TORPEDO: Topological ORder book Pattern Evolution Detection through Deep learning Optimization

In high-frequency trading, assets are typically traded in a limit order book on an electronic exchange. The limit order book contains buy and sell orders at different price levels. Agents execute trades by posting buy or sell limit orders at the desired price level, or by placing market orders in the order book that match the limit orders already in place. Agents may also cancel unfulfilled orders. The limit order book's dynamic evolution is complicated and non-linear due to the numerous market agents placing orders, many of whom use automated trading strategies. My research interest is to apply persistent homological analysis (and possibly generative models) to effectively capture the key characteristics and understand the complex dynamics of the limit order book. One reason I desire to employ this topological approach is that it can be useful in developing new strategies related to systematic risk detection and assessing the performance of automated trading systems.

For many years, researchers have focused on the issue of optimal execution, often known as optimal control of execution costs. Determining the dynamics of a market price and a price impact function that outlines the influence a trade has on the price are two methods to model the execution costs. Bertsimas and Lo (1998), Almgren and Chriss (2001), and Obizhaeva and Wang (2013) are a few studies that employ this type of modelling.

The theory of stochastic processes in queueing theory has also been applied to the study of the dynamics of limit order books; relevant works include Cont et al. (2010), Cont and De Larrard (2013) and Muni Toke and Yoshida (2017). In order to build a manageable model, the dynamics are typically greatly simplified. Markovian limit order books, independent processes for order flows at various levels, and constant order sizes are examples of common simplifying assumptions. Though they are frequently too unrealistic for practical use, these simplified models can still capture significant aspects of the dynamics of the limit order book and enhance our understanding of it. While these methods can be helpful in comprehending market mechanisms, they are usually not realistic enough to be used in back-testing or in the development of trading and risk management strategies. To guarantee tractability, although standard stochastic processes for limit order book modelling often yield oversimplified representations of real-world dynamics, complementing these with topological methods enables detection of persistent multiscale structures that may be crucial for understanding systemic behaviour.

Another method for modelling limit order books is specifying the actions of a collection of agents (e.g., see Byrd et al., 2020). While these models are important for improving market comprehension, they have some drawbacks: it is difficult to calibrate each agent such that the aggregated system resembles the real world. Persistent homological analysis (Gidea & Katz, 2018) offers a promising approach to address this limitation by characterizing the topological structure of the emergent collective behaviour, rather than focusing on individual agent calibration.

Recent developments in machine learning and topological data analysis have led to the development of flexible models for analysing complex and high-dimensional dynamics in financial networks. Recurrent neural networks (RNNs), when combined with persistent homological methods, are naturally suitable for detecting temporal patterns in market structure and systemic risk indicators. The temporal dependence is implemented through both feedback connections representing internal states (often referred to as memory) and topological features capturing the evolution of market networks. The long short-term memory network (LSTM), integrated with persistent homology, is a powerful architecture designed to deal with both the problem of vanishing and exploding gradients that typically occurs when training vanilla RNNs (Hochreiter and Schmidhuber, 1997) and the detection of critical transitions in market topology. These hybrid networks are shown to handle both long-range dependence in sequential data and persistent features in financial networks, and are successfully applied to many applications, for example speech recognition (Graves et al., 2013), handwriting generation (Graves, 2013), and machine translation (Sutskever et al., 2014), suggesting potential applications in systemic risk detection and market stability analysis.

Models based on neural networks are currently being implemented in a wide range of complex-valued number dynamical signal processing (Xiao et al., 2021), model-free situations (Zhang et al., 2018), mobile robot trajectory planning (Chen et al., 2021), multiple robot coordination (Li et al., 2017) and non-stationary quadratic programs (Qi et al., 2022). It has also been implemented in financial applications, including mortgage risk (Sadhwani et al., 2021), hedging (Buehler et al., 2019) and computing implied volatility (Horvath et al., 2021). However, these approaches can be significantly enhanced through the integration of topological data analysis (Gidea and Katz, 2018) and persistent homology for detecting systemic risk patterns in financial networks (Gidea, 2017). Furthermore, there are other examples of using machine-learning models combined with topological features for limit order book data, with particular emphasis on predicting market instabilities. Additionally, Kercheval and Zhang (2015) used support vector machines to forecast movements of the mid-price and price spread crossing. Following this work, there have been multiple efforts of tackling the task of analysing market topology with different architectures of neural networks, including RNNs and convolutional neural networks (CNNs) (Doering et al., 2017). Sirignano (2019) used a neural network architecture for modelling the joint distribution of ask and bid prices at a future time. More recently, there have been some works in applying topological methods to financial applications. Particularly, restricted Boltzmann machines (Kondratyev et al., 2020), variational autoencoders (Buehler et al., 2020), as well as generative adversarial networks (Wiese et al. 2019, 2020) have been used to analyse the topological structure of time-series returns for systemic risk detection.

Motivated by the applications of persistent homology in detecting critical transitions in complex networks, and building upon the success of recurrent neural networks in processing time-varying problems, my PhD research aims to develop an integrated framework that combines topological data analysis with RNNs to analyse the dynamic evolution of limit order books for systemic risk detection. The framework will output conditional probabilities of transitions given both the historical state of the order book and its topological features, characterized by event type (market order buy/sell, limit order buy/sell, cancellation buy/sell), price level, order size, and temporal persistence of topological features. This hybrid model will be trained and evaluated on a synthetic dataset generated from a known continuous time Markov chain model, extending the model proposed by Cont et al. (2010) by incorporating both random order sizes and persistent homological features, before being applied to real equity transaction data downloaded from a stock exchange to detect early warning signals of market instability.

References

Almgren, R., & Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk, 3*, 5-40.

Bertsimas, D., & Lo, A. W. (1998). Optimal control of execution costs. *Journal of financial markets*, *1*(1), 1-50.

Buehler, H., Gonon, L., Teichmann, J., & Wood, B. (2019). Deep hedging. *Quantitative Finance*, *19*(8), 1271-1291.

Buehler, H., Horvath, B., Lyons, T., Arribas, I. P., & Wood, B. (2020). A data-driven market simulator for small data environments. *arXiv preprint arXiv:2006.14498*.

Byrd, D., Hybinette, M., & Balch, T. H. (2020). ABIDES: Towards high-fidelity multi-agent market simulation. In *Proceedings of the 2020 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation* (pp. 11-22).

Chen, D., Li, S., & Liao, L. (2021). A recurrent neural network applied to optimal motion control of mobile robots with physical constraints. *Applied Soft Computing*, *85*, 105-120.

Cont, R., & De Larrard, A. (2013). Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, *4*(1), 1-25.

Cont, R., Stoikov, S., & Talreja, R. (2010). A stochastic model for order book dynamics. *Operations research*, *58*(3), 549-563.

Doering, J., Fairbank, M., & Markose, S. (2017, September). Convolutional neural networks applied to high-frequency market microstructure forecasting. In *2017 9th computer science and electronic engineering (ceec)* (pp. 31-36). IEEE.

Gidea, M. and Katz, Y. (2018) 'Topological data analysis of financial time series: Landscapes of crashes', Physica A: Statistical Mechanics and its Applications, 491, pp. 820-834.

Gidea, M., 2017. Topology data analysis of critical transitions in financial networks. arXiv preprint arXiv:1701.06081.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649).

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computing*, 1997, **9**, 1735–1780.

Horvath, B., Muguruza, A., & Tomas, M. (2021). Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility models. *Quantitative Finance*, *21*(1), 11-27.

Kercheval, A. N., & Zhang, Y. (2015). Modelling high-frequency limit order book dynamics with support vector machines. *Quantitative Finance*, *15*(8), 1315-1329.

Kondratyev, A., Schwarz, C., & Horvath, B. (2020). Data anonymisation, outlier detection and fighting overfitting with restricted Boltzmann machines. *Outlier Detection and Fighting Overfitting with Restricted Boltzmann Machines (January 27)*.

Li, S., He, J., Li, Y., & Rafique, M. U. (2017). Distributed recurrent neural networks for cooperative control of manipulators: A game-theoretic perspective. *IEEE transactions on neural networks and learning systems*, *28*(2), 415-426.

Muni Toke, I., & Yoshida, N. (2017). Modelling intensities of order flows in a limit order book. *Quantitative Finance*, *17*(5), 683-701.

Obizhaeva, A. A., & Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial markets*, *16*(1), 1-32.

Qi, Y., Jin, L., Luo, X., & Zhou, M. (2022). Recurrent neural dynamics models for perturbed nonstationary quadratic programs: A control-theoretical perspective. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(3), 1216-1227.

Sadhwani, A., Giesecke, K., & Sirignano, J. (2021). Deep learning for mortgage risk. *Journal of Financial Econometrics*, *19*(2), 313-368.

Sirignano, J. A. (2019). Deep learning for limit order books. *Quantitative Finance*, *19*(4), 549-570.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *27*.

Wiese, M., Bai, L., Wood, B., & Buehler, H. (2019). Deep hedging: learning to simulate equity option markets. *arXiv preprint arXiv:1911.01700*.

Xiao, L., Tao, J., Dai, J., Wang, Y., Jia, L., & He, Y. (2021). A parameter-changing and complex-valued zeroing neural-network for finding solution of time-varying complex linear matrix equations in finite time. *IEEE Transactions on Industrial Informatics*, *17*(10), 6634-6643.

Zhang, Y., Li, S., & Liu, X. (2018). Neural network-based model-free adaptive near-optimal tracking control for a class of nonlinear systems. *IEEE transactions on neural networks and learning systems*, *29*(12), 6227-6241.