Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

A Symbolic Closed-Form Solution to Sequential Market Making with Inventory

Shamin Kinathil,[1,2] Scott Sanner,[3] Sanmay Das,[4] Nicolás Della Penna[1]
[1]Australian National University, [2]Data61, [3]University of Toronto, [4]Washington University in St. Louis
{shamin.kinathil, nicolas.della-penna}@anu.edu.au, ssanner@mie.utoronto.ca, sanmay@wustl.edu

Abstract

Market-makers serve an important role as providers of liquidity and order in financial markets, particularly during periods of high volatility. Optimal market-makers solve a sequential decision making problem, where they face an exploration versus exploitation dilemma at each time step. A belief state MDP based solution was presented by Das and Magdon-Ismail [2008]. This solution however, was closely tied to the choice of a Gaussian belief state prior and did not take asset inventory into consideration when calculating an optimal policy. In this work we introduce a novel continuous state POMDP framework which is the first to solve, exactly and in closed-form, the optimal market-making problem with inventory, fixed asset value, arbitrary belief state priors, trader models and reward functions via symbolic dynamic programming. We use this novel model and solution to show that sequentially optimal policies are heavily inventory-dependent and calculate policies that operate with bounded loss guarantees under a variety of market models and conditions.

1 Introduction

Financial markets are well known for their volatility [Mandelbrot, 1963; Sornette, 2004; Volt, 2005]. A prime driver of high volatility are periods of great uncertainty, such as those following price shocks, about the value of the assets being traded. These periods can lead to sparse trading, where there are few offers and counter-offers, which effectively stop markets from functioning. Market-makers (MMs) can be used in these situations to attract trading and maintain order in markets. Examples of markets which employ MMs include the NASDAQ, which allows for multiple MMs to compete, and prediction markets [Hanson, 2003; 2007].

Market-makers do not perform their services *ex gratia*. They aim to extract a profit, which is a component of what is known as the bid-ask spread. When setting the bid and ask price, MMs must trade-off between setting prices to extract maximum profit from the next trade versus setting prices to get as much information about the new value of the asset so as to generate larger profits from future trades.

In this paper we recast the sequential market-making problem as a new general class of *continuous state* Partially Observable MDPs (POMDPs) [Kaelbling *et al.*, 1998] that subsumes our proposed market-making model and admits closed-form solutions. We also provide a Symbolic Dynamic Programming (SDP) [Boutilier *et al.*, 2001] based algorithm that can efficiently derive a value function and policy for this new class of POMDPs.

By utilising the POMDP framework we enable several new advances in algorithmic market-making:

- We are the first to address the open question of sequentially optimal market-making with *inventory based reasoning* by extending the seminal work of Glosten and Milgrom [1985], Amihud and Mendelson [1980] and Das and Magdon-Ismail [2008].
- Our solution is more general than previous solutions and allows the use of arbitrary belief state priors to represent the MM's uncertainty about the true value of the asset, asymmetric bid-ask spreads around the mean of the MM's belief, flexible models of trader behaviour and a variety of inventory-based rewards with both hard constraints and preferences.
- Our solution provides sequentially optimal policies which maximise MM profit under a variety of inventory conditions and trader models, while simultaneously bounding loss. Previous solutions which neglected inventory are prone to selecting sub-optimal bid-ask spreads which may lead to *unexpected losses*.

2 Background

Market-makers serve an important liquidity provision role in many financial markets by providing immediacy to transactions. In practice MMs often operate in tandem with a continuous double auction [Smith *et al.*, 2003] market structure in which both buyers and sellers submit limit orders in the form of bids and asks which are then matched algorithmically. Theoretical models of market-making typically abstract away limit orders from other traders and model the market as a *dealer market*, with the MM taking one side of every order. In these models it is common for the MM to quote a

bid price, a price at which they are willing to buy some number of shares, and an ask price, the price at which they are willing to sell some number of shares, at every point in time. The difference between the two prices, known as the bid-ask spread, serves three main purposes: (1) to provide an incremental profit to motivate the MM to actually provide their services; (2) to compensate the MM for the risk inherent in holding inventory [Amihud and Mendelson, 1980]; and (3) to compensate the MM for the adverse selection encountered in trading with a potentially more informed population [Glosten and Milgrom, 1985].

In recent history, research into algorithmic market-making has focussed on financial [Das, 2008] and prediction market settings [Chen and Pennock, 2007; Othman and Sandholm, 2012]. Market-making in prediction markets starts from the notion of a proper Market Scoring Rule (MSR) [Hanson, 2003]. Proper MSRs are both myopically incentive compatible, that is, they incentivize agents to move prices to their valuation, and are also loss-bounded, at least in bounded-payoff markets. Despite these attractive features, proper MSRs are, in general, loss-making [Pennock and Sami, 2007] and MMs using proper MSRs must be subsidised in order to perform their task of liquidity provision. Although there has been some theoretical work on the design of MMs that can adapt their spread over time [Othman *et al.*, 2010] they can be highly sensitive to the price path [Brahma *et al.*, 2012].

Learning based approaches to market-making in financial markets attempt to infer either underlying values or future prices based on trader models [Das, 2008; Das and Magdon-Ismail, 2008] or time-series analysis of prices [Chakraborty and Kearns, 2011]. Such models have been used to formulate efficient algorithms under zero-profit (competitive) conditions where, in perfect competition, the MM is pushed to setting bid and ask prices that yield a zero expected profit [Das, 2005]. An algorithm for a profit-maximising (monopolist) MM was presented by Das and Magdon-Ismail [2008]. In this setting with a fixed underlying asset value, bid and ask prices are set in a way that maximises the total discounted profit obtained by the MM. The profit-maximising market-making algorithm was cast as a belief state MDP where the MM's belief about the value of the asset was described by a Gaussian density function. The MM sets bid and ask prices based on the solution of the Bellman equation for a Gaussian belief state approximation of this MDP. While these models are promising in terms of their potential to provide liquidity while simultaneously operating at a profit, they come with no guarantees on worst case loss, since they do not take inventory into account when setting prices.

In this paper we study the problem of optimising profit while still maintaining inventory-sensitivity, the natural way to bound risk in the market-making setting. We present a new formalism for optimal sequential market-making based on continuous state POMDPs which allows for arbitrary belief state distributions, asymmetric bid-ask spreads, flexible trader models and a variety of inventory-based rewards, while still admitting exact closed-form solutions.

3 Market-Making as a POMDP

In this section we introduce our first contribution, a novel and flexible continuous state POMDP framework for sequentially optimal market-making with inventory that admits closed-form dynamic programming solutions.

3.1 POMDP Preliminaries

A POMDP is defined by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{O}, \Omega, \mathcal{H}, \gamma \rangle$ [Kaelbling *et al.*, 1998]. $\mathcal{S}$ specifies a potentially infinite (e.g., continuous) set of states while $\mathcal{A}$ and $\mathcal{O}$ respectively specify a finite set of actions and observations. The transition function $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ defines the effect of an action on the state. $\Omega : \mathcal{O} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is the observation function which defines the probability of receiving an observation given that an agent has executed an action and reached a new state. $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function which encodes the preferences of the agent. $\mathcal{H}$ represents the number of decision steps until termination and the discount factor $\gamma \in [0, 1)$ is used to geometrically discount future rewards.

Partial observability is encapsulated within a probabilistic observation model specified by $\mathcal{O}$ and Ω, which relates possible observations to states. POMDP agents maintain a belief $b \in \Delta(\mathcal{S})$, a probability distribution over $\mathcal{S}$. Beliefs provide sufficient statistics for the observable history of a POMDP agent without loss of optimality [Aström, 1965].

The objective of a POMDP agent is to find an optimal policy $\pi^* : \Delta(\mathcal{S}) \rightarrow \mathcal{A}$, which specifies the best action to take in every belief state b so as to maximise the expected sum of discounted rewards over the planning horizon $\mathcal{H}$. The value function under the optimal policy can be written as:

$$V^*(b) = \mathbb{E}_{\pi^*} \left[\sum_{h=0}^{\mathcal{H}} \gamma^h \sum_{s \in \mathcal{S}} \mathcal{R}\left(s, \pi^*(b^h(s))\right) \cdot b^h(s) \middle| b^0 = b \right].$$

The POMDP value function can be parameterized by a finite set of functions linear in the belief representation known as α-vectors and is convex in the belief [Sondik, 1971].

The value function at a given belief $b \in \Delta(\mathcal{S})$ can be calculated using:

$$V^h(b) = \max_{\alpha \in V^h} \sum_{s \in \mathcal{S}} \alpha(s) \cdot b(s).$$

Monahans exact Value Iteration [Monahan, 1982] algorithm has been extended to a more efficient class of point-based algorithms [Pineau *et al.*, 2003] which considers only a finite subset of the belief space $\mathcal{B}$. We cover such methods in Section 4.2. The results from the discrete state setting presented above can be generalised to the continuous case by replacing s, α-vectors and $\sum$ by their continuous state counterparts $\vec{x}$, α-functions and $\int$, respectively [Porta *et al.*, 2006].

In the next section we show how the expressive POMDP framework can be used to encapsulate an optimal sequential market-making model with inventory.

3.2 Optimal Sequential Market-Making Model

The market-making model used in this paper extends the seminal theoretical model of Glosten and Milgrom [1985] and the sequential model of Das and Magdon-Ismail [2008] by incorporating the inventory control conditions of Amihud and Mendelson [1980] as well as arbitrary belief state priors, trader models and asymmetric bid-ask spreads.

We begin by formulating the MM's sequential decision problem in the POMDP framework:

- $\mathcal{S} = \langle v, i \rangle$, where $v \in \mathbb{R}^+$ represents the value of the asset and $i \in \mathbb{N}$ represents the MM's inventory.

- $\mathcal{A} = \{(bid_1, ask_1), \ldots, (bid_N, ask_N)\}$ represents a finite set of allowed N bid-ask pairs, where $bid_n, ask_n \in \mathbb{R}^+$ and $bid_n < ask_n$.

The bid represents the price at which the MM is willing to buy one unit of the asset and the ask is the price at which the MM is willing to sell one unit of the asset.

- $\mathcal{O} = \{buy, sell, hold\}$ represents the possible actions of a trader at each time step.

At each time step the MM interacts with a trader t drawn from a heterogeneous population of traders with a known prior distribution. The trader has an uninformed estimate of the value of the asset $v_t = v + \epsilon_t$, where $\epsilon_t \in \mathbb{R}$ specifies the noise. In this paper we allow the traders to be specified according to two trader models: (1) Glosten-Milgrom (GM) [Glosten and Milgrom, 1985]; and (2) Discrete noise (D). In the traditional GM setting, traders are either $\mathrm{informed}$ or $\mathrm{uninformed}$, with ϵ_t set to 0.0. The Discrete setting involves $\mathrm{informed}$, $\mathrm{over\text{-}valuing}$ and $\mathrm{under\text{-}valuing}$ traders, with ϵ_t set to 0.0, constant $c \in \mathbb{R}$ and constant $-c \in \mathbb{R}$, respectively.

With the exception of $\mathrm{uninformed}$ traders, an arriving trader of type t will execute a buy order at an ask price ask_n if $v_t > ask_n$, a $sell$ order at the bid price bid_n if $v_t < bid_n$, and will $hold$ otherwise. We use an intermediate variable $u \sim \mathbb{P}(\mathcal{O})$, with $\mathbb{P}(u) = 1/3$ for $\mathrm{uninformed}$ trader orders.

- The transition function $\mathcal{T}$ for each state variable in $\mathcal{S}$ is given by:

$$\mathcal{T}\left(i'|i, v, bid, ask, t = \text{informed}, u\right) = \delta\left[i' - \begin{cases} (v > ask + \epsilon_t) \wedge (i \geq 1): & i - 1 \\ (v < bid + \epsilon_t): & i + 1 \\ \text{otherwise}: & i \end{cases}\right]$$

$$\mathcal{T}\left(i'|i, v, bid, ask, t = \text{uninformed}, u\right) = \delta\left[i' - \begin{cases} (u = buy) \wedge (i \geq 1): & i - 1 \\ (u = sell): & i + 1 \\ \text{otherwise}: & i \end{cases}\right]$$

$$\mathcal{T}\left(v'|i, v, bid, ask, t, u\right) = \delta\left[v' - v\right]$$

The Dirac function $\delta[\cdot]$ ensures that the transitions are valid conditional probability functions that integrate to 1.0. The value v is assumed to be fixed but unknown to the MM. The inventory i increments or decrements according to the observed trader action.

- The observation function Ω for each trader action is specified by:

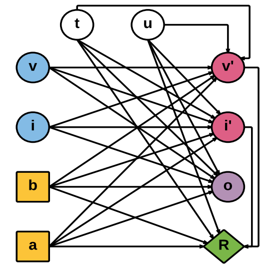


Figure 1: An optimal sequential market-making continuous state POMDP. $\mathcal{S} = \mathtt{v}, \mathtt{i}$. $\mathtt{b}, \mathtt{a} \in \mathcal{A}$, $\mathtt{o} \in \mathcal{O}$ and $\mathcal{R} = \mathtt{R}$. $\mathtt{t}$ and $\mathtt{u}$ are intermediate variables. Primed variables represent the "next state" of the variable. Arrows represent a dependency.

$$\Omega\left(buy|i, v, bid, ask, t = \text{informed}, u\right) = \begin{cases} (v > ask + \epsilon_t) \wedge (i \geq 1): & 1 \\ \text{otherwise}: & 0 \end{cases}$$

$$\Omega\left(sell|i, v, bid, ask, t = \text{informed}, u\right) = \begin{cases} (v < bid + \epsilon_t): & 1 \\ \text{otherwise}: & 0 \end{cases}$$

$$\Omega\left(hold|i, v, bid, ask, t = \text{informed}, u\right) = \begin{cases} (v > bid + \epsilon_t) \wedge (v < ask + \epsilon_t): & 1 \\ \text{otherwise}: & 0 \end{cases}$$

$$\Omega\left(o|i, v, bid, ask, t = \text{uninformed}, u\right) = \mathbb{I}\left[o = u\right]$$

Ω encodes the signal received by the MM upon the action taken by the trader. We note that information is conveyed only by the direction of the trade. In the case of an $\mathrm{uninformed}$ trader, the observation follows directly from their probabilistic choice determined in the transition function.

- The reward function $\mathcal{R}$, which constrains inventory to be non-negative, is specified as:

$$\mathcal{R}(i', v', bid, ask, t = \text{informed}, u) = \begin{cases} (v' + \epsilon_t > ask) \wedge (i' \geq 1): & ask \\ (v' + \epsilon_t < bid): & -bid \\ (v' + \epsilon_t > bid) \wedge (v' + \epsilon_t < ask): & 0 \\ (v' + \epsilon_t > ask) \wedge (i' < 0): & 0 \\ \text{otherwise}: & -\infty \end{cases}$$

$$\mathcal{R}(i', v', bid, ask, t = \text{uninformed}, u) = \begin{cases} u = buy \wedge (i' \geq 1): & ask \\ u = sell: & -bid \\ u = hold: & 0 \end{cases}$$

The reward received by the MM is dependent upon the action executed by the trader. In the case of a buy order, the MM receives the ask price, a $sell$ order results in the MM paying the bid price, whereas a $hold$ order results in no loss or gain. Our market-making model allows for an expressive class of reward functions that are piecewise linear in the value, inventory, and bid and ask prices. Hard constraints, such as the non-negative inventory condition, can be encoded using $-\infty$. Soft constraints, such as penalising linear deviations from a target inventory level, can be encoded using finite piecewise linear rewards.

The aim of the MM is to maximise its reward over a planning horizon $\mathcal{H}$. In order to do this the MM must trade-off profit taking, which can be seen as exploiting a certain bid-ask pair, and price discovery, where the MM explores other bid-ask pairs in $\mathcal{A}$. Figure 1 shows our model graphically.

As a critical insight in this work, we remark that by formulating the optimal sequential market-making model as a continuous state POMDP it is possible to separate belief state considerations from the dynamic programming solution. In subsequent sections we show that our model leads to an elegant piecewise linear structure in the α-functions derived through an SDP solution. This is in stark contrast with the approach of Das and Magdon-Ismail [2008] which required non-linear operations that complicate the consideration of inventory.

4 Symbolic Dynamic Programming

In this section we make our second contribution by showing how the continuous state POMDP based optimal sequential market-making model can be solved in closed-form via a Symbolic Dynamic Programming (SDP) [Boutilier *et al.*, 2001] based version of PBVI [Pineau *et al.*, 2003].

4.1 Symbolic Case Calculus

SDP assumes that all functions can be represented in case statement form [Boutilier *et al.*, 2001] as follows:

$$f = \begin{cases} \phi_1 : & f_1 \\ \vdots & \vdots \\ \phi_k : & f_k \end{cases}$$

Here, the ϕ_i are logical formulae defined over the state $\vec{x}$ that can consist of arbitrary logical combinations of boolean variables and linear inequalities ($\geq, >, <, \leq$) over continuous variables. We assume that the set of conditions $\{\phi_1, \ldots, \phi_k\}$ disjointly and exhaustively partition $\vec{x}$ such that f is well-defined for all $\vec{x}$. In this paper we restrict the f_i to be either constant or linear functions of the state variable $\vec{x}$. Henceforth, we refer to functions with linear ϕ_i and piecewise constant f_i as linear piecewise constant (LPWC) and functions with linear ϕ_i and piecewise linear f_i as linear piecewise linear (LPWL) functions.

Operations on case statements may be either unary or binary. All of the operations presented here are closed-form for LPWC and LPWL functions. We refer the reader to [Sanner *et al.*, 2011; Zamani and Sanner, 2012] for more thorough expositions of SDP for piecewise continuous functions.

Unary operations on a single case statement *f*, such as scalar multiplication $c \cdot f$ where $c \in \mathbb{R}$, are applied to each f_i ($1 \leq i \leq k$). Binary operations such as addition, subtraction and multiplication are executed in two stages. Firstly, the cross-product of the logical partitions of each case statement is taken, producing paired partitions. Finally, the binary operation is applied to the resulting paired partitions. The "cross-sum" $\oplus$ operation can be performed on two cases in the following manner:

$$\begin{cases} \phi_1 : f_1 \\ \phi_2 : f_2 \end{cases} \oplus \begin{cases} \psi_1 : g_1 \\ \psi_2 : g_2 \end{cases} = \begin{cases} \phi_1 \wedge \psi_1 : & f_1 + g_1 \\ \phi_1 \wedge \psi_2 : & f_1 + g_2 \\ \phi_2 \wedge \psi_1 : & f_2 + g_1 \\ \phi_2 \wedge \psi_2 : & f_2 + g_2 \end{cases}$$

"cross-subtraction" $\ominus$ and "cross-multiplication" $\otimes$ are defined in a similar manner but with the addition operator replaced by the subtraction and multiplication operators, respectively. Some partitions resulting from case operators may be inconsistent and are thus removed.

In principle, case statements can be used to represent all POMDP components, i.e., $\mathcal{R}(\vec{x}, a)$, $\mathcal{T}(\vec{x}, a, \vec{x}')$, $\Omega(o, a, \vec{x}')$, $\alpha(\vec{x})$ and $b(\vec{x})$. In practice, case statements are implemented using a more compact representation known as Extended Algebraic Decision Diagrams (XADDs) [Sanner *et al.*, 2011], which also support efficient versions of all of the aforementioned operations.

4.2 Closed-form Symbolic PBVI for Continuous State POMDPs

In this section we extend the Symbolic PBVI algorithm of Zamani et al. [2012] by relaxing its LPWC assumption to the more general LPWL case. We also show that the set of α-functions in the solution are LPWL functions that permit efficient computation. Symbolic PBVI for continuous state (and discrete $\mathcal{A}$) POMDPs can be written solely in terms of the following case operations:

$$\begin{aligned} \alpha_i^{a,o,h} &= \mathcal{R}(\vec{x}, a) \cdot \frac{1}{|\mathcal{O}|} + \\ &\quad \gamma \otimes \int \Omega(o, a, \vec{x}') \otimes \mathcal{T}(\vec{x}, a, \vec{x}') \otimes \alpha_i(\vec{x}') \, d\vec{x}', \\ &\quad \forall \alpha_i(\vec{x}') \in V^{h-1} \qquad (1) \\ \Gamma^{a,o,h} &= \bigcup_i \left\{ \alpha_i^{a,o,h} \right\} \\ \Gamma_b^{a,h} &= \sum_{o \in \mathcal{O}} \operatorname*{argmax}_{\alpha \in \Gamma^{a,o,h}} (\alpha \cdot b) \\ \alpha_b^h &= \operatorname*{argmax}_{\Gamma_b^{a,h}, a \in \mathcal{A}} \left(\Gamma_b^{a,h} \cdot b \right), \forall b \in \mathcal{B} \\ V^h &= \bigcup_{b \in \mathcal{B}} \left\{ \alpha_b^h \right\} \end{aligned}$$

To calculate the optimal h-stage-to-go value function we modify the Bellman backup in Equation (1) to the following form where we marginalize out intermediate variables for the trader type t and outcome u:

$$\begin{aligned} \alpha_i^{a,o,h} &= \bigoplus_u \bigoplus_t \mathbb{P}(t) \cdot \mathcal{R}(\vec{x}, a, t, u) \cdot \frac{1}{|\mathcal{O}|} + \\ &\quad \gamma \otimes \int \Omega(o, a, \vec{x}', t, u) \otimes \mathcal{T}(\vec{x}, a, \vec{x}', t, u) \otimes \alpha_i(\vec{x}') \, d\vec{x}', \\ &\quad \forall \alpha_i(\vec{x}') \in V^{h-1} \qquad (2) \end{aligned}$$

In Equation (2) the backup operation is calculated as an expectation over trader types t in a given trader model and observations $u \in \{buy, sell, hold\}$ from uninformed traders. We note that this algorithm can be easily extended to very long horizons through *model predictive control* methods [Qin and Badgwell, 2003] which optimise a single-step look-ahead followed by the execution of a static policy.

A critical insight in this work is that all operations used in the algorithm are closed-form for LPWC representations of $\Omega(\cdot)$ and LPWL representations of $\mathcal{T}(\cdot)$ and $\mathcal{R}(\cdot)$ [Sanner *et al.*, 2011; Zamani and Sanner, 2012]. A LPWC $\Omega(\cdot)$ ensures that the α-functions are LPWL and closed-form after the Bellman backup operation; the integration operation in Equation (2) results in a LPWL function. In contrast, a LPWL $\Omega(\cdot)$ would not result in a closed-form LPWL α-function. Therefore, by induction, the Symbolic PBVI value function V^h remains closed-form for arbitrary horizons. This result holds for the subclass of continuous state POMDPs with the aforementioned representations of $\Omega(\cdot)$, $\mathcal{T}(\cdot)$ and $\mathcal{R}(\cdot)$, of which the optimal sequential market-making model in Section 3.2 is an instance.

Symbolic PBVI gives a lower bound on the optimal exact VI [Monahan, 1982] value function and, if all belief states are enumerated to $\mathcal{H}$, the solution is optimal. In the context of optimal sequential market-making, the lower bound policy guarantees that the MM makes at least as much profit as indicated by the PBVI value function evaluated at a belief state.

5 Results

In this section we demonstrate that our novel optimal sequential market-making model: (1) is sequentially optimal; (2) utilises inventory based reasoning; (3) works with flexible trader models; and (4) is computationally tractable.

As far as we are aware, our model is the first to demonstrate these properties exactly and in closed-form. We note that the work of Das and Magdon-Ismail [2008] is restricted to Gaussian initial beliefs, the traditional GM trader model and does not reason about inventory.

5.1 Experimental Settings

For each of the analyses an initial set of beliefs over $\mathcal{S} = \langle v, i \rangle$ is used to generate the set of all reachable beliefs $\mathcal{B}$ to $\mathcal{H}$. Trader proportions under the Glosten-Milgrom and Discrete models were set to $\mathbb{P}(t) = 0.5$ and $\mathbb{P}(t) = 1/3$, respectively. The noise under the Discrete model for over-valuing and under-valuing traders was set to 10.0 and -10.0, respectively. While we can effectively solve for $\mathcal{H} = 30$ or more, we intentionally restrict the horizon to $\mathcal{H} = 3$ in the initial set of evaluations for purposes of interpretation and explanation.

5.2 Sequential Optimality

In Figure 2 we present the sequentially optimal policy of our MM operating within the Glosten-Milgrom trader model with an initial belief of $\langle \{\mathcal{U}[0.0, 50.0], \mathcal{U}[50.0, 100.0]\}, 1.0\rangle$ and $\mathcal{A} = \{(bid, ask)\,|(bid < ask), bid, ask \in \{0.0, 25.0, 49.0, 51.0, 75.0, 99.0\}\}$.

We can see that the MM changes the optimal $\mathcal{H} = 3$ bid-ask pair of $[0.0, 75.0]$ in response to trader actions: the ask price is raised in response to a buy, lowered following a $hold$ and unchanged in the event of a $sell$ order. The actions chosen at $\mathcal{H} = 2$ reflect the optimal choice given the MM's updated belief. At $\mathcal{H} = 1$ the MM chooses actions that ensure that a trader will buy, contingent upon having positive inventory. Furthermore, an optimal bid price of 0.0 is used by the MM for all $\mathcal{H}$, which is due to there being a non-zero probability that an uninformed trader will $sell$ to the MM at this bid price, a fact which the MM exploits. The sequence of actions show how the MM chooses to (asymmetrically) modulate its bid and ask prices to explore market value while protecting itself against inventory constraints.

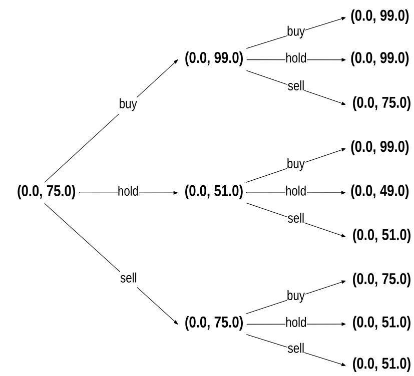


Figure 2: Sequentially optimal policy for an initial belief of $\langle \{\mathcal{U}[0.0, 50.0], \mathcal{U}[50.0, 100.0]\}, 1.0\rangle$. Nodes represent actions $a \in \mathcal{A}$ and arrows represent an observation $o \in \mathcal{O}$.

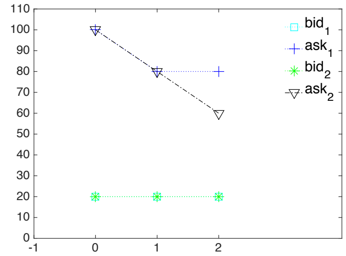
5.3 Inventory Sensitivity

In Figure 3 we present the optimal $\mathcal{H} = 3$ actions of our MM operating within the Glosten-Milgrom trading model with different initial beliefs and $\mathcal{A} = \{(bid, ask)|(bid < ask), bid, ask \in \{20.0, 40.0, 60.0, 80.0, 100\}\}$. Upon comparing Figure 3a and Figure 3b we notice that, with the exception of the $i = 0.0$ case, a narrow initial value belief has a narrower optimal bid-ask spread. It is also evident that the MM uses higher ask prices when the trader population comprises of a higher proportion of uninformed traders. In summary, from Figure 3 we can clearly see that the MM sets its bid and ask prices based on its initial beliefs, characteristics of the trader population and inventory levels.

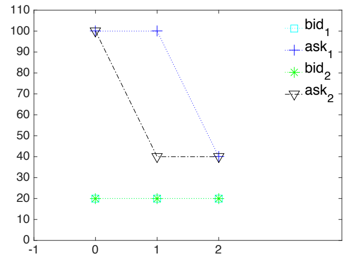
5.4 Trader Models

In the previous two experiments we showed how the optimal sequential market making-model presented in this paper can be used to calculate sequentially optimal and inventory sensitive policies, the two major contributions to algorithmic market making. In this experiment we demonstrate the flexibility of the model by examining the affect of different trader models on the MM's optimal bid and ask prices. In Figure 4 we present the optimal α-functions of our MM operating within the Glosten-Milgrom and Discrete trading model with an initial belief of $\langle \{\mathcal{U}[0.0, 50.0], \mathcal{U}[50.0, 100.0]\}, 1.0\rangle$ and $\mathcal{A} = \{(bid, ask)|(bid < ask), bid \in \{0.0, 25.0, 50.0, 75.0\}, ask \in \{50.0, 75.0, 100.0\}\}$.

In Figure 4a we can observe three distinct optimal actions in the α-function plots. If we set the width to 0.0 and vary the mean, we notice that $[0.0, 100.0]$ exploits the actions of uninformed traders who are equally likely to $sell$ or buy. In the regions where $[0.0, 50.0]$ and $[0.0, 75.0]$ are optimal, we note that as the mean increases above the ask price, the MM is more likely to receive buy orders from an informed trader. In Figure 4b we can see two interesting trends with respect

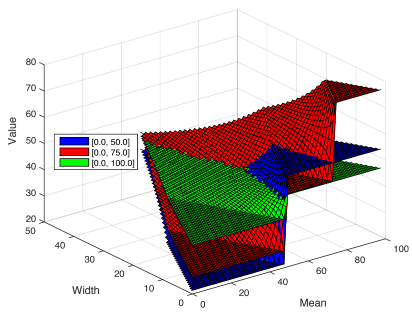


(a) Initial value belief of $v \sim \mathcal{U}[0.0, 100.0]$.

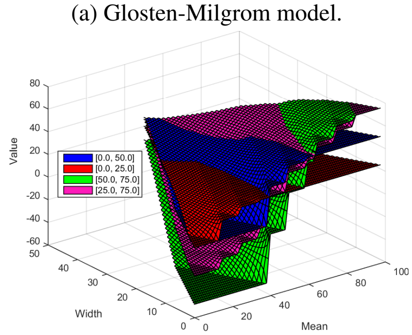


(b) Initial value belief of $v \sim \mathcal{U}[49.0, 51.0]$.

Figure 3: Optimal bid and ask prices under different market settings. The initial inventory level $i \in \{0, 1, 2\}$. (bid_1, ask_1) and (bid_2, ask_2) are the optimal prices for a trader population comprising of 50% and 10% uninformed traders, respectively.



(a) Glosten-Milgrom model.



(b) Discrete model.

Figure 4: Optimal α-functions for $i = 1.0$ under different trader models. The upper surface of the plot indicates the optimal α-function for the market-maker's value belief defined by $v \sim \mathcal{U}[m - w, m + w]$, for a given mean m and width w.

to the optimal actions. Firstly, when the MM is certain about the value, they select actions with increasing ask prices. This phenomenon also occurs when uncertainty about the mean value increases. For example, when the MM becomes increasingly uncertain about a mean value of 70.0, they use actions with a wider bid-ask spread. From Figure 4 it is evident that models of trader noise have a dramatic affect on the MM's optimal actions. Therefore, it is critical for an algorithmic market-making model to be flexible with regards to its reasoning about traders.

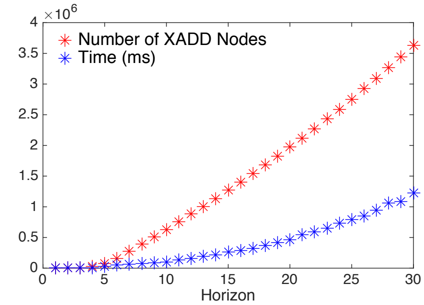


Figure 5: Time and space versus horizon $\mathcal{H}$. Space is measured in the total number of XADD nodes used to represent α-function case statements.

5.5 Time and Space Complexity

Figure 5 shows the relationship between horizon $\mathcal{H}$, computation time and space using a dynamic programming model predictive control approach. It is clear that computation for moderate $\mathcal{H}$ remains tractable. The space requirement grows at a rate greater than time due to an inflation in the number of observation partitions per α-function. This in turn increases the computation time per backup operation.

6 Conclusions and Future Work

In this paper, we have introduced the first optimal solution to sequential market-making with *inventory based reasoning*. We showed that by formulating the market-making problem as a new subclass of *continuous state* POMDPs, we substantially generalize the previous state of the art solution [Das and Magdon-Ismail, 2008] by allowing for more flexibility in the definition of prior beliefs (arbitrary in our formulation) and trader noise models, while also incorporating inventory constraints and operating with a bounded loss. We also provide a novel SDP solution, which allows us to solve this new subclass of POMDPs, and not just a particular market-making problem, in closed form.

There are a number of avenues for further research. Firstly, it is important to explore more expressive representations of the underlying market microstructure model such as incorporating different order sizes, their affect on market price and stealth trading by informed traders [Easley and O'Hara, 1987]. Another possible extension is to learn trader models from the order stream by using continuous extensions of Bayesian reinforcement learning methods for POMDPs [Poupart *et al.*, 2006]. In an orthogonal direction, we can explore POMDP Monte Carlo Tree Search (MCTS) [Silver and Veness, 2010] alternatives to PBVI. Although MCTS *cannot* guarantee lower bound policies like PBVI does, it may scale to larger and more complex POMDPs. These extensions may then be used collectively to model and solve optimal sequential market-making in more complex financial domains. In summary, the modelling and algorithmic contributions in this work opens up an entirely new way to investigate market microstructure, trader noise and inventory models in algorithmic sequential market-making.

Acknowledgments

Sanmay Das acknowledges support from US NSF award 1414452. We thank the anonymous reviewers for their comments and suggestions.

References

[Amihud and Mendelson, 1980] Yakov Amihud and Haim Mendelson. Dealership market: Market-making with inventory. *Journal of Financial Economics*, 8(1):31–53, 1980.

[Aström, 1965] K. J. Aström. Optimal control of markov decision processes with incomplete state estimation. *Journal of Mathematical Analysis and Applications*, 10:174–205, 1965.

[Boutilier *et al.*, 2001] Craig Boutilier, Ray Reiter, and Bob Price. Symbolic dynamic programming for first-order mdps. In *IJCAI*, pages 690–697, 2001.

[Brahma *et al.*, 2012] Aseem Brahma, Mithun Chakraborty, Sanmay Das, Allen Lavoie, and Malik Magdon-Ismail. A bayesian market maker. In *Proceedings of the Thirteenth ACM Conference on Electronic Commerce*, pages 215–232, 2012.

[Chakraborty and Kearns, 2011] Tanmoy Chakraborty and Michael Kearns. Market making and mean reversion. In *Proceedings of the Twelfth ACM Conference on Electronic Commerce*, pages 307–314, 2011.

[Chen and Pennock, 2007] Yiling Chen and David Pennock. A utility framework for bounded-loss market makers. In *UAI*, pages 49–56, 2007.

[Das and Magdon-Ismail, 2008] Sanmay Das and Malik Magdon-Ismail. Adapting to a market shock: Optimal sequential market-making. In *NIPS 22*, pages 361–368, 2008.

[Das, 2005] Sanmay Das. A learning market-maker in the glosten-milgrom model. *Quantitative Finance*, 5(2):169–180, 2005.

[Das, 2008] Sanmay Das. The effects of market-making on price dynamics. In *AAMAS*, pages 887–894, 2008.

[Easley and O'Hara, 1987] David Easley and Maureen O'Hara. Price, trade size, and information in securities markets. *Journal of Financial Economics*, 19(1):69–90, 1987.

[Glosten and Milgrom, 1985] Lawrence R. Glosten and Paul R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100, 1985.

[Hanson, 2003] Robin Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119, 2003.

[Hanson, 2007] Robin Hanson. Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets*, 1(1):3–15, February 2007.

[Kaelbling *et al.*, 1998] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Journal of Artificial Intelligence Research*, 101:99–134, 1998.

[Mandelbrot, 1963] Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36:394, 1963.

[Monahan, 1982] George E. Monahan. A survey of partially observable markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1):1–16, January 1982.

[Othman and Sandholm, 2012] Abraham Othman and Tuomas Sandholm. Rational market making with probabilistic knowledge. In *AAMAS*, pages 645–652, 2012.

[Othman *et al.*, 2010] Abraham Othman, Tuomas Sandholm, David Pennock, and Daniel Reeves. A practical liquidity-sensitive automated market maker. In *Proceedings of the Eleventh ACM Conference on Electronic Commerce*, pages 377–386, 2010.

[Pennock and Sami, 2007] David Pennock and Rahul Sami. In *Algorithmic Game Theory*, chapter Computational Aspects of Prediction Markets. Cambridge University Press, 2007.

[Pineau *et al.*, 2003] Joelle Pineau, Geoffrey J. Gordon, and Sebastian Thrun. Point-based value iteration: An anytime algorithm for POMDPs. In *IJCAI*, pages 1025–1030, 2003.

[Porta *et al.*, 2006] Josep M. Porta, Nikos Vlassis, Matthijs Spaan, and Pascal Poupart. Point-based value iteration for continuous POMDPs. *Journal of Machine Learning Research*, 7:2329–2367, December 2006.

[Poupart *et al.*, 2006] Pascal Poupart, Nikos Vlassis, Jesse Hoey, and Kevin Regan. An analytic solution to discrete bayesian reinforcement learning. In *ICML*, pages 697–704, 2006.

[Qin and Badgwell, 2003] S. Joe Qin and Thomas A. Badgwell. A survey of industrial model predictive control technology. *Control Engineering Practice*, 11(7):733 – 764, 2003.

[Sanner *et al.*, 2011] Scott Sanner, Karina Delgado, and Leliane Nunes de Barros. Symbolic dynamic programming for discrete and continuous state mdps. In *UAI*, pages 1–10, 2011.

[Silver and Veness, 2010] David Silver and Joel Veness. Monte-carlo planning in large pomdps. In *NIPS 23*, pages 2164–2172, 2010.

[Smith *et al.*, 2003] Eric Smith, J. Doyne Farmer, Laszlo Gillemot, and Supriya Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6):481–514, 2003.

[Sondik, 1971] Edward J. Sondik. *The Optimal Control of Partially Observable Markov Decision Processes*. PhD thesis, Stanford University, Palo Alto, California, USA, 1971.

[Sornette, 2004] Didier Sornette. *Why Stock Markets Crash: Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton University Press, 2004.

[Volt, 2005] Johannes Volt. *The Statistical Mechanics of Financial Markets*. Theoretical and Mathematical Physics. Springer Verlag, 2005.

[Zamani and Sanner, 2012] Zahra Zamani and Scott Sanner. Symbolic dynamic programming for continuous state and action mdps. In *AAAI*, pages 1–7, 2012.

[Zamani *et al.*, 2012] Zahra Zamani, Scott Sanner, Pascal Poupart, and Kristian Kersting. Symbolic dynamic programming for continuous state and observation pomdps. In *NIPS 25*, pages 1394–1402, 2012.