High-Frequency Trading Meets Reinforcement Learning

Exploiting the iterative nature of trading algorithms

Joaquin Fernandez-Tapia*

July 9, 2015

Abstract

We propose an optimization framework for market-making in a limit-order book, based on the theory of stochastic approximation. The idea is to take advantage of the iterative nature of the process of updating bid and ask quotes in order to make the algorithm optimize its strategy on a trial-and-error basis (i.e. *on-line learning*) using a variation of the stochastic gradient-descent algorithm. An advantage of this approach is that the exploration of the system by the algorithm is performed in run-time, so explicit specifications of the price dynamics are not necessary, as is the case in the stochastic-control approach [8]. For the price/liquidity modeling, we consider a discrete-time variant of the Avellaneda-Stoikov model [1] similar to its developent in the article of Laruelle, Lehalle and Pagès [10] in the context of optimal liquidation tactics. Our aim is to set the ground for more advanced reinforcement learning techniques and argue that the rationale of our method is generic enough to be extended to other classes of trading problems besides market-making.

Keywords: high-frequency trading, market making, limit order book, stochastic approximation, reinforcement learning

1 Introduction

One of the main problems in algorithmic and high-frequency trading, is the optimization of tactics whose main role is to interact with the limit-order book, during a short lapse of time, in order to perform a basic task: this can be the optimal posting of a child order in the order book, routing an order across different liquidity pools or a high-frequency market-maker posting bid and ask quotes during a couple of seconds. Among the main features of these tactics is that they are short lived, have a straightforward goal and they are repeated several times during the trading session. Moreover, most of the performance of algorithmic trading tactics depends, not necessarily in financial aspects (as the valuation of an asset) but in microstructural factors (e.g. auction mechanics, tick-size, liquidity available in the book, etc.). Another important aspect is that the performance of these tactics is measured on a statistical basis, as their execution is systematic and, their use, intensive.

In the last years, several factors have driven the trading industry towards automation and an increased focus on trading algorithms. Among these we can mention:

- *Market fragmentation:* Regulations such as MiFID and RegNMS which have liberalized the 'exchanges market', allowing the existence of new electronic market-places where assets can be traded (which before was centralized on primary exchanges) and the appearance of new market participants (such as high-frequency traders).

*Laboratoire de Probabilités et Modèles aleatoires, UMR 7599, Université Paris 6, 4, pl. Jussieu, F-75252 Paris Cedex 5, France. E-mail: `joaquin.fernandez@upmc.fr`

- *Electronic access:* The access to exchanges is mostly electronic (specially in equities, and increasingly more for other asset classes). Nowadays, trading arises at very low latency and a automated approach is paramount for any kind of strategy (short term and long term).
- *Technology:* Evolutions in computing power and the massive amounts of data that today are possible to collect in real-time, motivate new approaches to handle the optimization process such as via sophisticated control techniques and the use of data mining and machine learning algorithms.

From a research perspective, a large number of contributions have been published in the recent years in the field of optimizing algorithmic-trading tactics (see for instants [6]). More concretely in the case of market-making, after the seminal paper of Avellaneda and Stoikov [1] several others have extended its approach (e.g. [2, 7] for optimal liquidation with limit orders and [4, 8] for market-making). However, most of the approaches to analyze these problems have been through the lenses of stochastic-control techniques, which demand to define explicitly the statistical laws governing the price dynamics. The latter makes the model less flexible for applications, where we would like the algorithm learn by itself in order to adapt to the nature of the forces driving the price and liquidity (which at the intraday scale can evolve through different market regimes difficult to anticipate).

In this article, we aim to take advantage of the iterative nature of the trading process at short time-scales by proposing an on-line learning framework to analyze the market-making problem (which can be extended to more general trading tactics) based on the theory of stochastic approximation [3, 5]. In terms of modeling of price and liquidity, we follow a similar path that [10] (in the context of optimal liquidation) by considering a modified version of the Avellaneda-Stoikov model [1]. One of the main aims of this study is to set the research ground for more advanced reinforcement-learning techniques, in the field of algorithmic and high-frequency trading, consistent with market microstructure.

1.1 The limit-order book

To understand the modeling rationale, we should first look at the *auction mechanism*, i.e. the set of matching rules that defines how the interaction between buyers and sellers is carried out on the exchanges. For our purposes (mostly concerned with equity trading with visible liquidity), the two main auction phases existing during the trading day are (i) *fixing auctions* (orders are matched after being accumulated on a book during a certain period) and (ii) *continuous auctions* (orders are matched continuously as they appear on the market). Most of the volume is traded during the continuous phase (indeed, besides exceptional situations, fixing auctions only happens at the very beginning and at the end of the day). Here, we focus on the case of an agent (more specifically, a market-maker – see below) participating on continuous auctions. In the equity market, the virtual place where offer meets demand during the continuous auction phase is called *limit-order book*.

There are two main ways to send trading intentions to the order book in order to buy or sell assets:

- Market orders: The agent specifies a quantity (to buy or sell) to be immediately executed. i.e. the agent consumes liquidity at the (best) available price.
- Limit orders: The agent specifies a quantity and the price to trade, then waiting until a market order arrives as counterpart. The price improvement is balanced by the non-execution risk and accepting a worse price in the future.

We say that market orders are *aggressive* trades while limit orders are *passive* and the match of these orders (in the case of equity markets) is done first by price-priority then by time-priority. If a participant who sends a limit order is no longer interested in keeping that order in the book he can cancel his order before it gets executed.

In the order book, prices can only exists in a discrete grid in which the minimum difference between two prices is called *the tick-size*. Moreover, the order book can be divided into two different sides: the *bid-side* (passive buyers) and the *ask-side* (passive sellers). The highest proposed bid-price is called *best-bid* and the lowest proposed ask-price is called *best-ask*. By design, the best bid-price is always lower than the best ask-price. If were not the case, a trade would have already occurred (i.e. the seller would have already matched the buyer). In the case of market-making, orders are set passively at both bid and ask sides; the goal being to execute both orders and earn their price difference.

At this 'microscopic' level, there is no clear definition for 'the price' of an asset. It can represents the *mid-price* (i.e. the average between the best-bid and the best-ask) or the price of the last trade. From our point of view, we just need to define some *reference price* around which we set our bid and ask orders. For convenience, a natural choice for a reference price is a sub-sampling of the mid-price at a rate of few seconds.

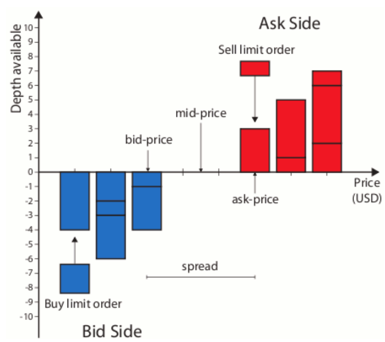

Figure 1: Graphical representation of the Limit Order Book.

1.2 The market-maker's problem

Throughout this work we consider a market-maker trading in an electronic limit order book. In a nutshell, the goal of a market maker is to provide liquidity by setting quotes at the bid and the ask sides of the order book. Each time one side of a bid/ask pair gets executed, the market-maker earns the price-difference between these two orders. Thus, the market-maker's algorithm would like to maximize the number of pairs of buy/sell trades executed, at the larger possible spreads and by holding the smallest possible inventory at the end of the trading session. Hence, the market-maker faces the following trade-off: it is expected that a large spread means a lower probability of execution while a narrower spread will mean a lower gain for each executed trade. On the other hand, if the trading algorithm only executes its orders on one side (because of price movements, for example), then its inventory moves away from zero, bearing the risk to eventually having to execute those shares at the end of the trading session at a worst price. The latter motivates the algorithm to center its orders around a 'fair price' in order to keep the inventory close to zero. Moreover, the more orders are executed, the larger the risk of ending the period with a large unbalanced inventory (variance effect), inducing still another trade-off.

From a modeling standpoint, one iteration of a market-making tactic can be seen as interacting with a black-box to which we apply, as input, the controls δ_a and δ_b (representing the positions in the order-book where to place orders with respect to some reference price) then, obtaining as output the variables N_a and N_b (depending on the controls and exogenous variables) representing the liquidity captured at each side of the spread during a time window of length ΔT (representing the duration for one iteration of the algorithm).

At the end of each iteration the payoff is represented by a random variable

$$\Xi(\delta_a, \delta_b) = \Pi(N_a(\delta_a, \xi), N_b(\delta_b, \xi), \xi) \tag{1}$$

where the ξ represent the exogenous variables influencing the payoff (e.g. price, spread).

In practice, ξ can both be a finite or an infinite dimensional random variable, and it is modeled in a way that it represents an *innovation* i.e. the part of the exogenous variables such as its realizations can be considered to be i.i.d. (or stationary ergodic or weakly dependent), which are the kind of property allowing the optimization procedure presented below, to converge. For example, instead of considering ξ being the price, from a modeling perspective, it will be more convenient to consider it to be the price returns.

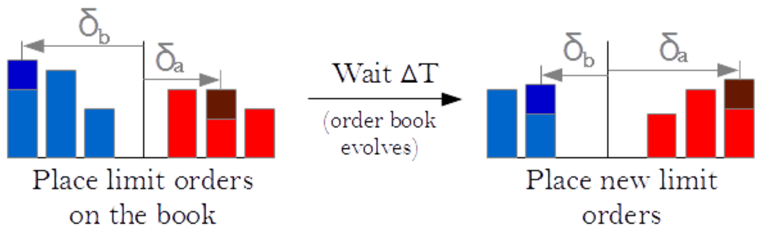

Figure 2: Scheme of the strategy. The market-maker places its orders in the limit order book, then he waits a lapse of ΔT. Market orders and limit orders arrive to the book during this period (hence, potentially representing a buying or sell of assets for the market-maker). After, this ΔT period and considering the new state of the order book and the performance of the strategy so far, the market-maker update his orders and the process is repeated.

1.3 On-line learning methods

Here we will want to maximize the expectation $\mathbb{E}[\Xi(\delta_a, \delta_b)]$. The question is, how, in order to find the solutions $\delta^* = (\delta_a^*, \delta_b^*)$, to exploit the iterative features of the algorithm (e.g. using errors as feedback) and, at the same time, to propose an adaptive framework less depending on an accurate modeling of the market-dynamics.

Our setting naturally joins the framework of stochastic approximation; where we define recursive procedures converging to the critical point of functions represented in the form of an expectation $h(\delta) = \mathbb{E}[H(\delta, \xi)]$, $\delta \in \mathbb{R}^d$, in cases when h cannot be computed, but where $H(\delta, \xi)$ and its derivatives can be evaluated at a reasonable cost and the random variable ξ can be simulated or obtained from a real-time data-stream (in the real-time case we call the procedure *on-line learning*). For our problem, we define $H(\delta_a, \delta_b, \xi) := \mathbb{E}[\Xi(\delta_a, \delta_b)|\xi]$, where $\Xi(\delta_a, \delta_b)$ is the realized payoff, defined on equation (1).

In order to maximize $h(\cdot)$, a probabilistic extension of the gradient method, namely

$$\delta_{n+1} = \delta_n + \gamma_{n+1} \nabla_\delta H(\delta_n, \xi_n), \quad \xi_n \sim \xi. \tag{2}$$

can be shown to converge towards the optimal value $\delta^* = (\delta_a^*, \delta_b^*)$.

As mentioned before, the advantage of the approach introduced and analyzed in this article, is its flexibility not only to approach problems where the price follows a Brownian diffusion (key hypothesis in the stochastic control approach) but also to much more general situations where the algorithm continuously extracts information from its environment without needing to further specify its dynamics. Moreover, the recursive aspect makes the procedure naturally adaptive and easily implementable. Notice also, that this framework can be generalized to other types of trading tactics (e.g. dark-pool trading, execution, routing to lit venues. See the works of Laruelle et al. [10, 9]), which can be seen as iterative problems where we control $\delta \in \mathbb{R}^d$, getting as output a 'liquidity captured' vector $N \in \mathbb{R}^p$, usually this N is represented by a vector of Poisson variables. Generically, the goal of a trading tactic is to maximize the expectation of a functional of the captured liquidity.

1.4 Outline

In this study we can look at stochastic approximation from two standpoints: first as a numerical procedure, in the sense that the innovation in the algorithm is a simulated random variable, and secondly, as a learning procedure in which the innovations are the observations of real-time data (or a historical data-set). We can also consider two different situations, depending on the way the market-maker valuates the inventory (mark-to-market or by adding a penalization function for the inventory). We will give an special focus on computing closed formulas in the case the price is Brownian, this is useful for several reasons: (i) exhibits numerical results (ii) highlights the relation between the different parameters and the solution and (iii) this can be used to compute 'first guesses' at the optimal solution and set them as a starting point for an algorithm working in a general case.

We start the next section by introducing in detail the model and the optimization problem. Section 2 focuses on our baseline situation (the mark-to-market situation) from a mathematical standpoint. Section 3 provides numerical examples and variations for the optimization method. Section 4 study extensions of the model (inventory control and relations with the dynamic programming approach). Finally we provide a conclusion, proof of the main results and a list of references.

2 Problem setting

2.1 Optimization problem

Let place ourselves on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$. We split the trading day $[0, T]$ into N periods of ΔT seconds each. The n-th period correspond to the time-interval $[(n-1)\Delta T, n\Delta T]$. We consider a market-making algorithm updating its bid and ask quotes at the beginning of each period. It posts its quotes at respective distances $\delta_a^{(n-1)}$ and $\delta_b^{(n-1)}$ around a reference price $S_{(n-1)\Delta T}$ fixed at the beginning of the ΔT-seconds period (this can be the mid-price at time $(n-1)\Delta T$, for instance). The orders remain unchanged until the end of the period.

Let us note $N_b^{(n)}$ and $N_a^{(n)}$ the respective number of orders the market-maker executes at the bid and at the ask during this n-th period. The realized payoff of the market-maker during this period is given by the following random-variable:

$$\Xi_n = \underbrace{N_a^{(n)} \left(S_{(n-1)T} + \delta_a^{(n-1)} \right)}_{\text{sold}} - \underbrace{N_b^{(n)} \left(S_{(n-1)T} - \delta_b^{(n-1)} \right)}_{\text{bought}} + \underbrace{\left(N_b^{(n)} - N_a^{(n)} \right) S_{nT}}_{\text{inventory valuation}}.$$

We introduce for each time period, the running price return

$$Y_t^{(n)} = S_{(n-1)\Delta T + t} - S_{(n-1)\Delta T}, \quad t \in [0, \Delta T].$$

In particular $Y_{\Delta T}^{(n)} = S_{nT} - S_{(n-1)T}$.

Now, let us introduce the variables

$$N^{(n)} = N_b^{(n)} + N_a^{(n)} \quad \text{and} \quad Q^{(n)} = N_b^{(n)} - N_a^{(n)}.$$

Also, we define the market-maker' *half-spread* as

$$\psi^{(n)} = \frac{\delta_a^{(n)} + \delta_b^{(n)}}{2}$$

and an *off-center* factor $\theta^{(n)}$ such as

$$\theta^{(n)} = \frac{\delta_a^{(n)} - \delta_b^{(n)}}{2}.$$

The latter variables have an intuitive interpretation: $N^{(n)}$ represents the total number of trades executed by the market-maker and $Q^{(n)}$ is the increase of inventory over the period $[(n-1)\Delta T, n\Delta T]$.

In our new variables, the realized payoff of the market-maker is given by:

$$\Xi_n = \psi^{(n)} N^{(n)} - \left(\theta^{(n)} - Y_{\Delta T}^{(n)}\right) Q^{(n)}.$$

The goal of the market-maker is to maximize the expectation of this quantity.

2.2 The one-period problem

Notice that if we suppose that the law of $Y^{(n)}$ (and $N^{(n)}$ and Q^n) does not depends on $S_{(n-1)\Delta T}$, we can always think in terms of a one-period problem as the function to maximize is always the same. i.e.

$$\max_{\psi \geq 0, \theta \in \mathbb{R}} \pi(\psi, \theta) := \mathbb{E}\left[\Xi_1\right]. \tag{3}$$

In this study, for the sake of putting the modeling forward, we focus on this *independent price-increments* situation leading to a *one-period problem*.

However, it is important to keep in mind that these hypothesis can be weakened and interpreted as an approximation for more general cases, namely, when price increments are stationnary, weakly dependent or in an ergodic setting in which we want to optimize the sum of future payoffs (from an asymptotic point of view).

In what follows, it will be interesting to consider the payoff conditioned to the trajectory of the price (as it represents an exogenous factor). Hence, for a function y_t, $t \in [0, \Delta T]$, let us define:

$$\Pi(\psi, \theta, (y_t)_{t \in [0, \Delta T]}) = \mathbb{E}_y\left[\Xi_n\right]. \tag{4}$$

In particular $\pi(\psi, \theta) = \mathbb{E}\left[\Pi(\psi, \theta, Y)\right]$, with $Y \sim Y_{\Delta T}^{(1)}$.

2.3 Stochastic-gradient method

The goal of the market maker is to maximize the expectation of the function $\mathbb{E}\left[\Pi(\psi, \theta, Y)\right]$. In a concave setting, this is equivalent of finding the zero of $\nabla \pi(\psi, \theta) = \nabla_{\psi, \theta} \mathbb{E}\left[\Pi(\psi, \theta, Y)\right]$. After showing that the solution of this zero-searching problem is unique, we can consider applying the Robbins-Monro theorem, which states that the zero of $\nabla \pi$, under a given set of hypothesis, can be found through an algorithm of the form:

$$\begin{aligned} \psi_{n+1} &= \psi_n + \gamma_{n+1} \partial_\psi \Pi(\psi_n, \theta_n, Y_{n+1}) \quad (5) \\ \theta_{n+1} &= \theta_n + \gamma_{n+1} \partial_\theta \Pi(\psi_n, \theta_n, Y_{n+1}). \quad (6) \end{aligned}$$

We will see that in our setting one of the conditions of the Robbins-Monro theorem will not be satisfied. Namely

$$\mathbb{E}\left[\left|\nabla_{\psi, \theta} \Pi(\psi, \theta, Y)\right|^2\right] \leq C(1 + \psi^2 + \theta^2)$$

however, by modifying the function by a multiplicative factor, as it is proposed in the article by Lemaire and Pagès [11], we can show that a procedure of the form:

$$\begin{aligned} \psi_{n+1} &= \psi_n + \gamma_{n+1} \rho(\psi_n, \theta_n) \partial_\psi \Pi(\psi_n, \theta_n, Y_{n+1}) \quad (7) \\ \theta_{n+1} &= \theta_n + \gamma_{n+1} \rho(\psi_n, \theta_n) \partial_\theta \Pi(\psi_n, \theta_n, Y_{n+1}) \quad (8) \end{aligned}$$

converges towards the solution (i.e. the point (ψ^*, θ^*) solution of $\nabla \pi(\psi, \theta) = 0$).

Here ρ is a strictly positive function (to be defined), mainly intended to control the behavior of the function Π for large values of θ.

2.4 Relating price and liquidity

Thus far we have introduced the optimization problem regardless of the explicit relations between price and liquidity. The interest of introducing a price dynamics is to relate statistically the price with the captured liquidity, i.e. N and Q. The latter is related to the fundamental relations between price direction and order imbalance for example.

In this chapter we will use the variation of the Avellaneda-Stoikov model as in [10].

We consider that, during a period $t \in [0, \Delta T]$ the probability to capture liquidity at a distance δ from the reference price Y_t is defined between t and $t + dt$, up to the second order, by the instantaneous intensity

$$\lambda(\delta) dt = A e^{-\delta k} dt. \tag{9}$$

and independent from the trajectory of Y.

Thus, the liquidity captured by an order placed at $S_0 \pm \delta$, during $[0, \Delta T]$, given the trajectory of the price is $(y_t)_{t \geq 0}$, will be Poisson random variables with intensities:

$$\lambda^-(\delta_b; (y_t)_{t \in [0, \Delta T]}) = \int_0^{\Delta T} A e^{-k(\delta_b + y_t)} dt \qquad \text{(bought shares)}, \tag{10}$$

$$\lambda^+(\delta_a; (y_t)_{t \in [0, \Delta T]}) = \int_0^{\Delta T} A e^{-k(\delta_a - y_t)} dt \qquad \text{(sold shares)}. \tag{11}$$

So, the captured liquidity $N^{(a)}$ and $N^{(b)}$ is modeled by Poisson random variables with intensities given (respectively) by λ^+ and λ^-.

The latter can be seen in two ways: (i) simply define the intensities of the Poisson variables $N^{(a)}$ in that $N^{(b)}$ as functionals of the realized price and simply using the preceding justification as a heuristic, or (ii) formally model how prices and liquidity are related in continuous-time (i.e. thinking in terms of the generalized Poisson processes $N_t^{(a)}$ and $N_t^{(b)}$). In the latter case we need to introduce a standard Poisson process $(N_u)_{u \geq 0}$ independent of Y such that the orders we capture are given by $\tilde{N}_s$ which correspond to a time change of N defined by $ds = A e^{-k(\delta + Y_u)} du$ or via a thinning operation.

Again, in order to work with symmetric variables, we define for every $y \in \mathcal{C}([0, \Delta T], \mathbb{R})$,

$$\lambda(\psi, \theta; y) = \lambda^- + \lambda^+ = 2 A e^{-k\psi} \int_0^{\Delta T} \cosh(k(\theta - y_t)) dt, \tag{12}$$

$$\mu(\psi, \theta; y) = \lambda^- - \lambda^+ = 2 A e^{-k\psi} \int_0^{\Delta T} \sinh(k(\theta - y_t)) dt. \tag{13}$$

with λ^- and λ^+ defined in (10) and (11). In particular we have $\mathbb{E}_Y[N] = \lambda$ and $\mathbb{E}_Y[Q] = \mu$. Furthermore, because negative spreads do not make sense in our problem, we suppose $\psi \geq 0$ and $\theta \in \mathbb{R}$. Moreover, in practice it does not makes sense to have passive bid quotes on the ask side (and vice-versa), which means we look for solutions on $|\theta| < C + \psi$, where C represents the half bid-ask spread (or an upper bound for it).

In this way we have:

$$\Pi(\psi, \theta; Y) = \psi \lambda(\psi, \theta; Y) - (\theta - Y_{\Delta T}) \mu(\psi, \theta; Y). \tag{14}$$

Throughout this stutyd it will be interested in comparing with the situation when the price is Brownian, where we can obtain closed formulas for the optimal quotes (see Theorem below). This have several advantages: (i) exhibit numerical results (ii) understand the relation between the different parameters and the solution and (iii) compute 'first guesses of the optimal solution' and set them as the starting point for an algorithm working on a general case. The following theorem gives the optimal quotes for the Brownian situation:

Theorem 1. *In the Brownian case, the optimal solution is given by*[1]*:*

$$\psi^* = \frac{1}{k}\left(\frac{k^2\sigma^2\Delta T - 1 + e^{-\frac{k^2\sigma^2\Delta T}{2}}}{1 - e^{-\frac{k^2\sigma^2\Delta T}{2}}}\right), \tag{15}$$

$$\theta^* = 0. \tag{16}$$

Proof. See Appendix A.1. □

3 Mark-to-Market inventory valuation

3.1 Introduction

First, we consider that no market-impact affecting the liquidation of the inventory. That is, the conditional expectation of the profits, given the price trajectory, is defined by:

$$\Pi(\psi, \theta; Y) = \psi\lambda(\psi, \theta; Y) - (\theta - Y_{\Delta T})\mu(\psi, \theta; Y). \tag{17}$$

while the function to maximize is defined by:

$$\pi(\psi, \theta) = \mathbb{E}\left[\Pi(\psi, \theta; Y)\right]. \tag{18}$$

Let us notice that in this case we can write the target function as

$$\pi(\psi, \theta) = 2Ae^{-k\psi}(a(\theta)\psi - b(\theta)), \tag{19}$$

where:

$$a(\theta) = \mathbb{E}\left[\int_0^{\Delta T} \cosh(k(\theta - Y_t))dt\right], \tag{20}$$

$$b(\theta) = \mathbb{E}\left[(\theta - Y_T)\int_0^{\Delta T} \sinh(k(\theta - Y_t))dt\right]. \tag{21}$$

3.2 Existence and uniqueness of the solution

The next theorem states the existence and uniqueness of the maximum and in particular, an unique solution for the equation $\nabla\pi(\psi, \theta) = 0$.

Theorem 2. *The function $\pi(\psi, \theta) = e^{-k\psi}(\psi a(\theta) - b(\theta))$ has an unique maximum.*

Proof. See Appendix A.2. □

With this result, we now need to compute the local derivatives of the function $\Pi(\psi, \theta; Y)$ then verify is the conditions of the Robbins-Monro theorem are satisfied.

[1]This result, as $\Delta T \to 0$, is consistent with the asymptotic solution in the article [8] in the case $\gamma \to 0$.

3.3 Stochastic gradient

In order to define the stochastic algorithm we start by computing the derivatives of the function $\Pi(\psi, \theta; y_t)$ with respect to ψ and θ.

Proposition 1. *The gradient $\nabla_{\psi,\theta}\Pi(\psi, \theta; y_t)$ is given by the equation:*

$$\begin{pmatrix} \frac{\partial}{\partial \psi}\Pi(\psi, \theta; y_t) \\ \frac{\partial}{\partial \theta}\Pi(\psi, \theta; y_t) \end{pmatrix} = \begin{pmatrix} (1 - k\psi) & k(\theta - y_{\Delta T}) \\ -k(\theta - y_{\Delta T}) & -(1 - k\psi) \end{pmatrix} \begin{pmatrix} \lambda(\psi, \theta; y_t) \\ \mu(\psi, \theta; y_t) \end{pmatrix}. \tag{22}$$

Proof. See Appendix A.3 □

In order to prove the convergence of the stochastic algorithm, we modify the procedure, as in [11], by multiplying the gradient by a factor $\rho(\psi, \theta)$ which help us to guarantee the convergence of the modified stochastic gradient algorithm:

$$\psi_{n+1} = \psi_n - \gamma_{n+1}\rho(\psi_n, \theta_n)\partial_\psi \Pi(\psi_n, \theta_n, Y_{n+1}) \tag{23}$$

$$\theta_{n+1} = \theta_n - \gamma_{n+1}\rho(\psi_n, \theta_n)\partial_\theta \Pi(\psi_n, \theta_n, Y_{n+1}) \tag{24}$$

Where $\rho(\psi, \theta) = e^{-k(|\theta| - \psi)}$.

Indeed the function $\rho(\psi, \theta)\nabla\Pi(\psi, \theta, Y)$ satisfies the condition of the Robbins-Monro algorithm as it is proven in the following theorem.

Theorem 3. *Let us consider the function $\rho(\psi, \theta) = e^{-k(|\theta| - \psi)}$, $\theta \in \mathbb{R}$, and, for a i.i.d. sequence of random trajectories Y_{n+1} (with same law as Y), the following recursive algorithm:*

$$\begin{aligned} \psi_{n+1} &= \psi_n + \gamma_{n+1}\rho(\psi_n, \theta_n)\partial_\psi \Pi(\psi_n, \theta_n, Y_{n+1}) \\ \theta_{n+1} &= \theta_n + \gamma_{n+1}\rho(\psi_n, \theta_n)\partial_\theta \Pi(\psi_n, \theta_n, Y_{n+1}), \end{aligned}$$

where ψ_0 and θ_0 are in $L^1(\mathbb{P})$, $\sum_n \gamma_n = +\infty$ and $\sum_n \gamma_n^2 < +\infty$.

Consider also the following technical conditions

$$\begin{aligned} \mathbb{E}\left[Y_T^2 \int_0^{\Delta T} e^{2k|Y_t|}dt\right] &< +\infty, \\ \mathbb{E}\left[\int_0^{\Delta T} e^{2k|Y_t|}dt\right] &< +\infty. \end{aligned}$$

Then, the recursive algorithm converges towards the point (ψ^*, θ^*) which correspond to the maximum of $\pi(\psi, \theta) = \mathbb{E}[\Pi(\psi_n, \theta_n, Y)]$ (which is the unique solution of the equation $\nabla\pi = 0$).

Proof. See Appendix A.3. □

A delicate issue that it has not been addressed up to this point is that the innovation is an infinite dimensional object (trajectory of the reference price), while in the Robbins-Monro theorem (see the proof of the last theorem) is set up for finite dimensional innovations. Fortunately, this is not a serious issue as in fact the algorithm only depends on the innovation through a finite dimensional functional of the trajectory (indeed, the innovation is a 3-dimensional object, after a change of variable).

Theorem 4. *The function $\Pi(\psi, \mu, Y)$ only depends on the stochastic process $(Y)_{t\in[0,T]}$ through the one-dimensional random variables Y_T and:*

$$b_k = \frac{1}{2}\log\left(\int_0^T e^{kY_t}dt \int_0^T e^{-kY_t}dt\right), \tag{25}$$

$$\rho_k = \frac{1}{2}\log\left(\frac{\int_0^T e^{-kY_t}dt}{\int_0^T e^{kY_t}dt}\right). \tag{26}$$

Moreover, we have the formulas

$$\lambda(\psi,\theta) = 2Ae^{-k\psi+b_k}\cosh(k\theta+\rho_k), \quad (27)$$

$$\mu(\psi,\theta) = 2Ae^{-k\psi+b_k}\sinh(k\theta+\rho_k). \quad (28)$$

Proof. See Appendix A.3. □

4 Numerical examples

4.1 Numerical example: Brownian case

We implement the stochastic optimization algorithm in the Brownian case and compare with the explicit solutions we just found. We consider an algorithm with projections in order to be always searching for the solution in the region where the payoff is expected to be positive.

4.1.1 Shape of the function to maximize

We set the following values for the models parameters

$$T = 30, \quad \sigma = 1.2, \quad A = 0.9, \quad k = 0.3.$$

The maximum is reached at:

$$\psi^* = 2.52024 \pm 10^{-6}, \theta^* = 0.0$$

The following figure shows a heatmap of the function:

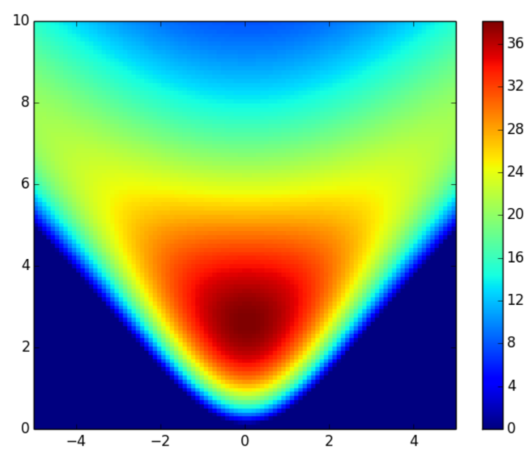

Figure 3: Heatmap of the target function in the Brownian situation (the abscissa correspond to θ and the ordinate correspond to ψ).

Three previous results has been helpful to reduce the computing cost:

- Separating variables on the representation of the function
- The innovations are in reality a three-dimensional variable
- Recognizing the admissible region (outside the admissible region the function has a very steeped derivative, creating numerical problems)

4.2 Stochastic algorithm

We apply the stochastic algorithm taking as initial point $(8, 1)$ and step $0.3 \times n^{-1}$.

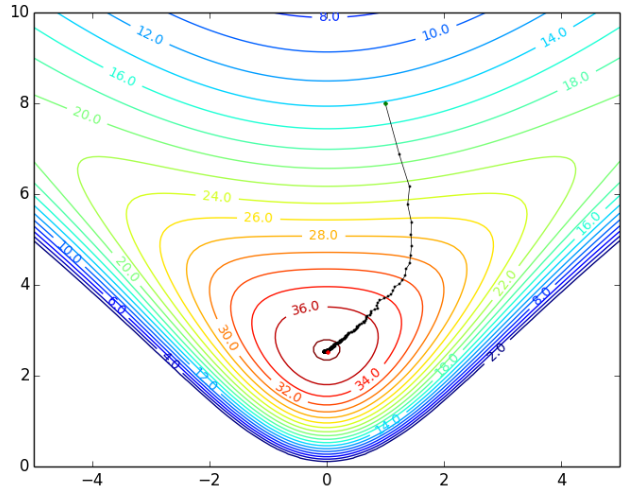

Figure 4: Example of the stochastic algorithm converging towards the solution.

4.2.1 Wider steps and Ruppert-Polyak

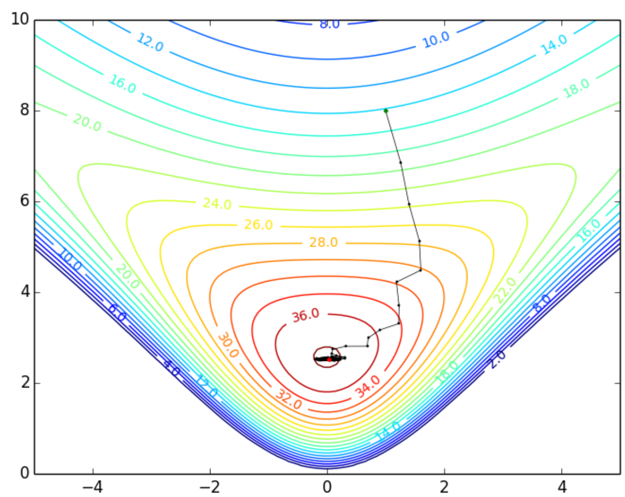

Figure 5: Convergence of the algorithm with exponent 0.6.

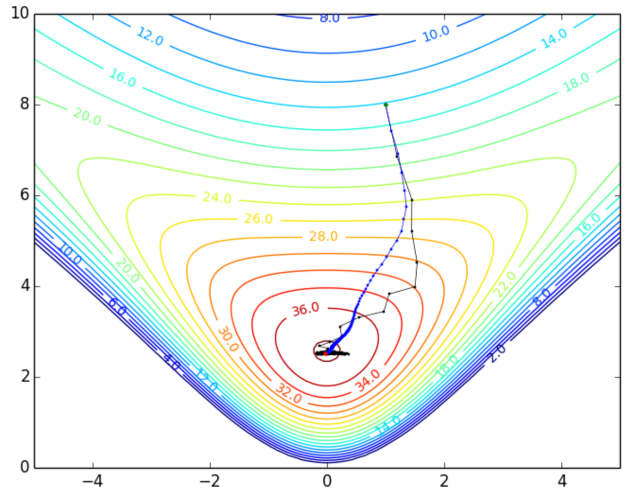

Figure 6: Convergence of the algorithm with exponent 0.6 in the Ruppert-Polyak situation.

4.2.2 Algorithm freeze-up

In the examples we just saw, the constant $C = 0.3$ was chosen manually by looking at previous numerical experimentation. One problem we can experience in practice, when the algorithm has a $\mathcal{O}(n^{-1})$ step, is that the algorithm freezes-up, that is, it starts to take too much iterations to converge as the step is gets too small for a large n.

The following figure shows how the algorithm freezes-up if we set $C = 0.1$.

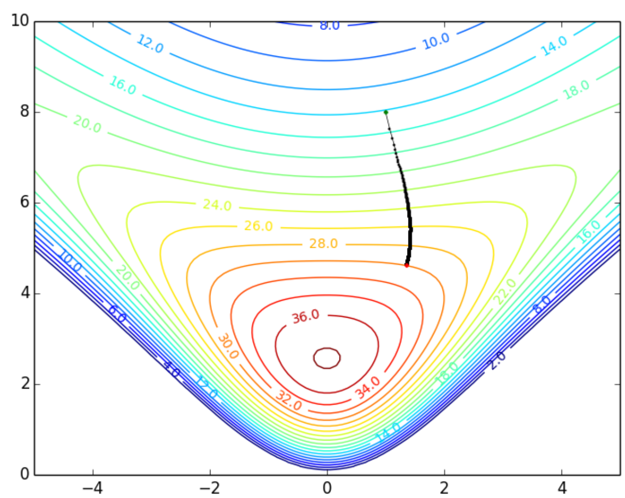

Figure 7: Example of a situation when the algorithm freezes up.

One of the advantages of the Ruppert-Polyak approach is that the algorithm gets the best of two worlds: it explores the environment on early stages in order to get closer to the solution, then the averaging improves the convergence once the algorithm is near the solution.

The following figure shows the algorithm with step Cn^{-6} (black path) and the Ruppert-Polyak averaged algorithm (blue path).

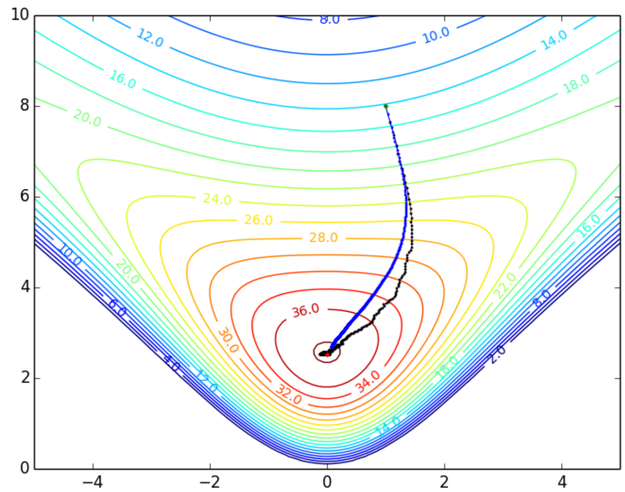

Figure 8: Ruppert-Polyak avoids the freeze-up situation.

5 Penalizing the inventory

In order to understand the reasons for penalizing the inventory, it is interesting to make a parallel with the stochastic-control approach as in [8].

5.1 Relation to the stochastic-control approach

A key hypothesis in our model is that the incoming information have a stationary dynamics and the function to maximize is the same at each iteration. We have already mentioned the advantages of this approach (e.g. adaptability, model-free) however it has the weakness of not having a view on the whole strategy (i.e. inventory risk) which is the strong point on the stochastic-control approach [8] as it is based on the dynamic-programming principle.

However, as it is mentioned in the papers [7, 8] the solutions of the Hamilton-Jacobi-Bellman equation solving the optimal quotes in market-making and optimal liquidation are close to those of the asymptotic regime if we are not close to the end of the trading session. Otherwise said, the hypothesis of stationary innovations is not a constraining one as in the stochastic-control situation; we are solving the 'asymptotic problem'.

Moreover, if we try to apply the dynamic-programming approach on our discrete version of the Avellaneda-Stoikov problem it is easy to see that the utility function we consider in our forward approach correspond to the situation when in the dynamic programming problem there is no risk-aversion (i.e. the market-maker just wants to maximize the PnL).

Otherwise said, the dynamic programming approach solves the inventory risk problem while the on-line learning approach focuses on adapting to changing market dynamics and eventually adverse selection. Adding penalization function on the remaining inventory has as goal add this 'inventory risk' dimension to the problem.

5.2 Changes in the structures of the target-function

So far, we have been valuating the inventory at the end of each period mark-to-market. In practice, when liquidating a given quantity in the market, agents incur in liquidation costs. Let us $\Phi(\cdot)$ be the cost of liquidating a remaining inventory Q. We suppose $\Phi(\cdot)$ positive, even and convex. Thus, we define:

$$\kappa(\lambda,\mu) = \mathbb{E}_Y\left[\Phi\left(Q\right)\right]. \tag{29}$$

The function to be maximized becomes

$$\pi(\psi,\theta) = \mathbb{E}\left[\psi\lambda - (\theta - Y_T)\mu - \kappa(\lambda,\mu)\right]. \tag{30}$$

When we add liquidation costs the dependency in λ and μ is no longer linear. Indeed, in most of the interesting situations there is no closed formula for $\kappa(\lambda,\mu)$.

There are two ways of thinking about the penalization in this context:

- Quantifying costs as if we were liquidating the inventory at the end of each period (to penalize the market-maker payoff), in this way, liquidation costs represent bid-ask spread costs and market impact costs. In that way we set:

$$\Phi(Q) = \underbrace{C|Q|}_{\text{bid}-\text{ask spread}} + \underbrace{\gamma|Q|^{1+\alpha}}_{\text{market impact}} \quad , \quad 0 \le \alpha < 1. \tag{31}$$

  An important situation is $\Phi(Q) = C|Q|$ (i.e. $\gamma = 0$) as it represents the case where the only cost is the bid-ask spread, quantified by the real number $C > 0$. Another situation of interest is $\Phi(Q) = CQ^2$ as it provides a case where we can obtain a closed formula for $\kappa(\lambda,\mu)$. We will look at these two situations in further detail.

- The other way, is to think that the penalization term in the inventory represents a function quantifying costs from the point of view of an (external) algorithm, controlling the inventory risk of the overall strategy. In other words, even though we are considering the 'one-period' problem, in practice the penalization function may be evolving over time (e.g. at a lower frequency than the refreshing of the algorithm).

  For example, we can think that early in the trading session we value the inventory in a mark-to-market way, but as we are near the end of the day, we increase the weight we give to the penalization term.

  For practical purposes we will address only the first situation here. However we will study general formulas for the expectation of liquidation costs, which can be useful in further studies in this case.

5.3 Outline of the stochastic algorithm

The idea is to apply the same recursive optimization procedure than in the precedent section, that is (some variation) of an algorithm of the form:

$$\psi_{n+1} = \psi_n + \gamma_{n+1}\partial_\psi \Pi(\psi_{n+1},\theta_{n+1},Y) \tag{32}$$
$$\theta_{n+1} = \theta_n + \gamma_{n+1}\partial_\theta \Pi(\psi_{n+1},\theta_{n+1},Y) \tag{33}$$

The following formula allows to easily compute the derivatives of liquidation costs when $\Phi(\cdot)$ is even and $\Phi(0) = 0$.

Proposition 2.

$$\partial_\lambda \mathbb{E}(\Phi(Q)) = \frac{1}{2}\left(\mathbb{E}(\Phi(Q+1)) - 2\mathbb{E}(\Phi(Q+1)) + \mathbb{E}(\Phi(Q-1))\right) \tag{34}$$

$$\partial_\mu \mathbb{E}(\Phi(Q)) = \frac{1}{2}\left(\mathbb{E}(\Phi(Q-1)) - \mathbb{E}(\Phi(Q+1))\right) \tag{35}$$

Proof. See Appendix A.4. □

5.4 Closed-formulas for the expectation

The following theorem provides useful closed formulas for $\mathbb{E}_Y[\Phi(Q)]$:

Theorem 5. *Let λ, μ and ε be defined by:*

$$\lambda(\psi,\theta;Y) = 2Ae^{-k\psi}\int_0^T \cosh(k(\theta - Y_t))dt, \tag{36}$$

$$\mu(\psi,\theta;Y) = 2Ae^{-k\psi}\int_0^T \sinh(k(\theta - Y_t))dt, \tag{37}$$

$$\varepsilon(\theta;Y) := \frac{\lambda(\psi,\theta;Y)}{\mu(\psi,\theta;Y)} = \frac{\int_0^T \sinh(k(\theta - Y_t))dt}{\int_0^T \cosh(k(\theta - Y_t))dt}. \tag{38}$$

Then, the following formulas hold:

$$\begin{aligned}
\mathbb{E}_Y[Q^2] &= \lambda + \lambda^2\varepsilon^2, \\
\mathbb{E}_Y[|Q|] &= \int_0^\lambda e^{-s} I_0\left(s\sqrt{1-\varepsilon^2}\right) ds \\
&+ 2|\varepsilon| \sum_{n=1}^{\infty} \sinh\left(\frac{n}{2}\log\left(\frac{1+|\varepsilon|}{1-|\varepsilon|}\right)\right) \int_0^\lambda e^{-s} I_n\left(s\sqrt{1-\varepsilon^2}\right) ds,
\end{aligned}$$

*where $I_n(\cdot)$ denotes the modified Bessel function of order n (see [**?**]).*

More generally, for $\Phi(\cdot)$ even, increasing on $\mathbb{R}^+$ and $\Phi(0) = 0$, we have:

$$\begin{aligned}
\mathbb{E}_Y[\Phi(Q)] &= \Phi(1)\int_0^\lambda e^{-s} I_0\left(s\sqrt{1-\varepsilon^2}\right) ds \\
&+ 2\sum_{n=1}^{\infty} D^2\Phi(n) \cosh\left(\frac{n}{2}\log\left(\frac{1+|\varepsilon|}{1-|\varepsilon|}\right)\right) \int_0^\lambda e^{-s} I_n\left(s\sqrt{1-\varepsilon^2}\right) ds \\
&+ 2|\varepsilon|\sum_{n=1}^{\infty} D\Phi(n) \sinh\left(\frac{n}{2}\log\left(\frac{1+|\varepsilon|}{1-|\varepsilon|}\right)\right) \int_0^\lambda e^{-s} I_n\left(s\sqrt{1-\varepsilon^2}\right) ds.
\end{aligned}$$

Where, for $n \in \mathbb{Z}$, we defined:

$$D^2\Phi(n) = \frac{\Phi(n+1) - 2\Phi(n) + \Phi(n-1)}{2},$$

$$D\Phi(n) = \frac{\Phi(n+1) - \Phi(n-1)}{2}.$$

Proof. See Appendix A.4. □

5.5 Bounds for liquidation costs

5.5.1 Upper bound for the liquidation-costs

Conditionally to Y we can place ourselves in the case where $\lambda(\psi, \theta; Y)$ and $\mu(\psi, \theta; Y)$ are considered as constant. As Q is always an integer, we get:

$$\mathbb{E}(\Phi(Q)) \leq (C + \gamma)\mathbb{E}(Q^2) = (C + \gamma)(\lambda + \mu^2)$$

5.5.2 Lower bound for the liquidation-costs

The first lower bound follows Jensen inequality: $\mathbb{E}(\Phi(Q)) \geq \Phi(\mu)$. However, this bound is not very useful when $\mu = 0$. A better bound can be obtained by observing the following two facts:

- Conditional to N the variable Q can be writen as $Q = 2B - N$ where B follows a binomial law with parameters (N, p).
- Secondly we use the fact that $\Phi(Q) \geq (C + \gamma)|Q|$; We want to find a lower bound for $\mathbb{E}[|Q|]$.

Theorem 6. *We have*

$$\mathbb{E}(|Q|) \geq \frac{\lambda}{2\sqrt{2(\lambda + 1)}} \geq \frac{1}{4} \min\left(\lambda, \sqrt{\lambda}\right).$$

Leading to:

$$\mathbb{E}(\Phi(Q)) \geq \left(\frac{C + \gamma}{4}\right) \min\left(\lambda, \sqrt{\lambda}\right).$$

Proof. See Appendix A.4. □

5.6 Existence of a minimum

As in the mark-to-market case, we want to show first that for a fixed θ there is a unique ψ where $\mathbb{E}[\Pi(\psi, \theta; Y)]$ has its maximum; in fact, we can show that this is true every time the function κ has the following form:

$$\kappa(\lambda, \mu) = \sum_{j=1}^{J} \lambda^j g_j\left(\frac{\mu}{\lambda}\right), \quad g_j\left(\frac{\mu}{\lambda}\right) \geq 0$$

Proposition 3. *If the conditional expectation of the liquidation costs is characterized by the function*

$$\kappa(\lambda, \mu) = \sum_{j=1}^{J} \lambda^j g_j\left(\frac{\mu}{\lambda}\right), \quad g_j(x) \geq 0, \forall x \in \mathbb{R}.$$

Then, for any fixed θ_0 such as $a(\theta_0) \geq b(\theta_0)$ we can find an unique ψ_{θ_0} maximizing $\mathbb{E}[\pi(\psi, \theta_0; Y)]$.

Proof. See Appendix A.4. □

6 Conclusion

In this article we provide a framework, based on the theory of stochastic approximation, for solving the problem of a market-maker participating on an electronic limit-order book. The idea is to take advantage of the iterative nature of trading tactics when proceeding algorithmically, on a high-frequency basis and when the performances are measured statistically. The advantage of our framework is that it is devised for the type of situations where we aim for a model-free approach in which the algorithm extract information from the environment during its execution, i.e. more adapted to the case of a liquid stock in which the velocity of order-book data allows the algorithm a rapid learning.

The mathematical proofs of the convergence of our learning algorithm are based on the Robbins-Monro theorem, and its formal development are provided in appendix as well as results guaranteeing that the problem is well-posed, i.e. it exists a solution, and this solution is unique.

We also studied the situation in which penalization costs are included in order to control inventory risk (maybe the only weakness of the framework, compared to the stochastic-control approach). We provided mathematical results on the analytical properties of the penalization cost function.

Among the possible future directions of research for this work are:

- Generalize the approach to a wider class of trading-tactics.
- Study in detail the relation with the approaches based on the dynamic programming principle (i.e. obtain the best of two worlds in terms of on-line learning and inventory control).
- Study stochastic algorithms in continuous time, not as a model approximation as in the stochastic-control approach, but, for example, by updating quotes in Poissonian time (as it is the natural way at which the algorithm aggregates information from markets). In this case the optimal quotes can be seen as controls of a Poisson process whose events update the control itself, so adding a source of self-reinforcement.

A Proof of the main results

A.1 Closed formulas in the Brownian case

Proof of Theorem 1:

Let us start by defining

$$\begin{aligned} a(\theta) &= \mathbb{E}\left[\int_0^{\Delta T} \cosh(k(\theta - Y_t))dt\right] \\ b(\theta) &= \mathbb{E}\left[(\theta - Y_T)\int_0^{\Delta T} \sinh(k(\theta - Y_t))dt\right]. \end{aligned}$$

The function to maximize can be writen as

$$\pi(\psi, \theta) = 2Ae^{-k\psi}(\psi a(\theta) - b(\theta)).$$

By classical calculus we can show that the maximum of this function satisfies

$$\psi = \frac{1}{k} + \frac{b(\theta)}{a(\theta)},$$

which in the Brownian case can be explicitly computed, as it is shown below.

Proposition 4. *If the reference price evolves following the stochastic differential equation $dS_t = \sigma dW_t$, then we have the following closed formulas:*

$$\begin{aligned} a(\theta) &= \frac{2\cosh(k\theta)}{k^2\sigma^2}\left(e^{\frac{k^2\sigma^2\Delta T}{2}} - 1\right), \\ b(\theta) &= \frac{2\theta\sinh(k\theta)}{k^2\sigma^2}\left(e^{\frac{k^2\sigma^2\Delta T}{2}} - 1\right) \\ &+ \frac{2\cosh(k\theta)}{k^3\sigma^2}\left(e^{\frac{k^2\sigma^2\Delta T}{2}}(k^2\sigma^2\Delta T - 2) + 2\right). \end{aligned}$$

In particular, we have the identity

$$\frac{1}{k} + \frac{b(\theta)}{a(\theta)} = \theta\tanh(k\theta) + \frac{1}{k}\left(\frac{k^2\sigma^2\Delta T - 1 + e^{-\frac{k^2\sigma^2\Delta T}{2}}}{1 - e^{-\frac{k^2\sigma^2\Delta T}{2}}}\right) > 0$$

Proof. Two types of integrals will appear in our computations:

$$I_{\Delta T}(\alpha) = \int_0^{\Delta T} \mathbb{E}\left[e^{\alpha W_t}\right] dt \quad \text{and} \quad J_{\Delta T}(\alpha) = \int_0^{\Delta T} \mathbb{E}\left[W_t e^{\alpha W_t}\right] dt.$$

Let us first compute $I_T(\alpha)$.

$$\begin{aligned} I_{\Delta T}(\alpha) &= \int_0^{\Delta T} \mathbb{E}\left[e^{\alpha W_t}\right] dt \\ &= \int_0^{\Delta T} e^{\frac{\alpha^2 t}{2}} dt \\ &= \frac{2}{\alpha^2}\left(e^{\frac{\alpha^2\Delta T}{2}} - 1\right). \end{aligned}$$

In order to compute $J_{\Delta T}(\alpha)$, we realise that $J_{\Delta T}(\alpha) = I'_{\Delta T}(\alpha)$, this implies:

$$J_{\Delta T}(\alpha) = \frac{2}{\alpha^3}\left(e^{\frac{\alpha^2\Delta T}{2}}(\alpha^2\Delta T - 2) + 2\right).$$

Three other identities will be used:

- If W is a Brownian motion, W and $-W$ have same laws, this yields:

$$I_{\Delta T}(-\alpha) = I_{\Delta T}(\alpha).$$

- By the same token:

$$J_{\Delta T}(-\alpha) = -J_{\Delta T}(\alpha).$$

- Finally, because W has independent increments:

$$\int_0^{\Delta T} \mathbb{E}\left[W_{\Delta T} f(W_t)\right] = \int_0^{\Delta T} \mathbb{E}\left[W_t f(W_t)\right], \quad \forall f \in \mathcal{C}([0, \Delta T])$$

Using the latter, we obtain:

$$\begin{aligned} a(\theta) &= \mathbb{E}\left[\int_0^{\Delta T} \cosh(k(\theta - \sigma W_t))dt\right] \\ &= \frac{1}{2}\int_0^{\Delta T} \mathbb{E}\left[e^{k\theta}e^{-k\sigma W_t} + e^{-k\theta}e^{+k\sigma W_t}\right] dt \\ &= \frac{e^{k\theta} I_{\Delta T}(-k\sigma) + e^{-k\theta} I_{\Delta T}(k\sigma)}{2} \\ &= \cosh(k\theta) I_{\Delta T}(k\sigma). \end{aligned}$$

and

$$\begin{aligned}
b(\theta) &= \mathbb{E}\left[(\theta - \sigma W_{\Delta T})\int_0^{\Delta T} \sinh(k(\theta - \sigma W_t))dt\right] \\
&= \theta\mathbb{E}\left[\int_0^{\Delta T} \sinh(k(\theta - \sigma W_t))dt\right] - \sigma\mathbb{E}\left[W_T\int_0^{\Delta T} \sinh(k(\theta - \sigma W_t))dt\right] \\
&= \theta k^{-1}a'(\theta) - \sigma\mathbb{E}\left[\int_0^{\Delta T} W_t \sinh(k(\theta - \sigma W_t))dt\right] \\
&= \theta \sinh(k\theta)I_{\Delta T}(k\sigma) - \sigma\left(\frac{e^{k\theta}}{2}J_{\Delta T}(-k\sigma) - \frac{e^{-k\theta}}{2}J_{\Delta T}(k\sigma)\right) \\
&= \theta \sinh(k\theta)I_{\Delta T}(k\sigma) + \sigma\cosh(k\theta)J_{\Delta T}(k\sigma)
\end{aligned}$$

By replacing the values of $I_{\Delta T}$ and $J_{\Delta T}$ we end the proof. □

Proposition 5. *(Brownian motion with trend) If $Y_t = \mu t + \sigma W_t$, we have*

$$\begin{aligned}
a(\theta) &= \int_0^{\Delta T} e^{\frac{k^2\sigma^2 t}{2}} \cosh(k\theta - k\mu t)dt \\
b(\theta) &= (\theta - \mu\Delta T)\int_0^{\Delta T} e^{\frac{k^2\sigma^2 t}{2}} \sinh(k\theta - k\mu t)dt \\
&- \frac{\sigma^2}{k}\int_0^{\Delta T} te^{\frac{k^2\sigma^2 t}{2}} \cosh(k\theta - k\mu t)dt.
\end{aligned}$$

Proof. Same reasoning as in the no-trend situation. □

A.2 Proof of existence and uniqueness

Before starting the proof of existence and uniqueness of the maximum for our target function (i.e. Theorem 2). We introduce the following concept which will be key in the proof.

A.2.1 Functional co-monotony

The functional co-monotony principle (see Pagès [**?**]) is the extension of the classical co-monotony principle for real-valued variables, to some stochastic processes such as Brownian diffusion processes, Processes with independent increments, etc. The classic co-monotony principle is stated as follows:

Proposition 6. *Let X be a real-valued random variable and $f, g : \mathbb{R} \to \mathbb{R}$ two monotone functions sharing the same monotony property. Then, if $f(X)$, $g(X)$ and $f(X)g(X)$ are in $L^1(\mathbb{P})$, the following inequality holds:*

$$\mathbb{E}\left[f(X)g(X)\right] \geq \mathbb{E}\left[f(X)\right]\mathbb{E}\left[g(X)\right]. \tag{39}$$

In order to extend this idea to functional of stochastic processes, we should first define an order relation between them. In that sense, we will consider that processes are random variables taking values in a path vector subspace, and define a (partial) order by saying that if α and β are two processes, then:

$$\alpha \leq \beta \quad \text{if} \quad \forall t \in [0, \Delta T], \alpha(t) \leq \beta(t).$$

Hence, we say that a functional is monotone if it is non-decreasing or non-increasing with the order relation defined above, and we will say that two functionals are co-monotone if they share the same monotony.

We can state now the functional co-monotony principle that will be useful to prove various inequalities:

Theorem 7. *(Functional co-monotonicity principle) If X is a pathwise-continuous Markov process, with a monotony preserving transition probabilities*[2]*, and F and G are two co-monotone functionals, continuous on $\mathcal{C}([0,T],\mathbb{R})$, then:*

$$\mathbb{E}\left[F(X)G(X)\right] \geq \mathbb{E}[F(X)]\mathbb{E}[G(X)]. \tag{40}$$

Proposition 7. *Under suitable conditions for $(Y)_{t\in[0,T]}$ we have the following identity:*

$$\mathbb{E}\left[(\theta - Y_{\Delta T})\int_0^{\Delta T}\sinh(k(\theta - Y_t))dt\right] \geq (\theta - \mathbb{E}\left[Y_{\Delta T}\right])\,\mathbb{E}\left[\int_0^{\Delta T}\sinh(k(\theta - Y_t))dt\right].$$

Proof. This comes from the fact that the two functionals involved in the expectation: $F(X) = X_{\Delta T}$, $G(X) = \int_0^{\Delta T}\sinh(kX_t)dt$, have same monotony. □

A.2.2 Existence and uniqueness of the maximum

We prove Theorem 2 by the following sequence of technical propositions.

Proof of Theorem 2:

For the following propositions we will define

$$a(\theta) = \mathbb{E}\left[\int_0^{\Delta T}\cosh(k(\theta - Y_t))dt\right] \tag{41}$$

$$b(\theta) = \mathbb{E}\left[(\theta - Y_T)\int_0^{\Delta T}\sinh(k(\theta - Y_t))dt\right] \tag{42}$$

$$c(\theta) = \mathbb{E}\left[\int_0^{\Delta T}\sinh(k(\theta - Y_t))dt\right] \tag{43}$$

$$d(\theta) = \mathbb{E}\left[(\theta - Y_T)\int_0^{\Delta T}\cosh(k(\theta - Y_t))dt\right] \tag{44}$$

Proposition 8. *The following assertions are true:*

1. *For every $\theta \in \mathbb{R}$, we have $a(\theta) > |c(\theta)|$.*
2. *$c(\theta)$ is strictly increasing, with limits*

$$\lim_{\theta\to+\infty} c(\theta) = +\infty \quad \text{and} \quad \lim_{\theta\to-\infty} c(\theta) = -\infty.$$

3. *It exists a positive constant C satisfying*

$$b(\theta) > 0, \qquad \forall\theta \text{ such as } |\theta| > C.$$

4. *It exists two real constants C_- and C_+ such as*

$$c(\theta)b(\theta) \leq a(\theta)b(\theta), \qquad \forall\theta \geq C_+$$

and

$$c(\theta)b(\theta) \geq -a(\theta)b(\theta), \qquad \forall\theta \leq C_-.$$

Proof. 1. It follows directly from Jensen inequality and that for every $x \in \mathbb{R}$ we have $\cosh(x) \geq |\sinh(x)|$.

2. It follows directly from the fact that $c'(\theta) \geq k\Delta T > 0$, for every $\theta \in \mathbb{R}$.

[2]In particular, this is the case for diffusions.

3. From the functional co-monotony principle we have

$$b(\theta) \geq (\theta - \mathbb{E}[Y_T])c(\theta).$$

   Then, from a direct application of the second result of this proposition, we have that it exists $C' > 0$ such as, for $\theta > C$, both factors of the left are positive. Similarly, it exists $C'' > 0$ such as, for $\theta < -C''$, both factors of the left are negative. We conclude our result by taking $C = \max(C', C'')$.

4. By a similar argument as in the last point, it exists C_+ such as, for $\theta \geq C_+$ both $c(\theta)$ and $b(\theta)$ are positive, so we can use the first result in this proposition which yields to $c(\theta)b(\theta) \leq a(\theta)b(\theta)$.

   The second inequality follows in the same way. We know it exists C_- such as, for $\theta \leq C_-$ we have $c(\theta)$ negative and $b(\theta)$ positive, so we can use the first result in this proposition which yields to $c(\theta)b(\theta) \geq -a(\theta)b(\theta)$.

□

Proposition 9. *Let us define the function*

$$f(\theta) = b(\theta)c(\theta) - a(\theta)d(\theta).$$

The following assertions are true

1. $f(\cdot)$ *is strictly decreasing.*
2. $\lim_{\theta\to+\infty} f(\theta) < 0$.
3. $\lim_{\theta\to-\infty} f(\theta) > 0$.

In particular, this implies that $f(\theta) = 0$ has a unique solution.

Proof. 1. First of all, observe that $a'(\theta) = kc(\theta)$, $c'(\theta) = ka(\theta)$, $b'(\theta) = c(\theta) + kd(\theta)$ and $d'(\theta) = a(\theta) + kb(\theta)$. This leads to

$$f'(\theta) = c^2(\theta) - a^2(\theta)$$

   which is strictly negative, due to the first result in the precedent proposition.

   Hence, f is strictly decreasing.

2. From the last result in the precedent proposition we have that if $\theta > C_+$ then

$$f(\theta) = b(\theta)c(\theta) - a(\theta)d(\theta) \leq a(\theta)(b(\theta) - d(\theta)).$$

   Otherwise said

$$f(\theta) \leq -a(\theta)\mathbb{E}\left[(\theta - Y_T)\int_0^{\Delta T} e^{-k(\theta-Y_t)}dt\right] = -a(\theta)e^{-k\theta}\mathbb{E}\left[(\theta - Y_T)\int_0^{\Delta T} e^{kY_t}dt\right]$$

   where the right side is negative for θ large enough.

3. From the last result in the precedent proposition we have that if $\theta < C_-$ then

$$f(\theta) = b(\theta)c(\theta) - a(\theta)d(\theta) \geq -a(\theta)(b(\theta) + d(\theta)).$$

   Otherwise said

$$f(\theta) \geq -a(\theta)\mathbb{E}\left[(\theta - Y_T)\int_0^{\Delta T} e^{k(\theta-Y_t)}dt\right] = -a(\theta)e^{k\theta}\mathbb{E}\left[(\theta - Y_T)\int_0^{\Delta T} e^{-kY_t}dt\right]$$

   In this case, as $-\theta$ becomes large enough, the right side becomes positive.

□

Proposition 10. *The function*

$$g(\theta) = a(\theta) \exp\left(-k\frac{b(\theta)}{a(\theta)}\right)$$

has a unique maximum.

Proof. We have

$$g'(\theta) = k \exp\left(-k\frac{b(\theta)}{a(\theta)}\right)\left(a'(\theta)\left(\frac{1}{k} + \frac{b(\theta)}{a(\theta)}\right) - b'(\theta)\right),$$

which is equivalent to

$$g'(\theta) = k^2 \exp\left(-k\frac{b(\theta)}{a(\theta)}\right)\left(\frac{c(\theta)b(\theta) - d(\theta)a(\theta)}{a(\theta)}\right).$$

Thus, the sign of $g'(\theta)$ is the same as the sign of $f(\theta)$ from the last proposition. Hence, $g'(\theta)$ is strictly decreasing and has a unique point where it becomes zero. Otherwise said, g is strictly concave and has a unique maximum. □

Proposition 11. *The function $\pi(\psi, \theta) = e^{-k\psi}(\psi a(\theta) - b(\theta))$ has an unique maximum.*

Proof. First of all, if we fix θ. The maximum is achieved in

$$\psi^*(\theta) = \frac{1}{k} + \frac{b(\theta)}{a(\theta)}.$$

The value of this maximum is given by

$$\pi(\psi^*(\theta), \theta) = k^{-1}e^{-1}e^{-k\frac{b(\theta)}{a(\theta)}}a(\theta)$$

By the precedent proposition, it exists a unique θ^* maximum of this function.
Thus, the maximum of $\pi(\psi, \theta)$ is achieved at the point $(\psi^*(\theta^*), \theta^*)$.

□

A.3 Convergence of the stochastic algorithm

Before proving the convergence of the stochastic algorithm, we recall the hypothesis of the Robbins-Monro theorem (see [12]) which is central to complete the proof.

A.3.1 The Robbins-Monro theorem

Let us consider an algorithm of the form

$$\delta_{n+1} = \delta_n + \gamma_{n+1}H(\delta_n, Y_{n+1}), \tag{45}$$

with $(Y_n)_{n\in\mathbb{N}}$ an i.i.d. sequence of ν-distributed $\mathbb{R}^q$-valued random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

In order to prove our main result, let us consider a random vector Y taking values in $\mathbb{R}^q$ with distribution ν and a Borel function $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$.
Following what precedes, we introduce into our analysis the following function:

$$h : \delta \mapsto \mathbb{E}\left[H(\delta, Y)\right]. \tag{46}$$

And, for this function to be well defined, we add the following condition:

$$\forall \delta \in \mathbb{R}^d, \quad \mathbb{E}\left[|H(\delta, Y)|\right] < +\infty. \tag{47}$$

Theorem 8. *Assume that the mean function h of the algorithm is continuous and satisfies*

$$\forall y \in \mathbb{R}^d, \delta \neq \delta^*, \quad \langle \delta - \delta^*, h(\delta) \rangle < 0 \tag{48}$$

(which implies that $\{h = 0\} = \{\delta^*\}$). Suppose furthermore that $Y_0 \in L^2(\mathbb{P})$ and that H satisfies

$$\forall \delta \in \mathbb{R}^d, \quad \mathbb{E}\left[\|H(\delta, Y)\|^2\right] \leq C(1 + \|\delta\|^2) \tag{49}$$

If the step sequence γ_n satisfies $\sum_n \gamma_n = +\infty$ and $\sum_n \gamma_n^2 < +\infty$, then:

$$\delta_n \to \delta^*, \quad \mathbb{P} - a.s. \tag{50}$$

and in every $L^p(\mathbb{P})$, $p \in (0, 2]$.

A.3.2 Computing the local gradient

Proof of Proposition 1:

Proof. By direct calculation we have:

$$\begin{aligned}
\frac{\partial}{\partial \psi} \lambda(\psi, \theta; y_t) &= -k\lambda(\psi, \theta; y_t), \\
\frac{\partial}{\partial \psi} \mu(\psi, \theta; y_t) &= -k\mu(\psi, \theta; y_t), \\
\frac{\partial}{\partial \theta} \lambda(\psi, \theta; y_t) &= k\mu(\psi, \theta; y_t), \\
\frac{\partial}{\partial \theta} \mu(\psi, \theta; y_t) &= k\lambda(\psi, \theta; y_t).
\end{aligned}$$

which applied to

$$\Pi(\psi, \theta; Y) = \psi\lambda(\psi, \theta; Y) - (\theta - Y_{\Delta T})\mu(\psi, \theta; Y).$$

lead directly to the result we want. □

A.3.3 Convergence of the stochastic algorithm

Proof of Theorem 3:

Proof. The main idea here is to apply the Robbins-Monro algorithm. However, as it was said, we cannot apply it directly on the local gradient $\nabla_{\psi,\theta}\Pi(\psi_n, \theta_n, Y)$, because of its behavior for larger values of θ. The latter makes impossible to obtain the Robbins-Monro condition

$$\mathbb{E}\left[\|\nabla_{\psi,\theta}\Pi(\psi_n, \theta_n, Y)\|^2\right] \leq C(1 + \psi^2 + \theta^2).$$

However, by multiplying by the factor $\rho(\psi, \theta) = e^{-k(|\theta| - \psi)}$, which does not impact the other Robbins-Monro conditions, does not changes the point to which the algorithm converges and allow us to retrieve the inequality we are looking for. Moreover, the choice of $\rho(\psi, \theta)$ is rather intuitive; it is based on the observation that the intensities λ and μ have terms with magnitudes $\cosh(k\theta)$ and $\sinh(k\theta)$ multiplied by a factor $e^{-k\psi}$, so our choice of $\rho(\psi, \theta)$ is quite natural.

Because the function in the algorithm corresponds to the gradient of a well-behaved function reaching a unique maximum, the only Robbins-Monro condition it remains to prove in order to obtain the convergence of the algorithm, is to show it exists a real constant C such as

$$\mathbb{E}\left[\|\rho(\psi, \theta)\nabla_{\psi,\theta}\Pi(\psi, \theta, Y)\|^2\right] \leq C(1 + \psi^2 + \theta^2).$$

By straightforward computation and regrouping terms, we have

$$\begin{aligned}
\|\nabla_{\psi,\theta}\Pi(\psi,\theta,Y)\|^2 &= \|(\theta - Y_T)\nabla\mu - \psi\nabla\lambda\|^2 + \lambda^2 + \mu^2 \\
&+ 2\psi(\lambda\partial_\psi\lambda - \mu\partial_\theta\lambda) + 2(\theta - Y_T)(\mu\partial_\theta\mu - \lambda\partial_\psi\mu) \\
&= \|(\theta - Y_T)\nabla\mu - \psi\nabla\lambda\|^2 + (\lambda^2+\mu^2)(1-2k\psi) + 4k\lambda\mu(\theta - Y_T) \\
&\le k^2(|\theta - Y_T|^2 + \psi^2)(\lambda^2+\mu^2) + (\lambda^2+\mu^2)(1-2k\psi) + 4k\lambda\mu(\theta - Y_T) \\
&= (\lambda^2+\mu^2)(k^2|\theta - Y_T|^2 + (1-k\psi)^2) + 4k\lambda\mu(\theta - Y_T) \\
&\le 2\lambda^2(k^2|\theta - Y_T|^2 + (1-k\psi)^2 + 4k(\theta - Y_T)) \\
&\le 2\lambda^2(3k^2|\theta - Y_T|^2 + (1-k\psi)^2 + 1) \\
&\le 12k^2\lambda^2(\theta^2 + \psi^2 + k^{-2} + Y_T^2)
\end{aligned}$$

On the other hand, by Jensen inequality we have

$$\lambda^2 \le 4A^2 e^{-2k\psi}\int_0^{\Delta T}\cosh^2(k(\theta - Y_t))dt \le 4A^2 e^{-2k\psi}\int_0^{\Delta T} e^{2k|\theta|}e^{2k|Y_t|}dt.$$

This leads to

$$\|\nabla_{\psi,\theta}\Pi(\psi,\theta,Y)\|^2 \le 48k^2A^2e^{2k(|\theta|-\psi)}(\theta^2+\psi^2+k^{-2}+Y_T^2)\int_0^{\Delta T} e^{2k|Y_t|}dt \tag{51}$$

which is equivalent to

$$\begin{aligned}
\mathbb{E}\left[\|\rho(\psi,\theta)\nabla_{\psi,\theta}\Pi(\psi,\theta,Y)\|^2\right] &\le 48k^2A^2\mathbb{E}\left[(\theta^2+\psi^2+k^{-2}+Y_T^2)\int_0^{\Delta T} e^{2k|Y_t|}dt\right] &(52)\\
&\le C(1+\theta^2+\psi^2). &(53)
\end{aligned}$$

with

$$C = 48k^2A^2\mathbb{E}\left[(1+k^{-2}+Y_T^2)\int_0^{\Delta T} e^{2k|Y_t|}dt\right]$$

which is bounded by hypothesis.

Hence the Robbins-Monro theorem can apply, which concludes the proof.

□

A.3.4 Dimensionality reduction for the innovation

Proof of Theorem 4:

Proof. Besides the end-value Y_T, all the dependency in the process Y is contained in the functions λ and μ (given by the equations (12) and (13)). To obtain our result, let us rewrite the following integral:

$$\int_0^T e^{k(\theta - Y_t)}dt = e^{k\theta}\sqrt{\frac{\int_0^T e^{-kY_t}dt}{\int_0^T e^{kY_t}dt}}\sqrt{\int_0^T e^{-kY_t}dt\int_0^T e^{kY_t}dt} = e^{k\theta+\rho_k}e^{b_k}. \tag{54}$$

By the same argument, we have

$$\int_0^T e^{-k(\theta - Y_t)}dt = e^{k\theta}\sqrt{\frac{\int_0^T e^{kY_t}dt}{\int_0^T e^{-kY_t}dt}}\sqrt{\int_0^T e^{-kY_t}dt\int_0^T e^{kY_t}dt} = e^{-k\theta-\rho_k}e^{b_k}. \tag{55}$$

The results follows by combining these quantities to obtain $\sinh(\cdot)$ and $\cosh(\cdot)$ as in the formulas for λ and μ ((12) and (13)). □

A.4 Penalizing the inventory

A.4.1 Derivatives for the penalization function

Proof of Proposition 2:

Proof. Conditional to λ and μ, the random variable Q is the difference of two independent Poisson random-variables N_b and N_a representing the liquidity capture at the bid and at the ask respectively. The variable N, on the other hand, represent the sum of these two variables. Because of the independence between N_b and N_a we know that conditional to N, the variable Q satisfies:

$$\mathbb{E}\left[\Phi(Q)|N\right] = \mathbb{E}\left[\Phi(2B-N)|N\right], \quad B \sim \mathrm{Bin}\left(N, \frac{\mu+\lambda}{2\lambda}\right).$$

Using that N is a Poisson variable with intensity λ we can write

$$\mathbb{E}\left[\Phi(Q)\right] = \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{e^{-\lambda}(\lambda+\mu)^k(\lambda-\mu)^{n-k}}{2^n k!(n-k)!}\Phi(2k-n) \tag{56}$$

$$= \sum_{n=0}^{\infty}\sum_{k=0}^{n+1} \frac{e^{-\lambda}(\lambda+\mu)^k(\lambda-\mu)^{n+1-k}}{2^{n+1} k!(n+1-k)!}\Phi(2k-n-1) \tag{57}$$

$$= \sum_{n=0}^{\infty}\sum_{k=0}^{n+1} \frac{e^{-\lambda}(\lambda-\mu)^k(\lambda+\mu)^{n+1-k}}{2^{n+1} k!(n+1-k)!}\Phi(2k-n+1) \tag{58}$$

The identity (57) arises from the hypothesis $\Phi(0) = 0$. The identity (58) arises just by inverting the order of summation.

Let us note $a = \lambda + \mu$ and $b = \lambda - \mu$, from (57) we obtain

$$\partial_a \mathbb{E}\left[\Phi(Q)\right] = \frac{1}{2}\mathbb{E}\left[\Phi(Q-1)\right]$$

and from (58)

$$\partial_b \mathbb{E}\left[\Phi(Q)\right] = \frac{1}{2}\mathbb{E}\left[\Phi(Q+1)\right].$$

This leads to the following formulas

$$\partial_\lambda \mathbb{E}\left[\Phi(Q)\right] = \frac{1}{2}\mathbb{E}\left[\Phi(Q+1) - \Phi(Q) + \Phi(Q-1)\right] \tag{59}$$

$$\partial_\mu \mathbb{E}\left[\Phi(Q)\right] = \frac{1}{2}\mathbb{E}\left[\Phi(Q-1) - \Phi(Q+1)\right]. \tag{60}$$

□

A.4.2 Closed-formulas for the penalization function

Proof of Theorem 5:

Proof. First of all, the idea is to exploit the following two facts:

1. The dependency of $\mathbb{E}_Y\left[\Phi(Q)\right]$ in the variables ψ and θ is totally contained on the functions λ and μ.
2. Under the information provided by the process Y, the inventory is the difference of two independent Poisson variables N^+ and N^- satisfying the following equations:

$$\mathbb{E}[N^- + N^+] = \lambda \tag{61}$$

$$\mathbb{E}[N^- - N^+] = \mu \tag{62}$$

This implies that we just need to do the analysis considering λ and μ as constants; the results can be immediately transferred to the case of the conditional expectation of $\Phi(Q)$ under the information provided by Y.

Now, let us consider a auxiliary Markov process $Q_s = N_s^{(1)} - N_s^{(2)}$ where $N_s^{(1)}$ and $N_s^{(2)}$ are two independent homogeneous Poisson processes with intensities $\lambda_1 = \frac{1+\varepsilon}{2}$ and $\lambda_2 = \frac{1-\varepsilon}{2}$ respectively. It is immediate that Q and Q_λ have the same law, thus, for any function $\Phi(\cdot)$, the quantities $\mathbb{E}\left[\Phi(Q_\lambda)\right]$ and $\mathbb{E}\left[\Phi(Q)\right]$ have the same value. To compute $\mathbb{E}\left[\Phi(Q_\lambda)\right]$ is done by using the infinitesimal generator and Dynkin's formula [**?**].

For $q \in \mathbb{Z}$ the infinitesimal generator of Q_s is

$$\mathcal{A}\Phi(q) = -\Phi(q) + \left(\frac{1+\varepsilon}{2}\right)\Phi(q+1) + \left(\frac{1-\varepsilon}{2}\right)\Phi(q-1)$$

By the Dynkin's formula we obtain:

$$\mathbb{E}\left[\Phi(Q)\right] = \Phi(0) + \int_0^\lambda \mathbb{E}\left[D^2\Phi(Q_t)\right]dt + \varepsilon\int_0^\lambda \mathbb{E}\left[D\Phi(Q_t)\right]dt$$

Where

$$D^2\Phi(q) = \frac{\Phi(q+1) - 2\Phi(q) + \Phi(q-1)}{2}$$
$$D\Phi(q) = \frac{\Phi(q+1) - \Phi(q-1)}{2}$$

The next step is to evaluate expectations of random variables that are defined as the difference between two Poisson random variables. Indeed, these kind of integer valued random variable are said to follow a Skellam distribution with parameters $\lambda_1, \lambda_2 > 0$ if they can be written as the difference of two independent Poisson random variables, N_1 and N_2, with respective parameters λ_1 and λ_2.

Let Q_s follows a Skellam distribution (see [**?**]) with mean εs and variance s; the support of Q_s is the whole set $\mathbb{Z}$. Elementary computations show that the distribution of the variable Q_s is defined by the following formula:

$$\mathbb{P}\left(Q = |q|\right) = 2e^{-s}\cosh\left(\frac{q}{2}\log\left(\frac{1+\varepsilon}{1-\varepsilon}\right)\right)I_{|q|}\left(s\sqrt{1-\varepsilon^2}\right) \quad , \quad \forall q \in \mathbb{Z}\setminus\{0\} \tag{63}$$
$$\mathbb{P}\left(Q = 0\right) = e^{-s}I_0\left(s\sqrt{1-\varepsilon^2}\right) \tag{64}$$

At this point we just need to replace these values in the equations obtained through the Dynkin's formula and using the proprerties of function Φ (these are, even, increasing in $\mathbb{R}^+$ and $\Phi(0) = 0$ (in particular $D\Phi(q) = -D\Phi(-q)$).

Finally, the equation for $\mathbb{E}[|Q|]$ is obtained by replacing $\Phi(\cdot)$ by $|\cdot|$.

The formula for $\mathbb{E}[Q^2]$ can be obtained by the same idea or by more elementary computations using the properties of the mean and variance of Poisson random variables. $\square$

A.4.3 Other properties of the penalization function

Proof of Proposition 4:

Proof. Because of the symmetry of the function $\Phi(\cdot)$ we can always suppose $p \leq 1/2$. In particular, we can consider a binomial random variable $\widetilde{B}$ with parameters $(N, 1/2)$ which stochastically dominates B. This is easy to explicitly generate:

$$B = \sum_{k=1}^{N} \mathbf{1}_{U_k \leq p} \leq \sum_{k=1}^{N} \mathbf{1}_{U_k \leq \frac{1}{2}} = \widetilde{B}$$

We obtain the following:

$$\mathbb{E}(|2B - N|) = 2\mathbb{E}\left(\left|B - \frac{N}{2}\right|\right) \geq 2\mathbb{E}\left(\left|B - \frac{N}{2}\right| \mathbf{1}_{\widetilde{B} \leq \frac{N}{2}}\right)$$

By stochastic domination and symmetry of $\widetilde{B}$ we obtain:

$$\mathbb{E}(|2B - N|) \geq 2\mathbb{E}\left(\left|\widetilde{B} - \frac{N}{2}\right| \mathbf{1}_{\widetilde{B} \leq \frac{N}{2}}\right) = \mathbb{E}\left(\left|\widetilde{B} - \frac{N}{2}\right|\right)$$

Moreover, we can use the following De Moivre's result for the absolute deviation of the Binomial distribution[**?**]:

$$\mathbb{E}\left(\left|\widetilde{B} - \frac{N}{2}\right|\right) = 2^{-N}\binom{N}{\lceil N/2 \rceil}\lceil N/2 \rceil$$

At this point, we consider the following elementary inequality:

$$2^{-N}\binom{N}{\lceil N/2 \rceil}\lceil N/2 \rceil \geq \frac{\sqrt{N}}{2\sqrt{2}}$$

Now let us take the expectation on N.

$$\mathbb{E}(\sqrt{N}) \geq \sum_{k=1}^{\infty} \sqrt{k} e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \mathbb{E}\left(\frac{1}{\sqrt{N+1}}\right) \geq \frac{\lambda}{\sqrt{\lambda + 1}}$$

(The last inequality comes from Jensen inequality) □

Proof of Proposition 5:

Proof. First of all, we have

$$\partial_\psi \pi(\psi, \theta) = \mathbb{E}\left[\lambda - k\psi\lambda - (\theta - Y_T)\mu + k\sum_{j=1}^{J} j\lambda^{j-1} g_j\left(\frac{\mu}{\lambda}\right)\right] \quad (65)$$

We are interested in the solution of $\partial_\psi \pi(\psi, \theta; Y) = 0$:

$$Ae^{-k\psi}(k\psi - 1)a(\theta) + Ae^{-k\psi}b(\theta) - k\sum_{j=1}^{J} je^{-k(j-1)\psi} c_j(\theta) = 0 \quad (66)$$

Here $c_j(\theta) = \mathbb{E}\left(A^{j-1} g_j\left(\frac{\mu}{\lambda}\right)\right)$.

We can divide by $e^{-k\psi}$ an re-arrange terms, this leads to

$$k\psi a(\theta) + (b(\theta) - a(\theta)) = c_1(\theta) + k\sum_{j=1}^{J-1}(j+1)A^j e^{-kj\psi} c_{j+1}(\theta) \quad (67)$$

The left side is an increasing linear function starting from a negative point, the left side a decreasing exponential starting from a positive point, so they intersect at some point. □

References

[1] Avellaneda, M., & Stoikov, S. (2008). High-frequency trading in a limit order book. Quantitative Finance, 8(3), 217-224.

[2] Bayraktar, E., & Ludkovski, M. (2014). Liquidation in limit order books with controlled intensity. Mathematical Finance, 24(4), 627-650.

[3] Benveniste, A., Métivier, M., & Priouret, P. (2012). Adaptive algorithms and stochastic approximations. Springer Publishing Company, Incorporated.

[4] Cartea, A., Jaimungal, S., & Ricci, J. (2014). Buy low, sell high: A high frequency trading perspective. SIAM Journal on Financial Mathematics, 5(1), 415-444.

[5] Duflo, M. (1997). Random iterative models (Vol. 22). Berlin: Springer.

[6] Fernandez-Tapia, J. (2015). Modeling, optimization and estimation for the on-line control of trading algorithms in limit-order markets. PhD Thesis. Laboratoire de Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie.

[7] Guéant, O., Lehalle, C. A., & Fernandez-Tapia, J. (2012). Optimal portfolio liquidation with limit orders. SIAM Journal on Financial Mathematics, 3(1), 740-764.

[8] Guéant, O., Lehalle, C. A., & Fernandez-Tapia, J. (2013). Dealing with the inventory risk: a solution to the market making problem. Mathematics and financial economics, 7(4), 477-507.

[9] Laruelle, S., Lehalle, C. A., & Pagès, G. (2011). Optimal split of orders across liquidity pools: a stochastic algorithm approach. SIAM Journal on Financial Mathematics, 2(1), 1042-1076.

[10] Laruelle, S., Lehalle, C. A., & Pagès, G. (2013). Optimal posting price of limit orders: learning by trading. Mathematics and Financial Economics, 7(3), 359-403.

[11] Lemaire, V., & Pagès, G. (2010). Unconstrained recursive importance sampling. The Annals of Applied Probability, 20(3), 1029-1067.

[12] Pagès, G. (2014) Introduction to Numerical Probability applied to finance. Technical Report.