Computational methods for market making algorithms

Olivier Guéant

1 Introduction

In finance, securities may be traded through exchanges or over-the-counter (OTC) directly between two parties. In OTC markets, some market participants are systematically providing liquidity to the others by showing/answering prices at which they agree to buy and sell the assets and contracts they cover. These market participants are called dealers or market makers and they play a central role in the functioning of markets.

With the rise of electronification and trading automation, the task of quoting assets on OTC markets must be carried out algorithmically by market makers. Market making models and algorithms have therefore been an important research topic in recent years, at the frontier between economics, quantitative finance, scientific computing, and machine learning.

In the 1980s, long time before algorithmic trading became necessary, economists proposed models where one or several risk-averse market makers optimized their pricing policy for managing their inventory risk (see [16, 17, 18]). More than twenty years later, Avellaneda and Stoikov revisited in [1] that literature with a quantitative finance viewpoint and proposed a model based on stochastic optimal control tools to help market makers determine their optimal quotes. This model inspired both academics and practitioners who then developed several realistic extensions – in particular for OTC markets (foreign exchange, bonds, etc.) although the paper was initially developed for stock markets. In [14], Guéant, Lehalle, and Fernandez-Tapia proved the first set of mathematical results on the Avellaneda-Stoikov model, derived closed-form approximations of the optimal quotes, and proposed extensions

Olivier Guéant
Université Paris 1 Panthéon Sorbonne, Centre d'Economie de la Sorbonne, 106 Boulevard de l'Hôpital, 75013 Paris, France. e-mail: olivier.gueant@univ-paris1.fr

to include a drift in the price dynamics and adverse selection. Cartea and Jaimungal, along with several researchers added many features to the initial models: alpha signals, ambiguity aversion, etc. (see [7, 8, 9, 10]). They also proposed a slightly different optimization framework where market makers maximize their expected profit minus a running penalty to avoid holding large inventories whereas [1] relied on an exponential utility function. To cite a few other extensions, general intensities and partial information in [6], persistence of the order flow in [19], multiple requested sizes in [4], client tiering and access to a liquidity pool in [2].

In most practical cases, market making algorithms must be built for entire portfolios whereas most models proposed in the literature have been single-asset ones until recently. Guéant and Lehalle were the first to touch upon a multi-asset extension of models *à la* Avellaneda-Stoikov in [13] and a complete analysis for the various objective functions present in the literature have been carried out in [12] (see also the book [11]). In spite of equations characterizing the optimal quotes, approximating numerically the optimal quotes remains a research problem, because of the curse of dimensionality. The goal of this chapter is to present a typical multi-asset market making model (Section 2), a theoretical characterization of the optimal quotes in that model (Section 3), and to discuss the various methods proposed in the literature that could be used in the financial industry (Section 4).

2 A multi-asset market making model

We present here a typical model for the market making of $d \geq 1$ assets.

For $i \in \{1,\dots,d\}$, the reference price of asset i is modeled by a process $(S^i_t)_{t\in\mathbb{R}_+}$ with dynamics $dS^i_t = \sigma^i dW^i_t$ and S^i_0 given, where $(W^1_t)_{t\in\mathbb{R}_+},\dots,(W^d_t)_{t\in\mathbb{R}_+}$ are d Brownian motions with correlation matrix $(\rho^{i,j})_{1\leq i,j\leq d}$ – hereafter we write $\Sigma = (\rho^{i,j}\sigma^i\sigma^j)_{1\leq i,j\leq d}$.

At each point in time, the market maker chooses the price at which she is ready to buy/sell each asset: for $i \in \{1,\dots,d\}$, we let her bid and ask quotes for asset i be modeled by two stochastic processes, respectively denoted by $(S^{i,b}_t)_{t\in\mathbb{R}_+}$ and $(S^{i,a}_t)_{t\in\mathbb{R}_+}$. For $i \in \{1,\dots,d\}$, we denote by $(N^{i,b}_t)_{t\in\mathbb{R}_+}$ and $(N^{i,a}_t)_{t\in\mathbb{R}_+}$ the two point processes modeling the number of transactions at the bid and at the ask, respectively, for asset i. In this simple model, the transaction size for asset i is constant and denoted by z^i. The inventory process of the market maker for asset i, denoted by $(q^i_t)_{t\in\mathbb{R}_+}$, has therefore the dynamics $dq^i_t = z^i dN^{i,b}_t - z^i dN^{i,a}_t$ with q^i_0 given, and we denote by $(q_t)_{t\in\mathbb{R}_+}$ the (column) vector process $\left(q^1_t,\dots,q^d_t\right)^\intercal_{t\in\mathbb{R}_+}$.

For each $i \in \{1,\dots,d\}$, we denote by $(\lambda_t^{i,b})_{t\in\mathbb{R}_+}$ and $(\lambda_t^{i,a})_{t\in\mathbb{R}_+}$ the intensity processes of $(N_t^{i,b})_{t\in\mathbb{R}_+}$ and $(N_t^{i,a})_{t\in\mathbb{R}_+}$, respectively.[1] We assume that the market maker stops proposing a bid (respectively ask) price for asset i when her position in asset i following the transaction would exceed a given threshold Q^i (respectively $-Q^i$). We assume that the intensities verify $\lambda_t^{i,b} = \Lambda^{i,b}(\delta_t^{i,b})1_{\{q^i_{t-}+z^i\le Q^i\}}$ and $\lambda_t^{i,a} = \Lambda^{i,a}(\delta_t^{i,a})1_{\{q^i_{t-}-z^i\ge -Q^i\}}$ where the processes $(\delta_t^{i,b})_{t\in\mathbb{R}_+}$ and $(\delta_t^{i,a})_{t\in\mathbb{R}_+}$ are defined by $\delta_t^{i,b} = S_t^i - S_t^{i,b}$ and $\delta_t^{i,a} = S_t^{i,a} - S_t^i$, for all $t \in \mathbb{R}_+$. Moreover, we assume that the functions $\Lambda^{i,b}$ and $\Lambda^{i,a}$ are twice continuously differentiable, decreasing[2] with $\forall\delta\in\mathbb{R}$, $\Lambda^{i,b/a\prime}(\delta) < 0$, and such that $\lim_{\delta\to+\infty}\Lambda^{i,b/a}(\delta) = 0$ and $\sup_\delta \frac{\Lambda^{i,b/a}(\delta)\Lambda^{i,b/a\prime\prime}(\delta)}{\left(\Lambda^{i,b/a\prime}(\delta)\right)^2} < 2$.

Finally, the process $(X_t)_{t\in\mathbb{R}_+}$ modelling the amount of cash on the market maker's cash account has the following dynamics:

$$dX_t = \sum_{i=1}^d S_t^{i,a} z^i dN_t^{i,a} - S_t^{i,b} z^i dN_t^{i,b} = \sum_{i=1}^d \left(\delta_t^{i,b} z^i dN_t^{i,b} + \delta_t^{i,a} z^i dN_t^{i,a}\right) - \sum_{i=1}^d S_t^i dq_t^i.$$

For the market maker, a classical optimization problem consists in maximizing the expected value of an exponential utility function (with risk aversion parameter $\gamma > 0$) applied to the mark-to-market (MtM) value of the portfolio at a given time T, i.e. the amount X_T plus the MtM value $\sum_{i=1}^d q_T^i S_T^i$ of the assets at time T:

$$\sup_{(\delta_t^{1,b})_t,\dots,(\delta_t^{d,b})_t,(\delta_t^{1,a})_t,\dots,(\delta_t^{d,a})_t\in\mathscr{A}} \mathbb{E}\left[-\exp\left(-\gamma\left(X_T + \sum_{i=1}^d q_T^i S_T^i\right)\right)\right],$$

where $\mathscr{A}$ is the set of predictable processes bounded from below.[3]

3 Theoretical results

The above problem is a stochastic optimal control problem that can be solved by using a Hamilton-Jacobi-Bellman (HJB) equation and a verification argument. In our case, the HJB equation is

[1] Intensities are instantaneous probabilities to trade in this context.

[2] The probability to trade with a client depends monotonically on the proposed price.

[3] Alternatively, we can consider a risk-adjusted expectation for the objective function:

$$\sup_{(\delta_t^{1,b})_t,\dots,(\delta_t^{d,b})_t,(\delta_t^{1,a})_t,\dots,(\delta_t^{d,a})_t\in\mathscr{A}} \mathbb{E}\left[X_T + \sum_{i=1}^d q_T^i S_T^i - \frac{1}{2}\gamma\int_0^T q_t^\intercal \Sigma q_t dt\right].$$

Results for one of the two optimization problems usually translate into results for the other.

$$0 = \partial_t u(t,x,q,S) + \frac{1}{2}\sum_{i,j=1}^{d} \rho^{i,j}\sigma^i\sigma^j \partial^2_{S^iS^j} u(t,x,q,S)$$
$$+ \sum_{i=1}^{d} 1_{\{q^i+z^i \le Q^i\}} \sup_{\delta^{i,b}} \Lambda^{i,b}(\delta^{i,b}) \left(u(t,x - z^iS^i + z^i\delta^{i,b}, q + z^ie^i, S) - u(t,x,q,S) \right)$$
$$+ \sum_{i=1}^{d} 1_{\{q^i-z^i \ge -Q^i\}} \sup_{\delta^{i,a}} \Lambda^{i,a}(\delta^{i,a}) \left(u(t,x + z^iS^i + z^i\delta^{i,a}, q - z^ie^i, S) - u(t,x,q,S) \right),$$

for all $(t,x,q,S) \in [0,T) \times \mathbb{R} \times \prod_{i=1}^d \left(z^i\mathbb{Z} \cap [-Q^i,Q^i]\right) \times \mathbb{R}^d$, where $\{e^i\}_{i=1}^d$ is the canonical basis of $\mathbb{R}^d$ and the terminal condition is

$$u(T,x,q,S) = -\exp\left(-\gamma\left(x + \sum_{i=1}^d q^iS^i\right)\right), \forall (x,q,S) \in \mathbb{R} \times \prod_{i=1}^d \left(z^i\mathbb{Z} \cap [-Q^i,Q^i]\right) \times \mathbb{R}^d$$

Using the ansatz $u(t,x,q,S) = -\exp\left(-\gamma\left(x + \sum_{i=1}^d q^iS^i + \theta(t,q)\right)\right)$, it is straightforward to verify that solving the above HJB equation boils down to solving the following system of nonlinear ordinary differential equations:[4]

$$0 = \partial_t \theta(t,q) - \frac{1}{2}\gamma q^\intercal \Sigma q$$
$$+ \sum_{i=1}^{d} 1_{\{q^i+z^i \le Q^i\}} z^i H^{i,b}_\xi \left(\frac{\theta(t,q) - \theta(t,q + z^ie^i)}{z^i} \right)$$
$$+ \sum_{i=1}^{d} 1_{\{q^i-z^i \ge -Q^i\}} z^i H^{i,a}_\xi \left(\frac{\theta(t,q) - \theta(t,q - z^ie^i)}{z^i} \right)$$

with terminal condition $\theta(t,q) = 0$, where, for each $i \in \{1,\dots,d\}$, $H^{i,b/a}(p) = \sup_\delta \frac{\Lambda^{i,b/a}(\delta)}{\gamma z^i}(1 - \exp(-\gamma z^i(\delta - p)))$.

Then, one can prove the following theorem using a verification argument (see [12], with slightly different notations):

Theorem 1. *There exists a unique function* $\theta : [0,T] \times \prod_{i=1}^d \left(z^i\mathbb{Z} \cap [-Q^i,Q^i]\right) \to \mathbb{R}$, *$C^1$ in time, solution of the above equation. Moreover, for* $i \in \{1,\dots,d\}$, *the optimal bid and ask quotes are characterized by*

$$\delta^{i,b*}_t = \tilde{\delta}^{i,b*}\left(\frac{\theta(t,q_{t-}) - \theta(t,q_{t-} + z^ie^i)}{z^i} \right) \quad \textit{for } q_{t-} + z^ie^i \in \prod_{j=1}^d \left(z^j\mathbb{Z} \cap [-Q^j,Q^j]\right),$$
$$\delta^{i,a*}_t = \tilde{\delta}^{i,a*}\left(\frac{\theta(t,q_{t-}) - \theta(t,q_{t-} - z^ie^i)}{z^i} \right) \quad \textit{for } q_{t-} - z^ie^i \in \prod_{j=1}^d \left(z^j\mathbb{Z} \cap [-Q^j,Q^j]\right),$$

[4] It is indeed a system of nonlinear ordinary differential equations because the variable q takes discrete values.

where the functions $\tilde{\delta}^{i,b*}(\cdot)$ *and* $\tilde{\delta}^{i,a*}_{\gamma}(\cdot)$ *are defined by*

$$\tilde{\delta}^{i,b/a*}(p) = {\Lambda^{i,b/a}}^{-1}\left(\gamma z^i H^{i,b/a}(p) - {H^{i,b/a}}'(p)\right).$$

4 Numerical methods

The above theorem states that finding the optimal quotes boils down to solving two problems:[5] (i) finding a numerical approximation of the function θ and (ii) computing the functions $\tilde{\delta}^{i,b/a*}$ for $i \in \{1,\ldots,d\}$. The latter problem does not raise any issue as the functions can be computed asset by asset by using classical optimization techniques. For the former, one needs to approximate numerically the solution of a system of nonlinear ordinary differential equations. For that purpose, two families of methods exist: grid methods where the solution is approximated at specific points and formula methods where the solution is approximated using simple or complex "combinations" of simple functions.

In the literature, it is common to see finite different methods on a grid to approximate θ. More precisely, Euler monotone schemes – explicit or implicit – are often used to solve this type of problems (see for instance [2, 4, 12]). Grid methods are very efficient in the one-asset case ($d = 1$) or when d is small (say $d \leq 3$). However, because they require a grid of dimension $d+1$ (one dimension of time and d dimensions for the assets) grid methods naturally suffer from the curse of dimensionality and cannot be used for larger d.

Grid methods can nevertheless be used if one reduces beforehand the dimensionality of the problem. An interesting way proposed in [4] consists of (i) approximating the covariance matrix Σ by a low-rank symmetric matrix by using a principal component analysis and keeping $k \leq 3$ risk factors – and therefore replacing the d-dimensional variable q by a low-dimensional one corresponding to the k risk factors – and (ii) replacing the risk limits in terms of assets $(Q^i)_i$ by risk limits in terms of factor exposures. By using this approximation, θ can be regarded as a function of time and risk factors and approximated using a grid of dimension $k+1$ and not $d+1$.

To beat the curse of dimensionality, formula methods can of course be used. Closed-form formulas have been proposed in [12] and more recently in [5]. In [5], the idea was to "approximate" the system of nonlinear ordinary differential equations by a multi-dimensional Riccati equation that can be solved in closed form. The approximation of θ turns out to be a polynomial of degree 2 in that case and approximations of the optimal quotes are then derived asset by asset using the above equations. This type of techniques provides great results, as exemplified in [5].

[5] For most extensions of the above model, these two problems remain the relevant ones.

In order to approximate the optimal quotes using a formula method, another interesting idea consists in looking for the function θ and sometimes the optimal quotes themselves in the form of neural networks using reinforcement learning techniques. Promising results in this line can be found in [15].

Numerical examples regarding credit indices are presented in [12]. For bonds, the papers [4] and [15] contain interesting illustrations. The case of foreign exchange has been tackled recently in [3] and should attract more interest in the near future.

References

1. Marco Avellaneda and Sasha Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.
2. Alexander Barzykin, Philippe Bergault, and Olivier Guéant. Algorithmic market making in foreign exchange cash markets with hedging and market impact. *Working paper*, 2021.
3. Alexander Barzykin, Philippe Bergault, and Olivier Guéant. Market making by an FX dealer: tiers, pricing ladders and hedging rates for optimal risk control. *Working paper*, 2022.
4. Philippe Bergault and Olivier Guéant. Size matters for OTC market makers: general results and dimensionality reduction technique. *Mathematical Finance*, 31(1):279–322 2020.
5. Philippe Bergault, David Evangelista, Olivier Guéant, and Douglas Vieira. Closed-form approximations in multi-asset market making. *Applied Mathematical Finance*, to appear.
6. Luciano Campi and Diego Zabaljauregui Optimal market making under partial information with general intensities. *Applied Mathematical Finance*–27, 2020.
7. Álvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. Algorithmic trading with model uncertainty. *SIAM Journal on Financial Mathematics*, 8(1):635–671, 2017.
8. Álvaro Cartea, Sebastian Jaimungal, and José Penalva. *Algorithmic and high-frequency trading*. Cambridge University Press, 2015.
9. Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. Buy low, sell high: A high frequency trading perspective. *SIAM Journal on Financial Mathematics*, 5(1):415–444, 2014.
10. Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. Algorithmic trading, stochastic control, and mutually exciting processes. *SIAM Review*, 60(3):673–703, 2018.
11. Olivier Guéant. *The Financial Mathematics of Market Liquidity: From optimal execution to market making*, volume 33. CRC Press, 2016.
12. Olivier Guéant. Optimal market making. *Applied Mathematical Finance*, 24(2):112–154, 2017.
13. Olivier Guéant and Charles-Albert Lehalle. General intensity shapes in optimal liquidation. *Mathematical Finance*, 25(3):457–495, 2015.
14. Olivier Guéant, Charles-Albert Lehalle, and Joaquin Fernandez-Tapia. Dealing with the inventory risk: a solution to the market making problem. *Mathematics and Financial Economics*, 7(4):477–507, 2013.
15. Olivier Guéant and Iuliia Manziuk. Deep reinforcement learning for market making in corporate bonds: beating the curse of dimensionality. *Applied Mathematical Finance*, 26(5):387–452, 2019.
16. Thomas Ho and Hans R Stoll. On dealer markets under competition. *The Journal of Finance*, 35(2):259–267, 1980.
17. Thomas Ho and Hans R Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics*, 9(1):47–73, 1981.
18. Thomas Ho and Hans R Stoll. The dynamics of dealer markets under competition. *The Journal of Finance*, 38(4):1053–1074, 1983.
19. Paul Jusselin. Optimal market making with persistent order flow. *arXiv preprint arXiv:2003.05958*, 2020.