\|\lambda\|_{\infty} := \lambda^L + \lambda^C + \lambda^M$ is a bound on the intensity rate of N_t. On the other hand, there exists a constant $c_2 > 0$ such that $\mathrm{d}(|a| + |b|)_t \leq c_2\,\mathrm{d}L_t$ so that: $\mathbb{E}^{\alpha}_{t,|a|_0+|b|_0}\left[|a|_t + |b|_t\right] \leq |a|_0 + |b|_0 + c_3\int_0^t \mathbb{E}\left[|a|_s + |b|_s\right]\mathrm{d}s$. Applying Gronwall's inequality, we then get:

$$\mathbb{E}^{\alpha}_{t,|a|_0+|b|_0}\left[|a|_t + |b|_t\right] \leq \left(|a|_0 + |b|_0\right)e^{c_3 t}. \qquad (3.4)$$

Plugging (3.4) into (3.3) finally leads to:

$$\mathbb{E}^{\alpha}_{t,0}\left[N_T\right] \leq c_4 e^{c_3 T}$$

wit c_3 and $c_4 > 0$ that do not depends on α, which proves that the first term in the r.h.s. of (3.2) is bounded. Also, its second term in the r.h.s. of (3.2) is bounded under **(Hrewards)**. Hence, the reward functional is bounded uniformly in α, which proves that the value function of the considered market-making problem is well-defined. □

3.2. *Markov Decision Process formulation of the market-making problem*

In this section, we first reformulate the market-making problem as a Markov Decision Process (MDP), and then characterize the value function as solution to a Bellman equation.

Let us denote (T_n) is the increasing sequence of the arrivals of market/limit/cancel order to the market; and let $Z_n := \phi^{a(Z_{T_n})}(Z_{T_n})$, where $\phi^a(z) \in E$ is the state of the order book at time t such that $T_n < t < T_{n+1}$, given that $Z_{T_n} = z$ and given that the strategy a has been chosen by the market maker at time T_n.

Let us consider the Markov Decision Process $(T_n, Z_n)_{n\in\mathbb{N}}$, which is characterized by the following information

$$\underbrace{[0,T]\times E}_{\text{state space}}\quad,\quad \underbrace{A_z}_{\text{market maker control}}\quad,\quad \underbrace{\lambda}_{\text{intensity of the jump}}\quad,\quad \underbrace{Q}_{\text{transitions kernel}}\quad,\quad \underbrace{r}_{\text{reward}}$$

where:

- $[0, T] \times E$ is the state space of the time-continuous controlled process $(T_n, Z_n)_{n\in\mathbb{N}}$; and $E :=$

$\mathbb{R}\times\mathbb{N}\times\mathbb{N}^K\times\mathbb{N}^K\times\mathbb{N}^{\bar{M}}\times\mathbb{N}^{\bar{M}}\times\mathbb{N}^{\bar{M}}\times\mathbb{N}^{\bar{M}}\times\mathbb{R}\times\mathbb{R}$ is the state space of (Z_t). For $z\in E$, $z=\big(x,y,\underline{a},\underline{b},\underline{na},\underline{nb},\underline{ra},\underline{rb},pa,pb\big)$ where: x is the cash held by the market maker, y their inventory; $\underline{a}$ and $\underline{b}$, introduced in Section 2.2.2. represent the orders in the ask and bid sides of the order book of all the participants except the market maker's; $\underline{na}$ (resp. $\underline{nb}$) is the $\bar{M}$-dimensional vector of the ranks of the market maker's sell (resp. buy) orders in the queues ; $\underline{ra}$ (resp. $\underline{rb}$) is the $\bar{M}$-dimensional vector of the number of ticks the $\bar{M}$ market maker's sell (resp. buy) orders are from the bid (resp. ask) price; pa (resp. pb) is the ask-price (resp. bid-price).

- A_z, for every state $z\in E$, is the set of the admissible actions (i.e.the actions the market maker can take) when the order book is at state z:

$$A_z=\Big\{\underline{ra},\underline{rb}\in\{1,...,K\}^{\bar{M}}\times\{1,...,K\}^{\bar{M}}\Big|rb_*,ra_*\geq i0\Big\},$$

where we define $c_*=\min_{1\leq i\leq K}\{c_i|c_i\neq -1\}$ and $c0=\underset{1\leq i\leq K}{\operatorname{argmin}}\{\underline{c}_i>0\}$ for $\underline{c}\in\mathbb{N}^{\bar{M}}$. We recall that this condition means that the market maker is not allowed to cross the spread.

- λ is the intensity of the controlled process (Z_t), and reads:

$$\lambda(z):=\lambda^{M^+}(z)+\lambda^{M^-}(z)+\sum_{1\leq j\leq K}\lambda^{L_j^+}(z)+\sum_{1\leq j\leq K}\lambda^{L_j^-}(z)+\sum_{1\leq j\leq K}\lambda^{C_j^+}(z)+\sum_{1\leq j\leq K}\lambda^{C_j^-}(z).$$

Observe that λ does not depend on the strategy α chosen by the market maker since we assumed that the general participants does not "see" the market maker's orders in the order book. Although we wrote z as argument for he intensity of the order book process, it cannot depend on any controlled component variable of the latter. To simplify, the reader can assume that the intensities only depend on the vectors $\underline{a}$ and $\underline{b}$.

- Q is the transition kernel of the MDP, which is defined as follows:

$$Q\big(B\times C|t,z,\alpha\big):=\lambda(z)\int_0^{T-t}e^{-\lambda(z)s}\mathbf{1}_B(t+s)Q'\big(C|\phi^\alpha(z),\alpha\big)\,\mathrm{d}s+e^{-\lambda(z)(T-t)}\mathbf{1}_{T\in B,z\in C},\quad(3.5)$$

for all Borelian sets $B\subset\mathbb{R}_+$ and $C\subset E$, for all $(t,z)\in[0,T]\times E$, for all $\alpha\in A$, and where Q' is the transition kernel of (Z_t) defined for all state z as:

$$Q'\big(z'|z,u\big)=\begin{cases}\frac{\lambda^{M^+}(z)}{\lambda(z)} & \text{if } z'=e^{M^+}(\phi^u(z))\\ \quad\vdots & \\ \frac{\lambda^{C^+}(z)}{\lambda(z)} & \text{if } z'=e^{C_K^+}(\phi^u(z)),\end{cases}$$

where $\phi^u(z)$ is the new state of the controlled order book when decision u as been taken and when the order book was at state z before the decision; $e^{M^+}(z)$ is the new state of the order book right after it received a buy market order, given that it was at state z before the jump; and $e^{C_i^\pm}(z)$ is the new state of the order book right after it received a cancel order from a general market participant on its i^{th} ask/bid limit, given that it was at state z.

- $r:[0,T]\times E^C\to\mathbb{R}$ is the running reward associated to the MDP with infinite horizon defined as follows:

$$r(t,z,a):=-c\big(z,a\big)e^{-\lambda(z)(T-t)}(T-t)\mathbf{1}_{t>T}+c\big(z,a\big)\Big(\frac{1}{\lambda(z)}-\frac{e^{-\lambda(z)(T-t)}}{\lambda(z)}\Big)+e^{-\lambda(z)(T-t)}g(z)\mathbf{1}_{t\leq T},\quad(3.6)$$

and its definition is motivated by Proposition 3.1 below.

The cumulated reward functional associated to the MDP $(T_n, Z_n)_{n\in\mathbb{N}}$ for an admissible policy $(f_n)_{n=0}^{\infty}$ is defined as:

$$V_{\infty,(f_n)}(t,z) = \mathbb{E}_{t,z}^{(f_n)}\left[\sum_{n=0}^{\infty} r\big(T_n, Z_n, f_n(T_n, Z_n)\big)\right],$$

and the associated value function is the supremum of the cumulated reward functional over all the admissible controls in $\mathbb{A}$, i.e.

$$V_{\infty}(t,z) = \sup_{(f_n)_{n=0}^{\infty}\in\mathbb{A}} V_{\infty,\alpha}(t,z), \qquad (t,z)\in[0,T]\times E, \tag{3.7}$$

Notice that we used the same notation for admissible controls of the MDP and those of the continuous-time control problem.

REMARK 3.2 *Q is defined as in (3.5) because*

$$\begin{aligned}\mathbb{P}\big(T_{n+1}-T_n\le t, Z_{n+1}\in B | T_0, Z_0, \dots, T_n, Z_n\big) &= \lambda(Z_n)\int_0^t e^{-\lambda(Z_n)s} Q'\big(B|Z_{T_n}, \alpha_{T_n}\big)\,\mathrm{d}s \\ &= \lambda(Z_n)\int_0^t e^{-\lambda(Z_n)s} Q'\big(B|Z_{T_n}, f_n(Z_n)\big)\,\mathrm{d}s,\end{aligned}$$

holds for any admissible policy $\alpha = (f_n)_{n=0}^{\infty}\in\mathbb{A}$, for all Borelian $B\subset E$, and for all $t\in[0,T]$.

In the sequel, we denote $\big([0,T]\times E\big)^C := \Big\{\big(t,z,a\big)\in E\times\{1,\dots,K\}^{2\bar{M}} \big| t\in[0,T], z\in E, a\in A_z\Big\}$, and $E^C := \Big\{\big(z,a\big)\in E\times\{1,\dots,K\}^{2\bar{M}} \big| z\in E, a\in A_z\Big\}$. Q' is the stochastic kernel from E^C to E that describes the distribution of the jump goals, i.e., $Q'\big(B|z,u\big)$ is the probability that the order book jumps in the set B given that it was at state $z\in E$ right before the jump, and the control action $u\in A_z$ has been chosen right after the jump time.

REMARK 3.3 *The MDP is defined in such a way that the control is feedback and constant between two consecutive arrivals of market/limit/cancel orders in the market, i.e. in the time-continuous setting: we restrict ourselves to the control $\alpha = (\alpha_t)$ which are entirely characterized by the decision functions $f_n : [0,T]\times E\to A$, and such that*

$$\alpha_t = f_n(T_n, Z_n) \text{ for } t\in\big(T_n, T_{n+1}\big]$$

By abuse of notation, we denote in the sequel by α the sequence of controls $(f_n)_{n=0}^{\infty}$.

The following Proposition 3.1 motivates the special choice of the running reward r as defined in (3.6):

PROPOSITION 3.1 *The value function of the MDP defined by (3.7) coincides with (3.1), i.e. we have for all $(t,z)\in E^C$:*

$$V_{\infty}(t,z) = V(t,z). \tag{3.8}$$

Proof. Let us show that for all $\alpha = (f_n)\in\mathbb{A}$ and all $(t,z)\in E^C$

$$V_{\alpha}(t,z) = V_{\infty}^{(f_n)}(t,z). \tag{3.9}$$

Let us first denote by $H_n := (T_0, Z_0, ..., T_n, Z_n)$. Notice then that for all admissible strategy α:

$$V_\alpha(t,z) = \mathbb{E}^\alpha_{t,z}\Bigg[\sum_{n=0}^{\infty} \mathbf{1}_{T>T_{n+1}}\big(T_{n+1}-T_n\big)c\big(Z_n,\alpha_n\big) + \mathbf{1}_{[T_n\le T<T_{n+1})}\Big(g(Z_T) - \eta {Y_T}^2 + (T-T_n)c\big(Z_n,\alpha_n\big)\Big)\Bigg]$$
$$= \sum_{n=0}^{\infty} \mathbb{E}^{(f_n)}_{t,z}\Big[r\big(T_n, Z_n, f_n(T_n,Z_n)\big)\Big], \tag{3.10}$$

where we conditioned by H_n between the first and the second line. We recognize $V^{(f_n)}_\infty$ in the r.h.s. of (3.10), so that the proof of (3.9) is completed.

It remains to take the supremum over all the admissible strategies $\mathbb{A}$ in (3.9) to get (3.8). □

From Proposition 3.1, we deduce that the value function of the market-making problem is the same as the value function V_∞ of the discrete-time MDP with infinite horizon. We now aim at solving the MDP control problem. To proceed, we first define the maximal reward mapping for the infinite horizon MDP:

$$(\mathcal{T}v)(t,z) := \sup_{a\in A_z}\Big\{r(t,z,a) + \int v(t',z')Q(t',z'|t,\phi^a(z),a)\Big\}$$
$$= \sup_{a\in A_z}\Big\{r(t,z,a) + \lambda(z)\int_0^{T-t} e^{-\lambda(z)s}\int v(t+s,z')Q'\big(dz'|\phi^a(z),a\big)ds\Big\}, \tag{3.11}$$

where we recall that:

- $\phi^\alpha(z)$ is the new state of the order book when the market maker follows the strategy α and the order book is at state z before the decision is taken.
- $\lambda(z)$ is the intensity of the order book process given that the order book is at state z.

We shall tighten assumption **(Hrewards)** in order to guarantee existence and uniqueness of a solution to (3.1), as well as characterizing the latter.

(HrewardsBis): The running and terminal rewards are at most quadratic w.r.t. the state variable, uniformly w.r.t. the control variable, i.e.

(i) The running reward f is such that $|c|$ is uniformly bounded by a quadratic in z function, i.e. there exists $c_5 > 0$ such that:

$$\forall (z,a)\in E\times A,\quad |f(z,a)| \le c_5(1+|z|^2).$$

(ii) The terminal reward g has no more than a quadratic growth, i.e. there exists $c_6 > 0$ such that:

$$\forall z\in E,\quad |g(z)| \le c_6(1+|z|^2).$$

REMARK 3.4 *Assumption **(HrewardsBis)** holds in the case where g is the terminal wealth of the market maker plus a penalization of their inventory, and where with no running reward, i.e. $f = 0$.*

The main result of this section is the following theorem that gives existence and uniqueness of a solution to (3.1), and moreover characterizes the latter as fixed point of the maximal reward operator defined in (3.11).

THEOREM 3.1 *$\mathcal{T}$ admits a unique fixed point v which coincides with the value function of the MDP. Moreover we have:*

$$v = V_\infty = V.$$

Denote by f^* the maximizer of the operator $\mathcal{T}$. Then $\big(f^*, f^*, ...\big)$ is an optimal stationary (in the MDP sense) policy.

REMARK 3.5 *Theorem 3.1 states that the optimal strategy is stationary in the MDP formulation of the problem, but of course, it is not stationary for the original time-continuous trading problem with finite horizon (3.1), since the time component is not a state variable anymore in the original formulation. Actually, given $n \in \mathbb{N}$ and the state of order book z at that time, the optimal decision to take at time T_n is given by $f^*\big(T_n, z\big)$.*

We devote the next section to the proof of Theorem 3.1.

3.3. *Proof of Theorem 3.1*

Remind first that we defined in the previous section $E^C := \Big\{\big(z,a\big) \in E \times \{1,\dots,K\}^{2\bar{M}} \big| z \in E, a \in A_z\Big\}$ and $\big([0,T]\times E\big)^C := \Big\{\big(t,z,a\big) \in [0,T]\times E \times \{1,\dots,K\}^{2\bar{M}} \big| t \in [0,T], z \in E, a \in A_z\Big\}$.

DEFINITION 3.1 *A measurable function $b : E \to \mathbb{R}_+$ is called a* bounding function *for the controlled process (Z_t) if there exists positive constants c_c, c_g, $c_{Q'}, c_\phi$ such that:*

(i) $|f(z,a)| \le c_c b(z)$ *for all* $(z,a) \in E^C$.
(ii) $|g(z)| \le c_g b(z)$ *for all* z *in* E.
(iii) $\int b(z')Q'(dz'|z,a) \le c_{Q'} b(z)$ *for all* $(z,a) \in E^C$.
(iv) $b(\phi_t^\alpha(z)) \le c_\phi b(z)$ *for all* $(t,z,\alpha) \in \big([0,T]\times E\big)^C$.

PROPOSITION 3.2 *Let b be such that :*

$$\forall z \in E, b(z) := 1 + |z|^2.$$

Then, b is a bounding function for the controlled process (Z_t), under Assumption ***(HrewardsBis)****.*

Proof. Let us check that b defined in Proposition 3.2 satisfies the four assertions in Definition 3.1.

- Assertion 1 and 2 of Definition 3.1 holds under **(HrewardsBis)**.
- First notice that $\underline{ra}, \underline{rb}$ are bounded by $\sqrt{\bar{M}}K$ (where we recall that K is the number of limits in each side of the order book, and $\bar{M}$ is the biggest number of limit orders that the market maker is allowed to send in the market). Secondly, $pa' \in B(pa, K), pb' \in B(pb, K)$, where $B(x,r)$ is the ball centered in x with radius $r > 0$, because of the limit conditions that we imposed in our LOB model. And last, we can see that $|\underline{a}'| \le |\underline{a}| + a_\infty K$. These three bounds are linear w.r.t. z so that assertion 3 holds.
- $\phi^\alpha(z) = z^\alpha$ only differs from z by its $\underline{na}$, $\underline{nb}$, and $\underline{ra}$, $\underline{rb}$ components.
  But $|\underline{na}| \le \sqrt{\bar{M}}\big(|\underline{a}| + \bar{M}\big)$ and $|\underline{nb}| \le \sqrt{\bar{M}}\big(|\underline{b}| + \bar{M}\big)$ are bounded by a linear function of $(\underline{a}, \underline{b})$,

also $|\underline{ra}|$ and $|\underline{rb}|$ are bounded by the universal constant $\sqrt{\bar{M}}K$, so assertion 4 in Definition 3.1 holds.

□

Let us define

$$\Lambda := (4K+2)\sup\left\{\frac{\lambda^{M^\pm}}{|\underline{a}|+|\underline{b}|}, \frac{\lambda^{L^\pm}}{|\underline{a}|+|\underline{b}|}, \frac{\lambda^{C^\pm}}{|\underline{a}(z)|+|\underline{b}(z)|}\right\},$$

which is well-defined under **(Harrivals)**.

PROPOSITION 3.3 *If b is a bounding function for (Z_t), then*

$$b(t,z) := b(z)e^{\gamma(z)(T-t)}, \text{ with } \gamma(z) = \gamma_0(4K+2)\Lambda\big(1+|\underline{a}|+|\underline{b}|\big) \text{ and } \gamma_0 > 0$$

is a bounding function for the MDP, i.e. for all $t \in [0,T], z \in E, a \in A_z$, we have:

$$|r(t,z,a)| \le c_g b(t,z),$$
$$\int b(s,z')Q(ds,dz'|t,z,a) \le c_\phi c_Q e^{C(T-t)}\frac{1}{1+\gamma_0}b(t,z),$$

with $C = \gamma_0\Lambda K(4K+2)\big(|a|_\infty + |b|_\infty\big)$.

Proof. Let $z' = \big(x', y', \underline{a}', \underline{b}', \underline{na}', \underline{nb}', \underline{ra}', \underline{rb}'\big)$ be the state of the order book after an exogenous jump occurs given that it was in state z before the jump. Since $|\underline{a}'| \le |\underline{a}| + a_\infty K$ and $|\underline{b}'| \le |\underline{b}| + b_\infty$, where a_∞ and b_∞ are defined as the border conditions of the order book, we have:

$$\gamma(z') \le \gamma(z) + C, \tag{3.12}$$

with $C = \gamma_0\Lambda K(4K+2)(a_\infty + b_\infty)$. Then, we get:

$$\begin{aligned}
\int b(s,z')Q(ds,dz'|t,\phi^\alpha(z),\alpha) &= \lambda(z)\int_0^{T-t} e^{-\lambda(z)s}\int b(t+s,z')Q'\big(dz'|\phi_s^\alpha(z),\alpha\big)\,\mathrm{d}s \\
&= \lambda(z)\int_0^{T-t} e^{-\lambda(z)s}\int b(z')e^{\gamma(z')(T-(t+s))}Q'\big(dz'|\phi_s^\alpha(z),\alpha\big)\,\mathrm{d}s \\
&\le \lambda(z)\int_0^{T-t} e^{-\lambda(z)s}\int b(z')e^{(\gamma(z)+C)(T-(t+s))}Q'\big(dz'|\phi_s^\alpha(z),\alpha\big)ds \\
&\le \lambda(z)\int_0^{T-t} e^{-\lambda(z)s}e^{(\gamma(z)+C)(T-(t+s))}\int b(z')Q'\big(dz'|\phi_s^\alpha(z),\alpha\big)ds \\
&\le \lambda(z)\int_0^{T-t} e^{-\lambda(z)s}e^{(\gamma(z)+C)(T-(t+s))}c_Q c_\phi b(z)ds \\
&\le \frac{\lambda(z)c_Q c_\phi}{\lambda(z)+\gamma(z)+C}e^{(\gamma(z)+C)(T-t)}\Big(1-e^{-(T-t)(\lambda(z)+\gamma(z)+C)}\Big)b(z) \\
&\le c_Q c_\phi\frac{\lambda(z)}{\lambda(z)+\gamma(z)+C}e^{C(T-t)}\Big(1-e^{-(T-t)(\lambda(z)+\gamma(z)+C)}\Big)b(t,z),
\end{aligned}$$

where we applied (3.12) at the third line. It remains to notice that

$$\frac{\lambda(z)}{\lambda(z)+\gamma(z)+C} = \frac{\lambda(z)}{\lambda(z)\big(1+\gamma_0\big)+\gamma_0\big[\underbrace{\Lambda(|a|+|b|)-\lambda(z)}_{\geq 0}\big]} \leq \frac{1}{1+\gamma_0},$$

to complete the proof of the proposition. □

Let us denote by $\|.\|_b$ the *weighted supremum norm* such that for all measurable function $v : E' \to \mathbb{R}$,

$$\|v\|_b := \sup_{(t,z)\in E'} \frac{|v(t,z)|}{b(t,z)},$$

and define the set:

$$\mathbb{B}_b := \Big\{ v : E' \to \mathbb{R} | v \text{ is measurable and } \|v\|_b < \infty \Big\}.$$

Moreover let us define

$$\alpha_b := \sup_{(t,z,\alpha)\in E'\times\mathcal{R}} \frac{\int b(s,z')Q(ds,dz'|t,\phi^\alpha(z),\alpha)}{b(t,z)}.$$

From the preceding estimations we can bound α_b as follows:

$$\alpha_b \leq c_Q c_\phi \frac{1}{1+\gamma_0} e^{CT},$$

so that, by taking: $\gamma_0 = c_Q c_\phi e^{CT}$, we get: $\alpha_b < 1$. In the sequel, we then assume w.l.o.g. that $\alpha_b < 1$. Recall that the maximal reward mapping for the MDP has been defined as:

$$\mathcal{T}v : (t,z) \mapsto \sup_{a\in A_z} \left\{ r(t,z,a) + \lambda(z) \int_0^{T-t} e^{-\lambda(z)s} \int v(t+s,z')Q'\big(dz'|\phi^a(z),a\big)\,\mathrm{d}s \right\}$$

It is straightforward to see that:

$$\|\mathcal{T}v - \mathcal{T}w\|_b \leq \alpha_b \|v-w\|_b, \tag{3.13}$$

which implies that $\mathcal{T}$ is contracting, since $\alpha_b < 1$.

Let $\mathcal{M}$ be the set of all the continuous function in $\mathbb{B}_b$. Since b is continuous, $(\mathcal{M}, \|.\|_b)$ is a Banach space.

$\mathcal{T}$ sends $\mathcal{M}$ to $\mathcal{M}$. Indeed, for all continuous function v in $\mathbb{B}_b$, $(t,z,a) \mapsto r(t,z,a) + \lambda(z)\int_0^{T-t} e^{-\lambda(z)s} \int v(t+s,z')Q'\big(dz'|\phi^a(z),a\big)ds$ is continuous on $[0,T]\times E^C$. A_z is finite, so we get the continuity of the application:

$$\mathcal{T}v : (t,z) \mapsto \sup_{a\in A_z} \left\{ r(t,z,a) + \lambda(z) \int_0^{T-t} e^{-\lambda(z)s} \int v(t+s,z')Q'\big(dz'|\phi^a(z),a\big)\,\mathrm{d}s \right\}.$$

PROPOSITION 3.4 *There exists a maximizer for $\mathcal{T}$, i.e. let $v \in \mathcal{M}$, then there exists a Borelian function $f : [0,T] \times E \to A$ such that for all $(t,z) \in E'$:*

$$\mathcal{T}v\Big(t,z,f\big(t,z\big)\Big) = \sup_{a\in A}\left\{r(t,z,a) + \lambda(z)\int_0^{T-t} e^{-\lambda(z)s}\int v(t+s,z')Q'\big(dz'|\phi^a(z),a\big)\,\mathrm{d}s\right\}$$

Proof. $D^*(t,z) = \Big\{a \in A \big| \mathcal{T}_a v(t,z) = \mathcal{T}v(t,z)\Big\}$ is finite, so it is compact. So $(t,z) \mapsto D^*(t,z)$ is a compact-valued mapping. Since the application $(t,z,a) \mapsto \mathcal{T}_a(t,z) - \mathcal{T}(t,z)$ is continuous, we get that $D^* = \Big\{(t,z,a) \in E'^C \big| \mathcal{T}_a v(t,z) = \mathcal{T}v(t,z)\Big\}$ is borelian. Applying the measurable selection theorem yields to the existence of the maximizer. (see Bäuerle and Rieder (2011) p.352) □

LEMMA 3.2 *The following holds:*

$$\sup_{\alpha\in\mathcal{A}} \mathbb{E}^{\alpha}_{t,z}\left[\sum_{k=n}^{\infty} |r(T_k,Z_k)|\right] \leq \frac{\alpha_b^n}{1-\alpha_b} b(t,z),$$

and in particular, we have:

$$\lim_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \mathbb{E}^{\alpha}_{t,z}\left[\sum_{k=n}^{\infty} \big|r\big(t_k,Z_k\big)\big|\right] = 0.$$

Proof. By conditioning we get $\mathbb{E}^{\alpha}_{t,z}\Big[\big|r(T_k,Z_k)\big|\Big] \leq c_g {\alpha_b}^k b(t,z)$ for $k \in \mathbb{N}$, and for all $\alpha \in \mathbb{A}$. It remains to sum this inequality to complete the proof of Lemma 3.2. □

We can now prove Theorem 3.1.

Proof. We divided the proof of Theorem 3.1 into four steps.
Step 1: Inequality (3.13) and Proposition 3.3 imply that $\mathcal{T}$ is a stable and contracting operator defined on the Banach space $\mathcal{M}$. Banach's fixed point theorem states that $\mathcal{T}$ admits a fixed point, i.e. there exists a function $v \in \mathbb{M}$ such that $v = \mathcal{T}v$, and moreover we have $v = \lim_{n\to\infty} \mathcal{T}^n 0$. Notice that $\mathcal{T}^N 0$ coincides with v_0 defined recursively by the following Bellman equation:

$$\begin{cases} v_N = 0 \\ v_n = \mathcal{T}v_{n+1} \text{ for } n = N-1, ..., 0. \end{cases} \tag{3.14}$$

The solution of the Bellman equation is always larger than the value function of the MDP associated (see e.g. Theorem 2.3.7 p.22 in Bäuerle and Rieder (2011)). Then we have: $\mathcal{T}^n 0 \geq \sup_{(f_k)} \mathbb{E}_n^{(f_k)}\left[\sum_{k=0}^{n-1} r(t_k,X_k)\right] =: J_n$, where J_n is the value function of the MDP with finite horizon n and terminal reward 0, associated to (3.14). Moreover, by Lemma 7.1.4 p.197 in Bäuerle and Rieder (2011), we know that $\big(J_n\big)_n$ converges as $n \to \infty$ to a limit that we denote by J. Passing at the limit in the previous inequality we get: $\lim_{n\to\infty} \mathcal{T}^n 0 \geq J$, i.e.

$$v \geq J. \tag{3.15}$$

Step 2: Let us fix a strategy $\alpha \in \mathbb{A}$, and take $n \in \mathbb{N}$. We denote $J_n(\alpha) := \mathbb{E}_0^{(\alpha_k)}\left[\sum_{k=0}^{n-1} r(t_k,X_k)\right]$, the reward functional associated to the control α on the discrete finite time horizon $\{0,\dots,n\}$. By

definition, we have $J_n(\alpha) \leq J_n$. We get by letting $n \to \infty$: $\lim_{n\to+\infty} J_n(\alpha) =: J_\infty(\alpha) \leq J$. Taking the supremum over all the admissible strategies α finally leads to:

$$V_\infty \leq J. \tag{3.16}$$

Step 3: Let us denote by f a maximizer of $\mathcal{T}$ associated to v, which exists, as stated in Proposition 3.4. v is the fixed point of $\mathcal{T}$ so that $v = \mathcal{T}_f^n(v)$, for $n \in \mathbb{N}$. Moreover $v \leq \delta$ where $\delta := \sup_{\alpha\in\mathcal{A}} \mathbb{E}\Big[\sum_{k=0}^{\infty} r^+(Z_k, \alpha_k)\Big]$, so that $\mathcal{T}_f^n(v) \leq \mathcal{T}_f^n 0 + \mathcal{T}_o^n \delta$, where $\mathcal{T}_o^n \delta = \sup_\alpha \mathbb{E}_n^\alpha\Big[\sum_{k=n}^{\infty} r^+(t_k, Z_k)\Big]$. Lemma 3.2 implies that $\mathcal{T}_o^n \delta \to 0$ as $n \to \infty$. Hence, we get:

$$v \leq J_f. \tag{3.17}$$

Step 4: Conclusion. Since

$$J_f \leq V_\infty \tag{3.18}$$

holds, we get by combining (3.15), (3.16), (3.17) and (3.18):

$$V_\infty \leq J \leq v \leq J_f \leq V_\infty. \tag{3.19}$$

All the inequalities in (3.19) are then equalities, which completes the proof of Theorem 3.1. □

4. Numerical resolution of the market-making control problem

In this section, we first introduce an algorithm to numerically solve a general class of discrete-time control problem with finite horizon, and then apply it on the trading problem (3.1).

4.1. *Framework*

Let us consider a general discrete-time stochastic control problem over a finite horizon $N \in \mathbb{N}\setminus\{0\}$. The dynamics of the controlled state process $Z^\alpha = (Z_n^\alpha)_n$ valued in $\mathbb{R}^d$ is given by

$$Z_{n+1}^\alpha \;=\; F(Z_n^\alpha, \alpha_n, \varepsilon_{n+1}), \quad n = 0, \ldots, N-1, \; Z_0^\alpha = z \in \mathbb{R}^d, \tag{4.1}$$

with $(\varepsilon_n)_n$ is a sequence of i.i.d. random variables valued in some Borel space $(E, \mathcal{B}(E))$, and defined on some probability space $(\Omega, \mathbb{F}, \mathbb{P})$ equipped with the filtration $\mathbb{F} = (\mathcal{F}_n)_n$ generated by the noise $(\varepsilon_n)_n$ ($\mathcal{F}_0$ is the trivial σ-algebra), the control $\alpha = (\alpha_n)_n$ is an $\mathbb{F}$-adapted process valued in $A \subset \mathbb{R}^q$, and F is a measurable function from $\mathbb{R}^d \times \mathbb{R}^q \times E$ into $\mathbb{R}^d$.

Given a running cost function f defined on $\mathbb{R}^d \times \mathbb{R}^q$, a terminal cost function g defined on $\mathbb{R}^d$, the cost functional associated to a control process α is

$$J(\alpha) \;=\; \mathbb{E}\left[\sum_{n=0}^{N-1} f(Z_n^\alpha, \alpha_n) + g(Z_N^\alpha)\right].$$

The set $\mathbb{A}$ of admissible controls is the set of control processes α satisfying some integrability conditions ensuring that the cost functional $J(\alpha)$ is well-defined and finite. The control problem,

also called Markov decision process (MDP), is formulated as

$$V_0(x_0) \ := \ \sup_{\alpha \in \mathcal{A}} J(\alpha), \tag{4.2}$$

and the goal is to find an optimal control $\alpha^* \in \mathcal{A}$, i.e., attaining the optimal value: $V_0(z) = J(\alpha^*)$. Notice that problem (4.1)-(4.2) may also be viewed as the time discretization of a continuous time stochastic control problem, in which case, F is typically the Euler scheme for a controlled diffusion process.

Problem (4.2) is tackled by the dynamic programming approach. For $n = N, \ldots, 0$, the value function V_n at time n is characterized as solution of the following backward (Bellman) equation:

$$\begin{cases} V_N(z) = g(z) \\ V_n(z) = \sup\limits_{a \in A} \left\{ f(z,a) + \mathbb{E}^a_{n,z} \left[V_{n+1}(Z_{n+1}) \right] \right\}, \quad z \in \mathbb{R}^d, \end{cases} \tag{4.3}$$

Moreover, when the supremum is attained in the DP formula at any time n by $a^*_n(z)$, we get an optimal control in feedback form given by: $\alpha^* = (a^*_n(Z^*_n))_n$ where $Z^* = Z^{\alpha^*}$ is the Markov process defined by

$$Z^*_{n+1} = F(Z^*_n, a^*_n(Z^*_n), \varepsilon_{n+1}), \quad n = 0, \ldots, N-1, \quad Z^*_0 = z.$$

There are two usual ways that have been studied in the literature, to solve numerically (4.3): one way is to use local regression methods, relying e.g. on quantization, k-nearest neighbors or kernel ideas to approximate the conditional expectations by cubature methods; another way is to rely on MC regress-now or later methods to regress the value functions V_{n+1} at time n for $n = 0, \ldots, N-1$ on basis functions or neural networks. See e.g. Kharroubi *et al.* (2014) for the regress-now and Balata and Palczewski (2017) for the regress-later methods for algorithms using basis functions, and e.g. Huré *et al.* (2018) for regression on neural networks based on regress-now or regress-later techniques.

4.2. *Presentation and rate of convergence of the Qknn algorithm*

In this section, we present an algorithm based on k-nn estimates for local non-parametric regression of the value function, and optimal quantization to quantize the exogenous noise, in order to numerically solve (4.3).

Let us first introduce some ingredients of the quantization approximation:

- We denote by $\hat{\varepsilon}$ a K-quantizer of the E-valued random variable $\varepsilon_{n+1} \sim \varepsilon_1$, that is a discrete random variable on a grid $\Gamma = \{e_1, \ldots, e_K\} \subset E^K$ defined by

$$\hat{\varepsilon} = \mathrm{Proj}_\Gamma(\varepsilon_1) \ := \ \sum_{\ell=1}^{K} e_l 1_{\varepsilon_1 \in C_i(\Gamma)},$$

  where $C_1(\Gamma)$, ..., $C_K(\Gamma)$ are Voronoi tesselations of Γ, i.e., Borel partitions of the Euclidian space $(E, |.|)$ satisfying

$$C_\ell(\Gamma) \subset \left\{ e \in E : |e - e_\ell| \ = \ \min_{j=1,\ldots,K} |e - e_j| \right\}.$$

The discrete law of $\hat{\varepsilon}$ is then characterized by

$$\hat{p}_\ell := \mathbb{P}[\hat{\varepsilon} = e_\ell] \;=\; \mathbb{P}[\varepsilon_1 \in C_\ell(\Gamma)], \quad \ell = 1, \ldots, K.$$

The grid points (e_ℓ) which minimize the L^2-quantization error $\|\varepsilon_1 - \hat{\varepsilon}\|_2$ lead to the so-called optimal L-quantizer, and can be obtained by a stochastic gradient descent method, known as Kohonen algorithm or competitive learning vector quantization (CLVQ) algorithm, which also provides as a byproduct an estimation of the associated weights $(\hat{p}_\ell)$. We refer to Pagès *et al.* (2004) for a description of the algorithm, and mention that for the normal distribution, the optimal grids and the weights of the Voronoi tesselations are precomputed and available on the website `http://www.quantize.maths-fi.com`

- Recalling the dynamics (4.1), the conditional expectation operator is equal to

$$P^{\hat{a}_n^M(z)} W(x) \;=\; \mathbb{E}\big[W(Z_{n+1}^{\hat{a}_n^M}) | Z_n = x\big] = \mathbb{E}\big[W(F(z, \hat{a}_n^M(z), \varepsilon_1))\big], \quad z \in,$$

that we shall approximate analytically by quantization via:

$$\widehat{P}^{\hat{a}_n^M(z)} W(z) := \mathbb{E}\big[W(F(z, \hat{a}_n^M(z), \hat{\varepsilon}))\big] \;=\; \sum_{\ell=1}^{K} \hat{p}_\ell W(F(z, \hat{a}_n^M(z), e_\ell)).$$

Let us secondly introduce the notion of training distribution that will be used to build the estimators of value functions at time n, for $n = 0, \ldots, N-1$. Let us consider a measure μ on the state space E. We refer to it in the sequel as the training measure. Let us take a large integer M, and for $n = 0, \ldots, N$, introduce $\Gamma_n = \left\{Z_1^{(1)}, \ldots, Z_n^{(M)}\right\}$, where $\left(Z_n^{(m)}\right)_{m=1}^{M}$ is a i.i.d. sequence of r.v. following law μ. Γ_n should be seen as a training sampling to estimate the value function V_n at time n.

The proposed algorithm reads as:

$$\left\{\begin{array}{rll} \hat{V}_N^Q(z) = & g(z), & \text{for } z \in \Gamma_N, \\ \hat{Q}_n(z,a) = & \sum_{\ell=1}^K p_\ell \left[f(z,a) + \hat{V}_{n+1}^Q\left(\mathrm{Proj}_{n+1}\left(F(z, e_\ell, a)\right)\right)\right], & \\ \hat{V}_n^Q(z) = & \sup\limits_{a \in A} \hat{Q}_n(z,a), & \text{for } z \in \Gamma_n,\ n = 0, \ldots, N-1. \end{array}\right. \tag{4.4}$$

where, for $n = 0, \ldots, N$, $\mathrm{Proj}_n(z)$ stands for the closest neighbor of $z \in E$ in the grid Γ_n, i.e. the operator $z \mapsto \mathrm{Proj}_n(z)$ is actually the euclidean projection on the grid Γ_n.

In the sequel, we refer to (4.4) as the Qknn algorithm.

We shall make the following assumption on the transition probability of $(Z_n)_{0 \le n \le N}$, to guarantee the convergence of the Qknn algorithm.

(Htrans) Assume that the transition probability $\mathbb{P}(Z_{n+1} \in A \big| Z_n = z, a)$ conditioned by $Z_n = z$ when control a is followed at time n admits a density r w.r.t. the training measure μ, which is uniformly bounded and lipschitz w.r.t. the state variable z, i.e. there exists $\|r\|_\infty > 0$ such that for all $z \in E$ and control u taken at time n:

$$|r(y; n, x, a)| \le \|r\|_\infty \quad \text{and} \quad |r(y; n, x, a) - r(y; n, x', a)| \le [r]_L |x - x'|$$

and r is defined as follows:

$$\mathbb{P}(Z_{n+1} \in O \big| Z_n = z, u) = \int_O r(y; n, x, a) d\mu(y).$$

and where we denoted by $[r]_L$ the Lipschitz constant of r w.r.t. x.

Denote by $\mathrm{Supp}(\mu)$ the support of μ. We shall assume smoothness conditions on μ and F to provide a bound on the projection error.

(Hμ) We assume $\mathrm{Supp}(\mu)$ to be bounded, and denote by $\|\mu\|_\infty$ the smallest real such that $\mathrm{Supp}(\mu) \subset B\left(0, \|\mu\|_\infty\right)$. Moreover, we assume $x \in E \mapsto \mu\big(B(x,\eta)\big)$ to be Lipschitz, uniformly w.r.t. η, and we denote by $[\mu]_L$ its Lipschitz constant.

(HF) For $x \in E$ and $a \in A$, assume F to be $\mathbb{L}^1$-Lipschitz w.r.t. the noise component ε, i.e., there exists $[F]_L > 0$ such that for all $x \in E$ and $a \in A$, for all r.v. ε and ε', we have:

$$\mathbb{E}\left[\left|F(x,a,\varepsilon) - F(x,a,\varepsilon')\right|\right] \leq [F]_L \mathbb{E}\left[\left|\varepsilon - \varepsilon'\right|\right]$$

We now state the main result of this section whose proof is postponed in Appendix A.

THEOREM 4.1 *Take $K = M^{2+d}$ points for the optimal quantization of the exogenous noise ε_n, $n = 1, \ldots, N$. There exist constants $[\hat{V}_n^Q]_L > 0$, that only depends on the Lipschitz coefficients of f, g and F, such that, under **(Htrans)**, it holds for $n = 0, ..., N-1$, as $M \to +\infty$:*

$$\|\hat{V}_n^Q(X_n) - V_n(X_n)\|_2 \;\leq\; \sum_{k=n+1}^{N} \|r\|_\infty^{N-k} \left[\hat{V}_k^Q\right]_L \left(\varepsilon_k^{proj} + [F]_L \varepsilon_k^Q\right) + \mathcal{O}\left(\frac{1}{M^{1/d}}\right), \tag{4.5}$$

where $\varepsilon_k^Q := \|\hat{\varepsilon}_k - \varepsilon_k\|_2$ stands for the quantization error, and

$$\varepsilon_n^{proj} := \sup_{a \in A} \|\mathrm{Proj}_{n+1}\left(F(X_n, a, \hat{\varepsilon}_n)\right) - F(X_n, a, \hat{\varepsilon}_n)\|_2$$

stands for the projection error, when decision a is taken at time n.

REMARK 4.1 *The constants $[\hat{V}_n^Q]_L > 0$ are defined in (A8).*

From Theorem 4.1, we can deduce consistency and provide a rate of convergence for the estimator $\hat{V}_n^Q$, $n = 0, \ldots, N-1$, under some rather tough yet usual compactness conditions on the state space.

COROLLARY 4.1 *Under **(Hμ)** and **(HF)**, the Qknn-estimator $\hat{V}_n^Q$ is consistent for $n = 0, \ldots, N-1$, when taking M^{d+1} points for the quantization; and moreover, we have for $n = 0, ..., N-1$, as $M \to +\infty$:*

$$\|\hat{V}_n^Q(X_n) - V_n(X_n)\|_2 \;=\; \mathcal{O}\left(\frac{1}{M^{1/d}}\right).$$

Proof. We postpone the proof of Theorem 4.1 to Appendix A. □

4.3. *Qknn agorithm applied to the order book control problem (3.1)*

We recall the expression of the time-continuous controlled order book process

$$Z_t = \big(X_t, Y_t, \underline{a}_t, \underline{b}_t, \underline{na}_t, \underline{nb}_t, pa_t, pb_t, \underline{ra}_t, \underline{rb}_t\big)$$

admits a representation as a MDP as shown in Section 3.2. In Section 3.3, we proved that the value function associated to the MDP is characterized as limit as $N \to \infty$ of the the Bellman equation (3.14). In this section, some implementation details on the Qknn algorithm are presented in order to numerically solve (3.14).

Training set design

Inspired by Fiorin *et al.* (2018), we use product-quantization method and randomization techniques to build the training set Γ_n on which we project (T_n, Z_n) that lies on $[0,T] \times E$, where T_n and Z_n stands for the n^{th} jump of Z and the state of Z at time t_n, i.e. $Z_n = Z_{T_n}$, for $n \geq 0$. This basic idea of Control Randomization consists in replacing in the dynamics of Z the endogenous control by an exogenous control $(I_{T_n})_{n\geq 0}$, as introduced in Kharroubi *et al.* (2014). In order to alleviate the notations, we denote by I_n the control taken at time T_n, for $n \geq 0$.

Initialization. Set: $\Gamma_0^E = \{z\}$ and $\Gamma_0^T = \{0\}$.

Randomize the control, using e.g. uniform distribution on A at each time step, and then simulate D randomized processes to generate $(T_n^k, Z_n^k)_{n=0,k=1}^{N,D}$.

For all $n = 1, \dots, N$, set $\Gamma_n^T = \{T_n^k, 1 \leq k \leq D\}$, which stands for the grid associated to the quantization of the n^{th} jump time T_n, and set $\Gamma_n^E = \{Z_n^k, 1 \leq k \leq D\}$ which stands for the grid associated to the quantization of the state Z_n of Z at time T_n.

Remark 4.2 *The way we chose our training sets is often referred to as an* exploration strategy *in the reinforcement learning literature. Of course, if one has ideas or good guess of where to optimally drive the controlled process, she should not follow an exploration-type strategy to build the training set, but should rather use the guess to build it, which is referred to as the* exploitation strategy *in the reinforcement learning and the stochastic bandits literature. We refer to Balata et al. (2019) for several other applications of the* exploration strategy *to build training sets. Note that this idea is the root of all the Q-learning based algorithms. See Sutton and Barto (2018) for more details on Q-learning.*

Let F and G be the Borelian functions such that $Z_n = F\big(Z_{n-1}, d_n, I_n\big)$ and $T_n = G\big(T_{n-1}, \epsilon_n, I_n\big)$, where $\epsilon_n \sim \mathcal{E}(1)$ stands for the temporal noise, and d_n is the state noise, for $n \geq 0$.

Let us fix $N \geq 1$ and consider $\big(\widehat{T}_n, \widehat{Z}_n\big)_{n=0}^N$, the dimension-wise projection of $\big(T_n, Z_n\big)_{n=0}^N$ on the grids $\Gamma_n^T \times \Gamma_n^E$, $n = 0, \dots, N$, i.e. $\widehat{T}_0 = 0$, $\widehat{Z}_0 = z$, and

$$\begin{cases} \widehat{T}_n = \text{Proj}\Big(G\big(\widehat{T}_{n-1}, \epsilon_n, I_n\big), \Gamma_n^T\Big), \\ \widehat{Z}_n = \text{Proj}\Big(F\big(\widehat{Z}_{n-1}, d_n, I_n\big), \Gamma_n^E\Big), \end{cases} \qquad \text{for } n = 1, \dots, N.$$

$\big(\widehat{T}_n, \widehat{Z}_n, I_n\big)_{n\in\{0,N\}}$ is a Markov chain.

Define then $\big(\widehat{T}_n^Q, \widehat{Z}_n^Q\big)_{n=0}^N$ as temporal noise-quantized version of $\big(\widehat{T}_n, \widehat{Z}_n, I_n\big)_{n=0}^N$. Note that we do not need to quantize the spacial noise since this noise already takes a finite number of states. Let $\hat{\varepsilon}_n$ be the quantized process associated to ϵ_n. The process $\big(\widehat{T}_n^Q, \widehat{Z}_n^Q\big)_{n=0}^N$ is then defined as follows:

$\widehat{Z}_0^Q = z$, $\widehat{T}_0^Q = 0$ and $\forall 1 \leq n \leq N$:

$$\begin{cases} \widehat{T}_n^Q = \mathrm{Proj}\Big(G\big(\hat{t}_{n-1}, \hat{\varepsilon}_n, I_n\big), \Gamma_n^T\Big), \\ \widehat{Z}_n^Q = \mathrm{Proj}\Big(F\big(\widehat{Z}_{n-1}, d_n, I_n\big), \Gamma_n^E\Big). \end{cases}$$

Denote by $\left(\widehat{V}_n^{Q,(N,D)}\right)_{n=0}^N$ the solution of the Bellman equation associated to $\left(\widehat{T}_n^Q, \widehat{Z}_n^Q\right)_{n=0}^N$:

$$(\widehat{B}_{N,D}^Q): \begin{cases} \widehat{V}_N^{Q,(N,D)} = 0 \\ \widehat{V}_n^{Q,(N,D)}(t,z) = r(t,z,a) + \sup\limits_{a \in A} \left\{ \mathbb{E}_{t,z}^a \left[\widehat{V}_{n+1}^{Q,(N,D)}\big(\widehat{T}_{n+1}^Q, \widehat{Z}_{n+1}^Q\big) \right] \right\}, \text{ for } n = 0, \ldots, N, \end{cases}$$

where $\mathbb{E}_{t,z}^a[.]$ stands for the expectation conditioned by the events $\widehat{T}_n^Q = t$,$\widehat{Z}_n^Q = z$ and when decision $I_n = a$ is taken at time t.

We wrote the pseudo-code of the Qknn algorithm to compute $(\widehat{B}_{N,D}^Q)$ in Algorithm 1.

Algorithm 1 Generic Qknn Algorithm

Inputs:

– N: number of time steps

– z: state in E at time $T_0 = 0$

– $\Gamma^\varepsilon = \{e_1, \ldots, e_L\}$ and $(p_\ell)_{\ell=1}^L$: the grid and the weights for the optimal quantization of $(\varepsilon_n)_{n=1}^N$.

– Γ_n and Γ_n^E the grids for the projection of respectively the time and the state components at time n, for $n = 0, \ldots, N$.

1: **for** $n = N-1, \ldots, 0$ **do**

2: Compute the approximated $Qknn$-value at time n:

$$\begin{aligned} \hat{Q}_n(z,a) = r\left(T_n, z, a\right) \\ + \sum_{\ell=1}^L p_\ell \widehat{V}_{n+1}^Q \left(\mathrm{Proj}\left(G(z, e_\ell, a), \Gamma_{n+1}^T\right), \mathrm{Proj}\left(F(z, e_\ell, a), \Gamma_{n+1}^E\right)\right), \\ \text{for } (z,a) \in \Gamma_n \times A_z; \end{aligned}$$

3: Compute the optimal control at time n

$$\hat{A}_n(z) \in \underset{a \in A_z}{\mathrm{argmin}}\ \hat{Q}_n(z,a), \quad \text{for } z \in \Gamma_n,$$

where the argmin is easy to compute since A_z is finite for all $z \in E$;

4: Estimate analytically by quantization the value function:

$$\widehat{V}_n^Q(z) = \hat{Q}_n\left(z, \hat{A}_n(z)\right), \qquad \forall z \in \Gamma_n;$$

5: **end for**

Output:

– $(\widehat{V}_0^Q)$: Estimate of $V(0,z)$;

We discuss in Remark 4.3 the reasons why we can apply Theorem 4.1.

REMARK 4.3 *When the number of jumps of the LOB $N \geq 1$ is fixed, the set of all the states that can take the controlled order book by jumping less than N times, denoted by $\mathcal{K}$ in the sequel, is finite. Hence, the reward function r, defined in (3.6), is bounded and Lipschitz on $\mathcal{K}$.*

The following proposition states that $\widehat{V}_n^{Q,(N,D)}$, built from the combination of time-discretization, k-nearest neighbors and optimal quantization methods, is a consistent estimator of the value function at time T_n, for $n = 0, \ldots, N-1$. It provides a rate of convergence for the Qknn-estimations of the value functions.

PROPOSITION 4.1 *The estimators of the value functions provided by Qknn algorithm are consistent. Moreover, it holds as $M \to +\infty$:*

$$\left\| \widehat{V}_n^{Q,(N,M)}\left(\widehat{T}_n, \widehat{Z}_n\right) - V_n\left(T_n, Z_n\right) \right\|_{M,2} = \mathcal{O}\left(\alpha^N + \frac{1}{M^{2/d}}\right), \quad \text{for } n = 0, \ldots, N-1,$$

where we denote by $\|.\|_{M,2}$ the $\mathbb{L}^2(\mu)$ norm conditioned by the training sets that have been used to build the estimator $\widehat{V}_{n+1}^{Q,(N,M)}$.

Proof. Splitting the error of time cutting and quantization, we get:

$$\begin{aligned} \|V_n\left(T_n, Z_n\right) - \widehat{V}_n^{(N,M)}\left(\widehat{T}_n, \widehat{Z}_n\right)\|_{M,2} \leq \|V_n\left(T_n, Z_n\right) - V_n^{(N)}\left(T_n, Z_n\right)\|_{M,2} \\ + \|V_n^{(N)}\left(T_n, Z_n\right) - \widehat{V}_n^{(N,M)}\left(\widehat{T}_n, \widehat{Z}_n\right)\|_{M,2}. \end{aligned} \tag{4.8}$$

Step 1: Applying Lemma 3.2, we get the following bound on the first term in the r.h.s. of (4.8):

$$\|V_n\left(T_n, Z_n\right) - V_n^{(N)}\left(T_n, Z_n\right)\|_{M,2} \leq \frac{\alpha^N}{1-\alpha} \|b\|_\infty, \tag{4.9}$$

where $\|b\|_\infty$ stands for the supremum of b over $[0,T] \times E$.

Step 2: Note that the assumptions of Theorem 4.1 are met as noticed in Remark 4.3, so that the latter provides the following bound for the second term in the r.h.s. of (4.8):

$$\left\| V_n^{(N)}\left(T_n, Z_n\right) - \widehat{V}_n^{Q,(N,M)}\left(\widehat{T}_n, \widehat{Z}_n\right) \right\|_{M,2} \underset{M\to\infty}{=} \mathcal{O}\left(\frac{1}{M^{2/d}}\right). \tag{4.10}$$

It remains to plug (4.9) and (4.10) into (4.8) to complete the proof of Proposition 4.1. □

4.4. *Numerical results*

In this section, we propose several settings to test the efficiency of Qknn on simulated order books. We take no running reward, i.e. $f = 0$, and take the wealth of the market maker after liquidating their inventory as terminal reward, i.e. $g(z) = x + L(y)$. The intensities are taken constant in the first tests, and state dependent in the second ones. The values of the state dependent intensities are similar to the ones in Huang *et al.* (2015) (see). Although the intensities are assumed to be uncontrolled in section 3 for predictability reasons, the latter are controlled processes in this section, i.e. the intensities of the order arrivals depends on the orders in the order book from all the participant plus the ones of the market maker. The optimal trading strategies have been computed

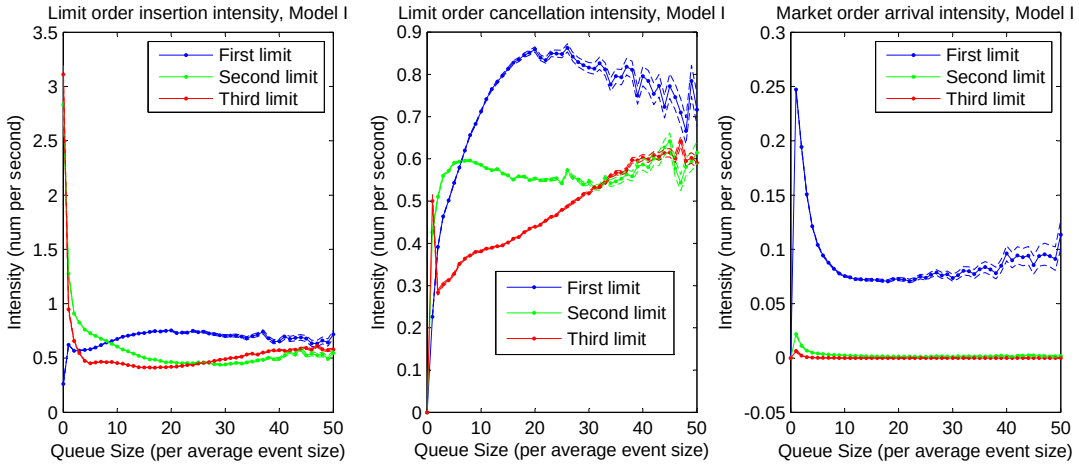


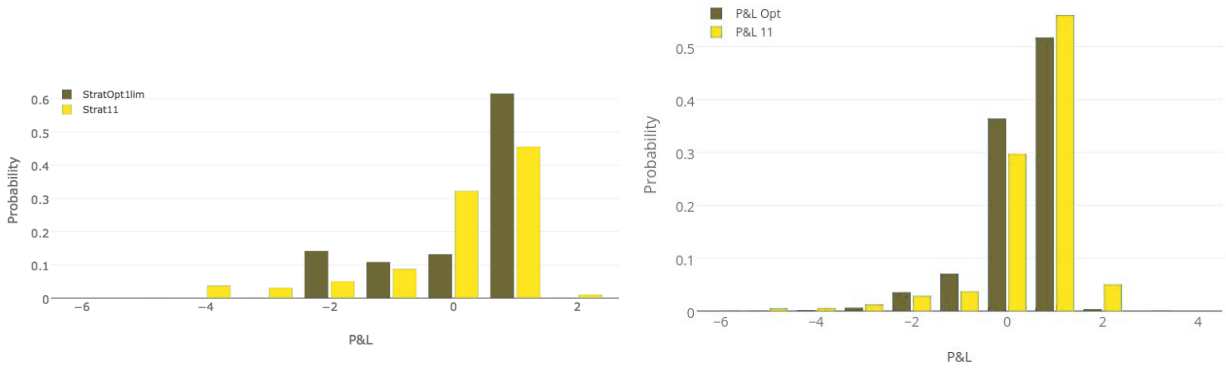
Figure 2.: Values of the intensities w.r.t. the queue size used in the numerical tests of Section 4.4. The plot is taken from Huang *et al.* (2015) (see its Figure 13). The intensities are estimated by the authors using their data from Alcatel Lucent.

among two different classes of strategies: in Section 4.4.1, we tested the algorithm to approximate the optimal strategy among those where the market maker is only allowed to place orders only at the best bid and the best ask; in Section 4.4.2, we computed the optimal trading strategy among the class of the strategies where the market maker allows herself to place orders on the two best limits on each side of the order book. Note that the second class of controls is more general than the first one. The code is available on `https://github.com/comeh`.

The search of the k nearest neighbors, that arises when estimating the conditional expectations using the Qknn algorithm, is very time-consuming; especially in the considered market-making problem which is of dimension more than 10. The efficiency of Qknn then highly depends on the algorithm used to find the k nearest neighbors in high-dimension. We implemented the method using the Fast Library for Approximate Nearest Neighbors algorithm (FLANN), introduced in Muja and Lowe (2009) and already available as a library of C++, Python, Julia and many other languages. This algorithm is based on tree methods. Note that recent algorithms based on graph also proved to perform well and can also be used.

4.4.1. **Case 1: The market maker only place orders at the best ask and best bid.** Denote by A1lim the class of controls where the placements of orders are allowed on the best ask and best bid exclusively. We implement the Qknn algorithm to compute the optimal strategy among those in A1lim. We then compared the optimal strategy with a naive strategy which consists in always placing one order at the best bid and one order at the best ask. The naive strategy is called 11 in the plots, and can be seen as a benchmark. The naive strategy is a good benchmark when the model for the intensities of order arrivals is symmetrical, i.e. the intensities for the bid and the ask sides are the same. Indeed, in this case, the market maker can expect to earn the spread in average.

In Figure 3a, we take constant intensities to model the limit and market orders arrivals, and linear intensity to model the cancel orders. In this setting, as we can see in the figure, the strategy computed using Qknn algorithm performs as well as the naive strategy. Note that, obviously, the market maker has to take enough points for the state quantization in order for Qknn algorithm to perform well. In Figure 3b, we plotted the P&L of the market maker when the latter compute the optimal strategy using only 6000 points for the state space discretization, and for such a low number of points for the grid, Qknn algorithm performs poorly. In this setting, notice that the



(a) 10^5 points for the grids.

(b) 6000 points for the grids.

Figure 3.: Histogram of the P&L when following the Qknn estimated optimal strategy (Opt) and the naive strategy (11). We took symmetrical and constant intensities, and a short terminal time $T = 1$. Notice that the Qknn strategy looks to improve the P&L by reducing the losses when enough points are taken to build the grids (see Figure 3a), and that its performance is worse if less points are taken to build the grids (see Figure 3b).

naive strategy seems to perform well.

In Figure 4, we plotted the empirical histogram of the P&L of the market maker using the Qknn-estimated optimal strategy, computed with grids of size $N = 10^3, 10^4, 10^5, 10^6$ for the state space discretization; and the empirical histogram of the P&L of the market maker using the naive strategy. We took intensities that are state dependent. One can see that the larger the size of the grids are, the better the Qknn-estimation of the optimal strategy is.

In Figure 5, we plot the P&L of the market maker following the Qknn-estimated optimal strategy and the naive strategy. We took the same parameters as in Figure 4 to run the simulations except from the terminal time that we set to be equal to $T = 10$. In this setting, the Qknn-estimated optimal strategy performs much better than the naive strategy, which highlights the fact that the naive strategy is not optimal.

We plotted in Figure 6 the reaction of Qknn when a trend is added in the dynamic of the market. In this example, we took a higher intensity for the sell market order than the one for the buy market order, which creates an artificial positive trend in the dynamic of the price. Observe that Qknn understood correctly that it is better not to sell when the price goes up.

4.4.2. **Case 2: the market maker place orders on the first two limits of the Orders Book.** We extend the class of admissible controls to the ones where the market maker places order on the first two limits on the bid and ask sides of the order book. Denote by $A2lim$ the latter. We run simulations to test the Qknn algorithm on A2lim. In figure 7 and figure 8, we plot the empirical distributions of the P&L when the market maker follows the three different strategies:

- Qknn-estimated optimal strategy among those in A2lim (PLOpt2lim).
- Qknn-estimated optimal strategy among those in A1lim (PLOpt1lim).
- naive strategy, i.e. always place orders on the best bid and best ask queues (PL11).

Note that the P&L of the market maker is always better when the class of admissible controls is extended, see Figure 7; but in some numerical tests, the extended set of controls does not seem to improve the P&L. Indeed, we observed that the terminal P&L estimated using Qknn among A2lim and A1lim have the similar empirical distribution in the tests whose results are presented in Figure 8.

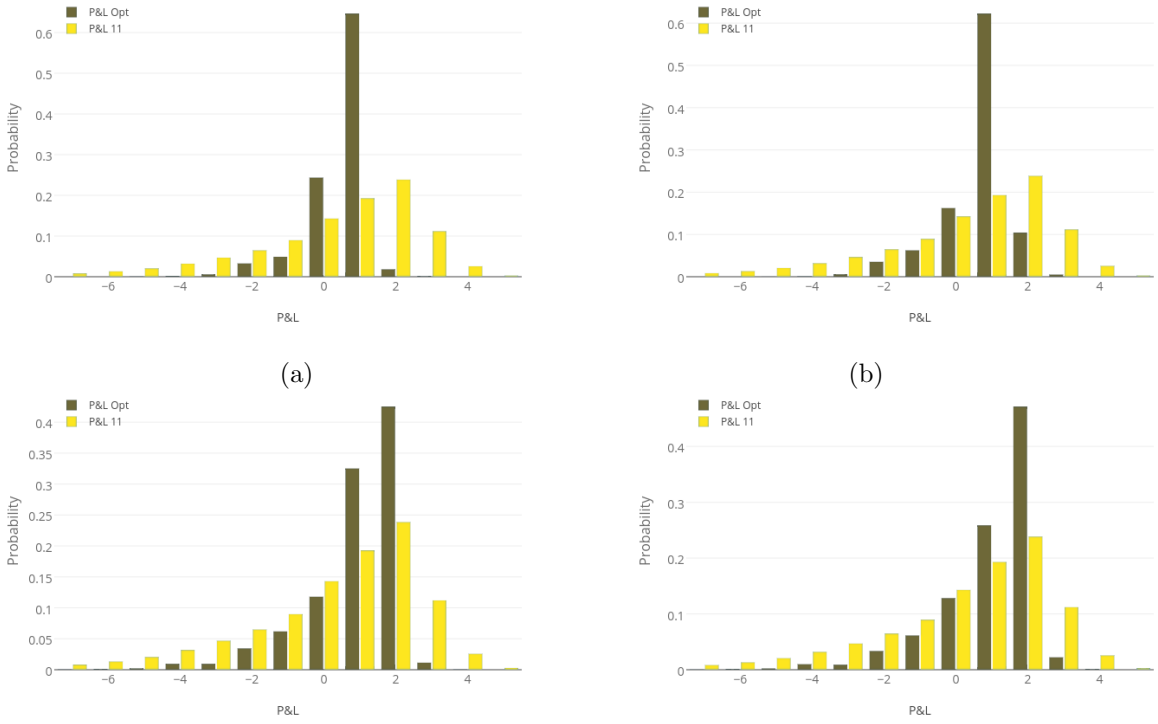


Figure 4.: Histogram of the P&L when the market maker follow the Qknn estimated optimal strategy (Opt) and the naive strategy (11). The intensities λ^M , λ_i^L and λ_i^C depend on the state of the order book. The P&L of the market maker when following the Qknn-estimated optimal strategy is computed using 10^3, 10^4, 10^5, and 10^6 points for the grids: see respectively Figures 4a, 4b, 4c and 4d. The reader can see that the market maker increases their expected terminal wealth (after liquidation) by taking more and more points for the state space discretization. Also, the naive strategy is beaten when the intensities are state dependent.

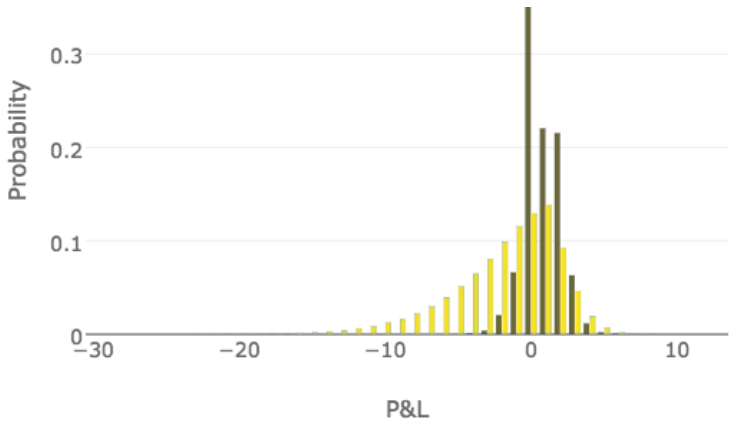


Figure 5.: Distribution of the P&L when the market maker follows the optimal strategy (dark blue) and the naive strategy (light blue). We took symmetrical and state dependent intensities; and long terminal time: $T = 10$. Notice that the Qknn strategy does better than the naive strategy when the intensities are state dependent.

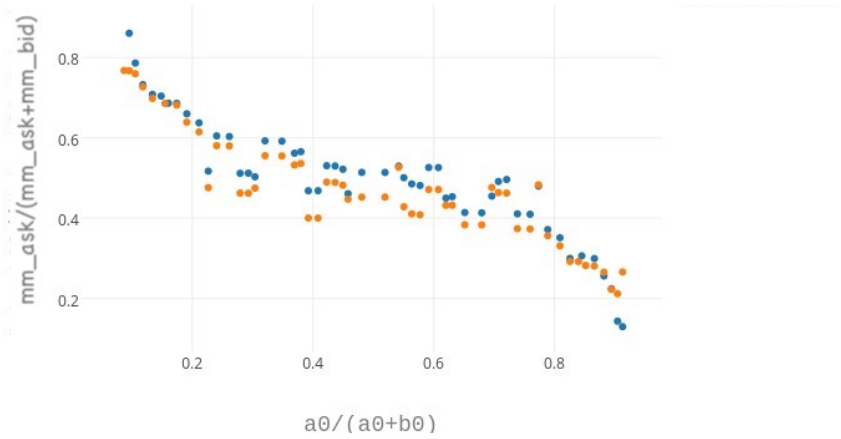


Figure 6.: Reaction of Qknn to an artificial positive trend in the LOB. The y-axis represents the ratio of orders sent on the ask side by the market maker to their orders sent in the ask and bid sides in the order book. The x-axis is represents the ratio of the size of the best-ask limit to the sum of the sizes of the best-bid and best-ask limits. The blue points are those without trend in the market. The orange points are those with a positive trend in the market. We can see that Qknn took correctly the trend into account in its decisions: for two same order books, Qknn is less willing to sell when the price is expected to increase.

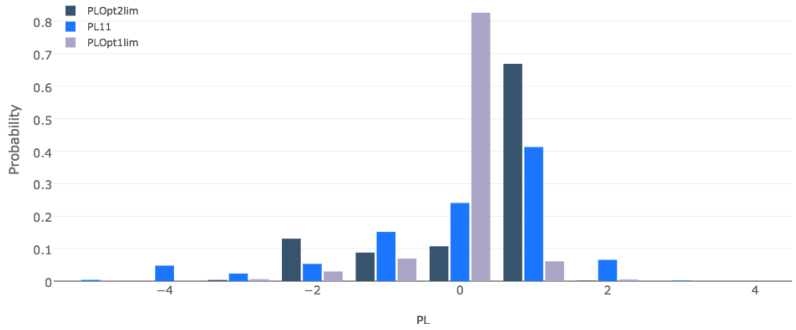


Figure 7.: P&L of the market maker who follows optimal strategies and the naive strategy (PL11). Short Terminal Time. asymmetrical intensities for the market order arrivals: the intensity for the buying market order process is taken higher than the one for the selling market order process. The wealth of the market maker is greater when she places orders on the two first limits of each sides of the order book, rather than when she places orders only on the best limits at the bid and ask sides.

5. Model extension to Hawkes Processes

We consider in this section a market maker who aims at maximizing a function of their terminal wealth, penalizing their inventory at terminal time T in the case where the orders arrivals are driven by Hawkes processes. Let us first present the model with Hawkes processes for the LOB.

Model for the LOB:

We assume that the order book receives limit, cancel, and market orders. We denote by L^+ (resp. L^-) the limit order arrivals process the ask (resp. bid) side; by C^+ (resp. C^-) the cancel order on the ask (resp. bid) side; and by M^+ (resp. M^-) the buy (resp. sell) market order arrivals processes. In this section, the limit orders arrivals are assumed to follow Hawkes processes dynamics, and moreover we assume the kernel to be exponential. The order arrivals are then modeled by a (4K+2)-variate Hawkes process (N_t) with a vector of exogenous intensities λ_0 and exponential kernel ϕ, i.e. $\phi^{ij}(t) = \alpha_{ij}\beta e^{\beta t}\mathbf{1}_{t\geq 0}$.

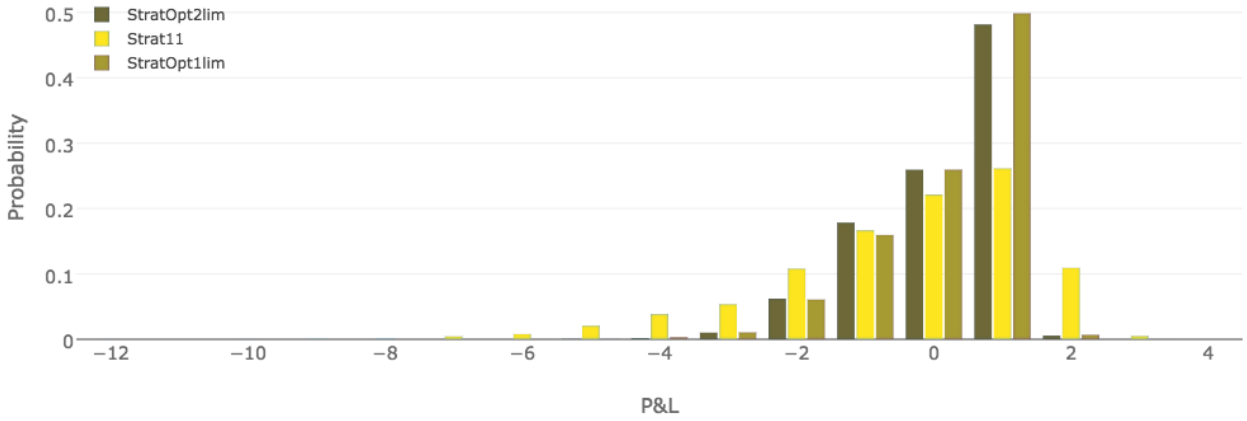


Figure 8.: P&L when following the optimal strategy or the naive strategy (PL11). Long Terminal Time. Symmetrical intensities for the arrival of market orders. 4.10^4 points for the quantization. Notice that the Qknn strategy computed on the extended class of controls, i.e. order placements on the two first limits (StratOpt2lim), performs as well as the one computed on the original class of controls, i.e. order placements on the best-bid and best-ask (StratOpt1lim).

Let $\alpha=(\alpha^{ij})_{i,j}$. We assume: $(\alpha)_{i,j}$ to have spectral radius strictly smaller than 1, which is a sufficient condition to guarantee stationarity of the model and convergence of $\mathbb{E}[\lambda_u|\mathcal{F}_t]$ as $u \to +\infty$, as shown e.g. in Hawkes (1971).

Note that in the presented model, the following holds:

(Hλ) λ is assumed to be independent of the control.

Denoting by $D = 4K+2$ the dimension of (N_t), the m^{th} component of the intensity λ of N_t writes, under **(Hλ)**:

$$\lambda_t^m = \lambda_0^m + \sum_{j=1}^{D} \alpha_{mj} \int_0^t e^{-\beta(t-s)}\, \mathrm{d}N_s^j, \qquad \text{for } m = 1, \dots, D,$$

It is well-known that for this choice of intensity, the couple $(N_t, \lambda_t)_{t\geq 0}$ becomes Markovian, see e.g. Lemma 6 in Massoulié (1998) for a proof of this result, and moreover we have:

$$\mathrm{d}\lambda_t^m = -\beta D\big(\lambda_t^m - \lambda_0^m\big)dt + \sum_{j=1}^{D} \alpha_{mj}\, \mathrm{d}N_t^j, \qquad \text{for } m = 1, \dots, D,$$

with given initial conditions: $\lambda_0^m \in \mathbb{R}_+^*$ for $m = 1, \dots, D$.

We can now rewrite the control problem (3.1) in the particular case where the order book is driven by Hawkes processes, there is no running reward, i.e. $f = 0$, and where the terminal reward G stands for the terminal wealth of the market maker penalized by their inventory. We then consider the following problem in this section:

$$V(t, \lambda, z) := \sup_{\alpha \in \mathbb{A}} \mathbb{E}^{\alpha}_{t,z,\lambda}\big[G\big(Z_T\big)\big], \tag{5.1}$$

where $G(z)$ denotes the wealth of the market maker when the controlled order book is at state z, plus a term of penalization of their inventory; and where $\mathbb{A}$ is the set of the admissible controls, i.e. the predictable decisions taken by the market maker until a terminal time $T > 0$.

We now present the main result of this section.

THEOREM 5.1 *V is characterized as the unique solution of the following HJB equation:*

$$\begin{cases} f(T,z,\lambda) = G(z), \ \textit{for } z \in E \\ \qquad 0 = \dfrac{\partial f}{\partial t}(t,z,\lambda) - D\beta \displaystyle\sum_{m=1}^{D} \Big[(\lambda^m - \lambda_0^m) \dfrac{\partial f}{\partial \lambda^m}(t,z,\lambda) \\ \qquad\qquad + \lambda^m \sup_{a \in A_z} \left[f\big(t, e_m^a(z), \lambda + \alpha_m\big) - f\big(t,z,\lambda\big) \right] \Big], \\ \qquad\qquad\qquad \textit{for } 0 \le t < T, \quad \textit{and} \quad (t,z,\lambda) \in \mathbb{R}_+ \times E \times \mathbb{R}_+^* \,. \end{cases} \tag{5.2}$$

where $\alpha_m = (\alpha_{1m}, \dots, \alpha_{Dm})$. Moreover, V admits the following representation:

$$V(t,z,\lambda) = \sup_{\alpha \in \mathbb{A}} \sum_{n=0}^{\infty} \mathbb{E}_{t,z,\lambda}^{\alpha} \Bigg[1_{T_n \le T} G\left(Z_{T_n}^{\alpha}\right) \exp\Bigg\{ -|\lambda_0|(T - T_n) \\ + \sum_{m=1}^{D} \frac{\lambda_{T_n}^m - \lambda_0^m}{D\beta} \left(e^{-\sum_{j=1}^{D} \beta_{mj}(T - T_n)} - 1 \right) \Bigg\} \Bigg], \tag{5.3}$$

where, for $n \ge 0$, T_n stands for the n^{th} jump time of Z after time t, and $(Z_{T_n}^{\alpha}, \lambda_{T_n})_{n=0}^{\infty}$ is as a MDP controlled by $\alpha \in \mathbb{A}$; and where $\mathbb{E}_{t,z,\lambda}^{\alpha}[.]$ stands for the expectation conditioned by $Z_t = z, \lambda_t = \lambda$ when the control α is followed.

REMARK 5.1 *V is characterized in (5.3) as the value function associated to an MDP with infinite horizon, where the running reward reads:*

$$r(t,z,\lambda) \ = \ 1\!\!1_{t \le T} \, G\left(z\right) \exp\left\{ -|\lambda_0|_1 (T-t) + \sum_{m=1}^{D} \frac{\lambda^m - \lambda_0^m}{D\beta} \left(e^{-D\beta(T-t)} - 1 \right) \right\},$$

where $|.|_1$ denotes the $\mathbb{L}^1\big(\mathbb{R}^D\big)$ norm.

Proof: (of Theorem 5.1)

Step 1: Let us check that (5.3) holds, where V is defined as solution of (5.1).

First notice that $(\lambda_t, Z_t)_t$ is a PDMDP[1], i.e. $(\lambda_t, Z_t)_t$ is deterministic between two jumping times. We then aim at rewriting the expression of the value function defined in (5.1) as the value function associated to a infinite horizon control problem of the PDMDP $(\lambda_t, Z_t)_t$. To do so, we first notice

[1] PDMDP stands for Piecewise Deterministic Markov Decision Process, which is a MDP whose dynamic is deterministic between jumps.

that by conditioning on the time jumps we get:

$$\begin{aligned} V(t,z,\lambda) &= \sup_{\alpha\in\mathbb{A}} \mathbb{E}^{\alpha}_{t,z,\lambda}\Big[G\big(Z^{\alpha}_{T}\big)\Big] \\ &= \sup_{\alpha\in\mathbb{A}} \mathbb{E}^{\alpha}_{t,z,\lambda}\Big[\sum_{n=0}^{\infty} 1_{T_n\le T<T_{n+1}} G\big(Z^{\alpha}_{T_n}\big)\Big] \\ &= \sup_{\alpha\in\mathbb{A}} \sum_{n=0}^{\infty} \mathbb{E}^{\alpha}_{t,z,\lambda}\Big[1_{T_n\le T} G\big(Z^{\alpha}_{T_n}\big)\mathbb{P}\Big(T-T_n\le T_{n+1}-T_n\big|T_n\Big)\Big], \end{aligned} \tag{5.4}$$

where $\big(T_n\big)_n$ is the sequence of jump times of N. This process is a jump process with intensity $\mu_s = \sum_{m=1}^{D}\lambda_s^m$. Since it holds, conditioned on $\mathcal{F}_{T_n}$:

$$\mu_s = \sum_{m=1}^{D}\lambda_0^m + \big(\lambda^m_{T_n}-\lambda^m_0\big)e^{-D\beta(s-T_n)}, \qquad \text{for } s\in[T_n,T_{n+1}),$$

then, we have:

$$\begin{aligned} \mathbb{P}\Big(T_{n+1}-T_n\ge T-T_n\big|T_n\Big) &= \int_{T-T_n}^{\infty}\mu_s e^{-\int_0^s\mu_u du}\,\mathrm{d}s \\ &= \exp\Big\{-|\lambda_0|(T-T_n)+\sum_{m=1}^{D}\frac{\lambda^m_{T_n}-\lambda^m_0}{D\beta}\Big(e^{-D\beta(T-T_n)}-1\Big)\Big\}. \end{aligned} \tag{5.5}$$

Plugging (5.5) into (5.4), the value function rewrites:

$$\begin{aligned} V(t,z,\lambda) = \sup_{\alpha\in\mathbb{A}}\sum_{n=0}^{\infty}\mathbb{E}^{\alpha}_{t,z,\lambda}\Bigg[1_{T_n\le T}G\left(Z^{\alpha}_{T_n}\right)\exp\Bigg\{&-|\lambda_0|(T-T_n) \\ &+\sum_{m=1}^{D}\frac{\lambda^m_{T_n}-\lambda^m_0}{D\beta}\left(e^{-D\beta(T-T_n)}-1\right)\Bigg\}\Bigg], \end{aligned} \tag{5.6}$$

which completes the step 1. The r.h.s of (5.6) can be seen as the value function of an infinite horizon control problem associated to the PDMDP.

Step 2: Let us show that V is the unique solution to (5.2).

Notice first that the solutions to the following HJB equation

$$\begin{cases} G(z) = f(T,z,\lambda) \\ 0 = \frac{\partial f}{\partial t} - \sum_{m=1}^{D} D\beta\big(\lambda^m-\lambda_0^m\big)\frac{\partial f}{\partial\lambda^m} + \lambda^m \sup\limits_{a\in A_z}\big[f\big(t,e^a_m(z),\lambda+\alpha_m\big)-f\big(t,z,\lambda\big)\big], \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } 0\le t<T. \end{cases}$$

are the fixed points of the operator $\mathcal{T}=\mathcal{T}_1\circ\mathcal{T}_2$ where $\mathcal{T}_1$ and $\mathcal{T}_2$ are defined as follows:

$$\mathcal{T}_1 : F\mapsto f \text{ solution of } \begin{cases} \frac{\partial f}{\partial t} - D\beta\sum_{m=1}^{D}\big(\lambda^m-\lambda_0^m\big)\frac{\partial f}{\partial\lambda^m} = F(t,z,\lambda) \\ f(T,z,\lambda)=G(z), \end{cases}$$

and:

$$\mathcal{T}_2 : f \mapsto -\sum_{m=1}^{D} \lambda^m \sup_{a\in A_z} \big[f\big(t, e_m^a(z), \lambda+\alpha_m\big) - f\big(t,z,\lambda\big)\big].$$

We now use the characteristic method to rewrite the image of $\mathcal{T}_1$.
Let us take function F, and define $f = \mathcal{T}_1(F)$. Let us fix $t \in [0,T]$ and $\lambda \in (\mathbb{R}_+)^D$, and denote by g the function $g(s,z) = f(s,z,\lambda_s^1,...,\lambda_s^D)$ where, for $m = 1,\dots,D$, $s \mapsto \lambda_s^m$ is a differentiable function defined on $[t,T]$ as solution to the following ODE:

$$\begin{cases} \frac{d\lambda_s^m}{ds} = -D\beta\big(\lambda_s^m - \lambda_0^m\big), & \text{for all } t < s \le T, \\ \lambda_t^m = \lambda^m. \end{cases} \tag{5.7}$$

For $m = 1,\dots,D$, basic theory on ODE provides existence and uniqueness of a solution to (5.7), which is given by:

$$\lambda_s^m \;=\; \lambda_0^m + \big(\lambda^m - \lambda_0^m\big)e^{-D\beta(s-t)}, \qquad \text{for } s \in [t,T], \text{ and } m = 1,\dots,D.$$

Since $\frac{\partial g}{\partial s} = \frac{\partial f}{\partial s} + \sum_{m=1}^{D} \frac{d\lambda_s^m}{ds}\frac{\partial f}{\partial \lambda^m}$, then $g(t,z) = G(z) - \int_t^T F(s,z,\lambda_s)\,\mathrm{d}s$, which finally leads to the following expression of $\mathcal{T}_1(F)$:

$$\mathcal{T}_1(F) = f(t,z,\lambda) = G(z) - \int_t^T F\big(s,z,\lambda_s\big)\,\mathrm{d}s. \tag{5.8}$$

Replacing F by $\mathcal{T}_2(f)$ in (5.8), we get that f is fixed point of $\mathcal{T}_1 \circ \mathcal{T}_2$ if and only if:

$$f(t,\lambda,z) + \sum_{m=1}^{D}\int_t^T \lambda_s^m f\big(s,z,\lambda_s\big)\,\mathrm{d}s = G(z) - \sum_{m=1}^{D}\int_t^T \lambda_s^m \sup_{a\in A_z} f\big(s,e_m^a(z),\lambda_s+\alpha_m\big)\,\mathrm{d}s.$$

Notice

$$\frac{\partial f(s,\lambda_s,z)e^{-\sum_{j=1}^{D}\int_t^s \lambda_u^j du}}{\partial s} = -\sum_{m=1}^{D} \lambda_s^m e^{-\sum_{j=1}^{D}\int_t^s \lambda_u^j du} \sup_{a\in A_z} f\big(s,e_m^a(z),\lambda_s+\alpha_m\big),$$

so that:

$$\begin{aligned} f(t,\lambda,z) &= G(z)e^{-\sum_{m=1}^{D}\int_t^T \lambda_s^m\,\mathrm{d}s} + \sum_{m=1}^{D}\int_t^T \lambda_s^m e^{-\int_t^s \lambda_u du} \sup_{a\in A_z} f\big(s,e_m^a(z),\lambda_s+\alpha_m\big)\,\mathrm{d}s \\ &= G(z)e^{-\sum_{m=1}^{D}\int_t^T \lambda_s^m\,\mathrm{d}s} + \sup_{a\in A_z} \mathbb{E}^a_{t,\lambda,z}\left[f\big(T_1,Z_1,\lambda_{T_1}+\alpha_m\big)\right], \end{aligned} \tag{5.9}$$

where T_1 is the first jump time of N larger than t, we denote $Z_1 = Z_{T_1}$. Equation (5.9) shows that the fixed point of $\mathcal{T}_1 \circ \mathcal{T}_2$ is characterized as the fixed point of the operator $\mathcal{T}$ defined for any smooth enough function f by:

$$\mathcal{T}(f) = G(z)e^{-\sum_{m=1}^{D}\int_t^T \lambda_s^m\,\mathrm{d}s} + \sup_{a\in A_z} \mathbb{E}^a_{t,\lambda,z}\Big[f\big(T_1,Z_1,\lambda_{T_1}+\alpha_m\big)\Big],$$

where $\mathbb{E}^a_{t,\lambda,z}[.]$ stands for the expectation conditioned by the events $\lambda_t = \lambda$ and $Z_t = z$, when decision a is taken at time t. We recognize here the maximal reward operator of the value function

defined in (5.6). Basic theory on PDMDP shows that the maximal reward operator $\mathcal{T}$ admits V as unique fixed point, which completes step 2. □

Conclusion

In this paper, we solved theoretically and numerically a general market-making problem with different microstructural models of order books, by rewriting the problem as a Markov decision process with infinite horizon. This new representation offers a nice and simple characterization of the optimal strategy of market-making, which is implementable relying for example on some quantization and control randomization ideas as proposed in this paper. Others algorithms, like those based on reinforcement learning (see e.g. Chapter 6.5 of Sutton and Barto (2018) for an introduction to (deep) Q-learning, and/or Guéant and Manziuk (2019) for its application to market-making), look particularly well-adapted to solve the market-making problem using its MDP representation, especially in the context of high-dimension.
The proposed methodology can be adapted to solve theoretically and numerically control problems associated to a general class of controlled point processes.

Acknowledgment

We would like to thank the anonymous referees for their valuable comments on the first version of the paper.

References

Abergel, F., Anane., A., Chakraborti, A., Jedidi, A. and Muni Toke, I., *Limit Order Books*, 2016, Cambridge University Press.

Avellaneda, M. and Stoikov, S., High-frequency trading in a limit order book. *Quantitative Finance*, 2007, **8**, 217–224.

Balata, A., Huré, C., Laurière, M., Pham, H. and Pimentel, I., A class of finite-dimensional numerically solvable McKean-Vlasov control problems. *ESAIM Proceedings and surveys*, 2019, **65**.

Balata, A. and Palczewski, J., Regress-Later Monte Carlo for optimal control of Markov processes. *eprint arXiv:1712.09705*, 2017.

Baradel, N., Bouchard, B., Evangelista, D. and Mounjid, O., Optimal inventory management and order book modeling. *arXiv:1802.08135*, 2018.

Bäuerle, N. and Rieder, U., *Markov Decision Processes with Applications to Finance*, 2011, Springer.

Cartea, A. and Jaimungal, S., Modeling Asset Prices for Algorithmic and High Frequency Trading. *Applied Mathematical Finance*, 2010, **20(6)**, 512–547.

Cartea, A. and Jaimungal, S., Risk metrics and fine tuning of high-frequency trading strategies.. *Mathematical Finance*, 2013, **25(3)**, 576–611.

Cartea, A., Jaimungal, S. and Ricci, J., Buy Low Sell High: a High Frequency Trading Perspective. *SIAM Journal on Financial Mathematics*, 2014, pp. 415–444.

Cartea, A., Penalva, J. and Jaimungal, S., *Algorithmic and High-frequency trading*, 2015, Cambridge University Press.

Cont, R., Stoikov, S. and Talreja, R., A stochastic model for order book dynamics. *Operations Research*, 2007, **58**, 549–563.

El Aoud, S. and Abergel, F., A stochastic control approach for options market making.. *World scientific publishing company*, 2015, **1(1)**.

Fiorin, L., Pagès, G. and Sagna, A., Product Markovian quantization of a diffusion process with applications to finance. *Methodology and Computing in Applied Probability*, 2018, pp. 1–32.

Fodra, P. and Pham, H., High Frequency trading and asymptotics for small risk aversion in a Markov renewal model. *SIAM Journal of Financial Mathematics*, 2015a, **6(1)**, 656–684.
Fodra, P. and Pham, H., Semi Markov model for market microstructure. *Applied Mathematical Finance*, 2015b, **22(3)**, 261–295.
Graf, S. and Luschgy, H., *Foundations of quantization for probability distributions*, Vol. 1730, , 2000, Springer-Verlag Berlin Heidelberg.
Guéant, O., *The Financial Mathematics of Market Liquidity: From optimal execution to market making*, 2016, Chapman and Hall/CRC.
Guéant, O., Lahalle, C.A. and Fernandez-Tapia, J., Dealing with the Inventory Risk. *Mathematics and Financial Economics*, 2012, **7**, 477–507.
Guéant, O. and Manziuk, I., Deep reinforcement learning for market making in corporate bonds: beating the curse of dimensionality. *arXiv preprint arXiv:1910.13205*, 2019.
Guilbaud, F. and Pham, H., Optimal High Frequency Trading with limit and market orders. *Quantitative Finance*, 2013, **13(1)**, 79–94.
Gyorfi, L., Kohler, M., Krzyzak, A. and Walk, H., *A Distribution-Free Theory of Nonparametric Regression*, 2002, Springer.
Hawkes, A.G., Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 1971, **58**, 83–90.
Ho, T. and Stoll, H., Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 1979, **9**, 47–73.
Huang, W., Lehalle, C.A. and Rosenbaum, M., Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 2015, **110(509)**, 107–122.
Huré, C., Pham, H., Bachouch, A. and Langrené, N., Deep neural networks algorithms for stochastic control problems on finite horizon: convergence analysis. *arXiv:1812.04300*, 2018.
Jacquier, A. and Liu, H., Optimal Liquidation in a level-I Limit order book for large tick stocks. *SIFIN*, 2018, **9(1)**, 875–906.
Kharroubi, I., Langrené, N. and Pham, H., A numerical algorithm for fully nonlinear HJB equations: an approach by control randomization. *Monte Carlo Methods and Applications*, 2014, **20**, 145–165.
Lehalle, C.A., Othmane, M. and Rosenbaum, M., Optimal liquidity-based trading tactics. *arXiv:1803.05690*, 2018.
Massoulié, L., Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic Processes and their Applications*, 1998, **75(1)**, 1–30.
Muja, M. and Lowe, D., Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2009.
Pagès, G., Pham, H. and Printems, J., Optimal quantization methods and applications to numerical problems in finance. In *Handbook on Numerical Methods in Finance*, edited by S.T. Rachev and G.A. Anastassiou, chap. 7, pp. 253–298, 2004 (Birkhäuser: Boston).
Rosu, I., A dynamic model of the limit order book. *Review of Financial Studies*, 2008, **22**, 4601–4641.
Sutton, R.S. and Barto, A.G., *Reinforcement Learning: An Introduction*, Second , 2018, The MIT Press.

Appendix A: Proof of Theorem 4.1 and Corollary 4.1

We divided the proofs of Theorem 4.1 and Corollary 4.1 into the following Lemmas.

Lemma A.1 aims at bounding the projection error. It relies on Gyorfi *et al.* (2002), see p.93, as well as Zador's theorem, stated in Section B for the sake of completeness.

LEMMA A.1 *Assume $d \geq 3$, and take $K = M^{d+2}$ points for the optimal quantization of ε_n, then it holds under **(Hμ)** and **(HF)**, as $M \to +\infty$,*

$$\varepsilon_n^{proj} \quad = \quad \mathcal{O}\left(\frac{1}{M^{1/d}}\right), \tag{A1}$$

where we remind that $\varepsilon_n^{proj} := \sup_{a\in A} \|\mathrm{Proj}_{n+1}\left(F(X_n,a,\hat{\varepsilon}_n)\right) - F(X_n,a,\hat{\varepsilon}_n)\|_2$ *stands for the average projection error.*

Proof. Let us take $\eta > 0$, and observe that

$$\mathbb{P}\left(\left|\mathrm{Proj}_{n+1}\big[F(X_n,a,\hat{\varepsilon}_{n+1})\big] - F(X_n,a,\hat{\varepsilon}_{n+1})\right|^2 > \eta\right) = \mathbb{E}\left[\prod_{m=1}^{M}\mathbb{E}\left[\mathbf{1}_{|X_{n+1}^{t,(m)}-F(X_n,a,\hat{\varepsilon}_{n+1})|>\sqrt{\eta}}\middle| X_n,\hat{\varepsilon}_{n+1}\right]\right]$$

$$= \mathbb{E}\left[\Big(1-\mu\big[B\big(F(X_n,a,\hat{\varepsilon}_{n+1}),\sqrt{\eta}\big)\big]\Big)^M\right],$$

where for all $x\in E$ and $\eta>0$, $B(x,\eta)$ denote the ball of center x and radius η. Since $x\mapsto(1-x)^M$ is M-Lipschitz, we get by application of Zador's theorem:

$$\begin{aligned}
&\mathbb{P}\left(\left|\mathrm{Proj}_{n+1}\big[F(X_n,a,\hat{\varepsilon}_{n+1})\big] - F(X_n,a,\hat{\varepsilon}_{n+1})\right|^2 > \eta\right)\\
&\qquad\le M[F]_L[\mu]_L\,\|\hat{\varepsilon}_{n+1}-\varepsilon_{n+1}\|_2 + \mathbb{E}\left[\Big(1-\mu\big(B\big(F(X_n,a,\varepsilon_{n+1}),\sqrt{\eta}\big)\big)\Big)^M\right]\\
&\qquad= \frac{M[F]_L[\mu]_L}{K^{1/d}} + \mathbb{E}\left[\Big(1-\mu\big(B\big(F(X_n,a,\varepsilon_{n+1}),\sqrt{\eta}\big)\big)\Big)^M\right] + \mathcal{O}\left(\frac{M}{K^{1/d}}\right),
\end{aligned}$$

as the number of points for the quantization of the exogenous noise K goes to $+\infty$, and where M stands for the size of the grids Γ_n.

Let us introduce $A_1,...,A_{N(\eta)}$, a cubic partition of $\mathrm{Supp}(\mu)$, which is bounded under **(Hμ)**, such that for all $j=1,\dots,N(\eta)$, A_j has diameter η. Also, Notice that there exists $c>0$, which only depends on $\mathrm{Supp}(\mu)$, such as

$$N(\eta) \le \frac{c}{\eta^d}. \tag{A2}$$

If $x\in A_j$, then $A_j\subset B(x,\eta)$, therefore:

$$\begin{aligned}
\mathbb{E}\Big[\left(1-\mu\left(B(X_n,\eta)\right)\right)^M\Big] &= \sum_{j=1}^{N(\eta)}\int_{A_j}\Big(1-\mu(B(x,\eta))\Big)^M\mu(dx)\\
&\le \sum_{j=1}^{N(\eta)}\int_{A_j}\Big(1-\mu(A_j)\Big)^M\mu(dx).
\end{aligned} \tag{A3}$$

Also notice that:

$$\sum_{j=1}^{N(\eta)}\mu(A_j)\Big(1-\mu(A_j)\Big)^M \le \sum_{j=1}^{N(\eta)}\max_z z(1-z)^M \le \frac{e^{-1}N(\eta)}{M}. \tag{A4}$$

Combining (A3) and (A4) leads to

$$\mathbb{E}\Big[\left(1-\mu\left(B(X_n,\eta)\right)\right)^M\Big] \le \frac{e^{-1}N(\eta)}{M}. \tag{A5}$$

Let $L = 2\|\mu\|_\infty$ stands for the diameter of the support of μ. We then get, as $M \to +\infty$,

$$\begin{aligned}
&\mathbb{E}\Big[\,\big|\mathrm{Proj}_{n+1}\big[F(X_n,a,\hat{\varepsilon}_{n+1})\big] - F(X_n,a,\hat{\varepsilon}_{n+1})\big|^2\,\Big] \\
&\qquad = \int_0^\infty \mathbb{P}\Big(\,\big|\mathrm{Proj}_{n+1}\big[F(X_n,a,\hat{\varepsilon}_{n+1})\big] - F(X_n,a,\hat{\varepsilon}_{n+1})\big|^2 > \eta\Big)\,\mathrm{d}\eta \\
&\qquad \le \int_0^{L^2} \frac{M[F]_L[\mu]_L}{K^{2/d}} + \mathbb{P}\Big(\big|\mathrm{Proj}_{n+1}\big[F(X_n,a,\hat{\varepsilon}_{n+1})\big] - F(X_n,a,\epsilon_{n+1})\big| > \sqrt{\eta}\Big)\,\mathrm{d}\eta \\
&\qquad = \int_0^{L^2} \min\left(1, \frac{e^{-1}N(\sqrt{\eta})}{M}\right)\mathrm{d}\eta + \mathcal{O}\left(\frac{M}{K^{1/d}}\right) \\
&\qquad = \int_0^{L^2} \min\left(1, \frac{c\eta^{-d/2}}{eM}\right)\mathrm{d}\eta + \mathcal{O}\left(\frac{M}{K^{1/d}}\right) \\
&\qquad = \int_0^{(c/(eM))^{(2/d)}} 1 d\eta + \int_{(c/(eM))^{(2/d)}}^{L^2} \frac{c\eta^{-d/2}}{eM}\,\mathrm{d}\eta + \mathcal{O}\left(\frac{M}{K^{1/d}}\right) \\
&\qquad = \frac{\tilde{c}^2}{M^{2/d}} + \mathcal{O}\left(\frac{M}{K^{1/d}}\right), \qquad \text{(A6)}
\end{aligned}$$

where $\tilde{c}$ is defined as $\tilde{c} := \sqrt{\frac{d}{d-2}}\left(\frac{c}{e}\right)^{1/d}$, and where we used (A5) and (A2) to go from the second to the third line. It remains to take $K = M^{d+1}$ points for the optimal quantization of the exogenous noise, and then take square root of equality (A6), in order to derive (A1). □

Lemma A.2 *Assume $d \ge 3$, take $K = M^{d+2}$ points for the optimal quantization of ε_n, and let $x \in E$. Then it holds under **(Hμ)** and **(HF)**, as $M \to +\infty$:*

$$\varepsilon_n^{proj}(x) \quad = \quad \mathcal{O}\left(\frac{1}{M^{1/d}}\right),$$

where $\varepsilon_n^{proj}(x)$, defined as $\varepsilon_n^{proj}(x) := \sup_{a\in A} \|\mathrm{Proj}_{n+1}\left(F(x,a,\hat{\varepsilon}_n)\right) - F(x,a,\hat{\varepsilon}_n)\|_2$, stands for the later-projection error at state x.

Proof. Following the same steps as those used to prove Lemma A.1, we show that:

$$\begin{aligned}
&\mathbb{P}\left(\big|\mathrm{Proj}_{n+1}\big[F(x,a,\hat{\varepsilon}_{n+1})\big] - F(x,a,\hat{\varepsilon}_{n+1})\big|^2 > \eta\right) \\
&\qquad = \frac{M[F]_L[\mu]_L}{K^{1/d}} + \mathbb{E}\left[\Big(1 - \mu\big(B\big(F(x,a,\varepsilon_{n+1}),\sqrt{\eta}\big)\big)\Big)^M\right] + \mathcal{O}\left(\frac{M}{K^{1/d}}\right),
\end{aligned}$$

as $K \to +\infty$, and moreover,

$$\mathbb{E}\left[\Big(1 - \mu\big(B\big(F(x,a,\varepsilon_{n+1}),\sqrt{\eta}\big)\big)\Big)^M\right] \le \frac{e^{-1}N(\eta)}{M},$$

holds, which is enough to complete the proof of Lemma A.2. □

LEMMA A.3 *Under* ***(HF)****, for* $n = 0, \dots, N$ *there exists constant* $\left[\hat{V}_n^Q\right]_L > 0$ *such that for* $x, x' \in E$*, it holds as* $M \to \infty$*:*

$$\left|\hat{V}_n^Q(x) - \hat{V}_n^Q(x')\right| \leq \left[\hat{V}_n^Q\right]_L |x - x'| + \mathcal{O}\left(\frac{1}{M^{1/d}}\right). \tag{A7}$$

Moreover, following bounds holds on $\left[\hat{V}_n^Q\right]_L$*, for* $n = 0, \dots, N$*:*

$$\begin{cases} \left[\hat{V}_N^Q\right]_L & \leq [g]_L \\ \left[\hat{V}_n^Q\right]_L & \leq [f]_L + [F]_L \left[\hat{V}_{n+1}^Q\right]_L, \qquad \text{for } n = 0, ..., N-1. \end{cases} \tag{A8}$$

Proof. Let us show that by induction that $\hat{V}_N^Q$ is Lipschitz. First, notice that (A7) holds at terminal time $n = N$, if one define $\left[\hat{V}_N^Q\right]_L$ as $\left[\hat{V}_N^Q\right]_L = [g]_L$. Let us take $x, x' \in E$. Assume $\left|\hat{V}_{n+1}^Q(x) - \hat{V}_{n+1}^Q(x')\right| \leq \left[\hat{V}_{n+1}^Q\right]_L |x - x'| + \mathcal{O}\left(\frac{1}{M^{1/d}}\right)$ holds for some $n = 0, \dots, N-1$. Let us show that

$$\left|\hat{V}_n^Q(x) - \hat{V}_n^Q(x')\right| \leq \left[\hat{V}_n^Q\right]_L |x - x'| + \mathcal{O}\left(\frac{1}{M^{1/d}}\right),$$

where $\left[\hat{V}_n^Q\right]_L$ is defined in (A8). Notice that, by the dynamic programming principle and the triangular inequality, it holds:

$$\begin{aligned} |\hat{V}_n^Q(x) - \hat{V}_n^Q(x')| &\leq [f]_L |x - x'| \\ &\quad + \sup_a \mathbb{E}_n^a \left[\left|\hat{V}_{n+1}^Q\big(\text{Proj}_{n+1}\left(F(x, a, \hat{\varepsilon}_{n+1})\right)\big) - \hat{V}_{n+1}^Q\big(\text{Proj}_{n+1}\left(F(x', a, \hat{\varepsilon}_{n+1})\right)\big)\right|\right] \\ &\leq [f]_L |x - x'| + \left[\hat{V}_{n+1}^Q\right]_L \sup_a \mathbb{E}\left[\left|\text{Proj}_{n+1}\left(F(x, a, \hat{\varepsilon}_{n+1})\right) - F(x, a, \hat{\varepsilon}_{n+1})\right|\right] \\ &\quad + \mathcal{O}\left(\frac{1}{M^{1/d}}\right) \\ &\leq \left([f]_L + \left[\hat{V}_{n+1}^Q\right]_L [F]_L\right) |x - x'| + \mathcal{O}\left(\frac{1}{M^{1/d}}\right) \\ &\leq \left[\hat{V}_n^Q\right]_L |x - x'| + \mathcal{O}\left(\frac{1}{M^{1/d}}\right), \end{aligned}$$

which completes the proof of (A7). □

We now proceed to the proof of Theorem 4.1.

Proof. (of Theorem 4.1) Combining inequality $|u_1 + u_2 + u_3|^2 \leq 3\left(|u_1|^2 + |u_2|^2 + |u_3|^2\right)$ that holds for all $u_1, u_2, u_3 \in \mathbb{R}$ with inequality $\left|\sup_{i \in I} a_i - \sup_{i \in I} b_i\right| \leq \sup_{i \in I} |a_i - b_i|$ that holds for all families $(a_i)_{i \in I}$

and $(b_i)_{i\in I}$ of reals, and all set I, we have:

$$\|\hat{V}_n^Q(X_n) - V_n(X_n)\|_2^2 \leq 3\,\mathbb{E}\Bigg[\sup_{a\in A}\mathbb{E}_{n,X_n}\left|\hat{V}_{n+1}^Q\left(\mathrm{Proj}_{n+1}\left(F(X_n,a,\hat{\varepsilon}_{n+1})\right)\right) - \hat{V}_{n+1}^Q(F(X_n,a,\hat{\varepsilon}_{n+1}))\right|^2$$
$$+\sup_{a\in A}\mathbb{E}_{n,X_n}\left|\hat{V}_{n+1}^Q(F(X_n,a,\hat{\varepsilon}_{n+1})) - \hat{V}_{n+1}^Q(F(X_n,a,\varepsilon_{n+1}))\right|^2$$
$$+\sup_{a\in A}\mathbb{E}_{n,X_n}\left|\hat{V}_{n+1}^Q(F(X_n,a,\varepsilon_{n+1})) - V_{n+1}(F(X_n,a,\varepsilon_{n+1}))\right|^2\Bigg]$$

where $\mathbb{E}_{n,X_n}$ stands for the expectation conditioned by the state X_n at time n. It holds as $M\to+\infty$, using Lemma A.3:

$$\|\hat{V}_n^Q(X_n) - V_n(X_n)\|_2^2 \leq 3\,\left[\hat{V}_n^Q\right]_L\mathbb{E}\Bigg[\sup_a\mathbb{E}_{n,X_n}\left[\left|\mathrm{Proj}_{n+1}\left(F(X_n,a,\hat{\varepsilon}_{n+1})\right) - F(X_n,a,\hat{\varepsilon}_{n+1})\right|^2\right]$$
$$+\sup_a\mathbb{E}_{n,X_n}\left[\left|F(X_n,a,\hat{\varepsilon}_{n+1}) - F(X_n,a,\varepsilon_{n+1})\right|^2\right]\Bigg]$$
$$+3\,\|r\|_\infty\mathbb{E}\left[\left|\hat{V}_{n+1}^Q(X_{n+1})) - V_{n+1}(X_{n+1}))\right|^2\right] + \mathcal{O}\left(\frac{1}{M^{1/d}}\right) \tag{A9}$$

Under **(HF)**, (A9) can then be rewritten as:

$$\|\hat{V}_n^Q(X_n) - V_n(X_n)\|_2^2 \;\leq\; 3\left[\hat{V}_n^Q\right]_L\left([F]_L^2(\epsilon_n^Q)^2 + (\epsilon_n^{proj})^2\right)$$
$$+3\|r\|_\infty\|\hat{V}_{n+1}^Q(X_{n+1}) - V_{n+1}(X_{n+1})\|_2^2 + \mathcal{O}\left(\frac{1}{M^{1/d}}\right).$$

(4.5) then follows by induction, which completes the proof of Theorem 4.1. □

Proof. (of Corollary 4.1)
Corollary 4.1 is straightforward by plugging the bound for the projection error provided by Lemma A.1 and the one of the quantization error provided by the Zador's Theorem into (4.5). □

Appendix B: Zador's Theorem

THEOREM B.1 (Zador's theorem) *Let us take $n = 0,\dots,N$, and denote by K the number of points for the quantization of the exogenous noise ε_n.*
Assume that $\mathbb{E}\left[|\varepsilon_n|^{2+\eta}\right] < +\infty$ for some $\eta > 0$. Then, there exists a universal constant $C > 0$ such that:

$$\lim_{M\to+\infty}\left(M^{\frac{1}{d}}\|\hat{\varepsilon}_n - \varepsilon_n\|_2\right) = C$$

Proof. We refer to Graf and Luschgy (2000) for a proof of Theorem B.1. □