Single- and Multiplayer Trade Execution Strategies under Transient Price Impact

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
der Universität Mannheim

vorgelegt von

Elias Strehle
aus Soltau

Mannheim, 2017

Dekan:	Dr. Bernd Lübcke, Universität Mannheim
Referent:	Professor Dr. Alexander Schied, Universität Mannheim
Korreferent:	Professor Dr. Martin Schlather, Universität Mannheim

Tag der mündlichen Prüfung: 21. Dezember 2017

To my friends

Abstract

The problem of optimal execution is to trade a fixed amount of a financial asset over a fixed time horizon in a way that minimizes costs from price impact and transaction costs. Three types of price impact can be distinguished: Temporary, transient and permanent price impact. While mathematical models of optimal execution under temporary and permanent price impact can be analyzed with standard methods from the calculus of variations, models featuring transient price impact are more complex.

This thesis studies optimal execution under transient price impact for a single investor and for multiple investors. Assuming that trading incurs quadratic transaction costs, existence and uniqueness of optimal execution strategies and Nash equilibria is established for a large class of transient price impact functions. Closed-form representations of Nash equilibria are derived under the assumption that price impact decays exponentially. These representations are studied in detail to arrive at an economic evaluation of order anticipation strategies and predatory trading.

A second focus of this thesis is the intimate connection between problems of optimal execution and Fredholm integral equations. It is shown that, given information about certain characteristics of transient price impact, one can deduce qualitative features of optimal execution strategies, such as nonnegativity and convexity, from the corresponding Fredholm integral equations without obtaining an explicit solution.

Zusammenfassung

Soll eine vorgegebene Menge eines Finanzprodukts über einen vorgegebenen Zeitraum ge- oder verkauft werden, stellt sich die Frage nach der optimalen Handelsstrategie, die Kosten durch Preiseinfluss und Transaktionskosten minimiert. Hierbei können drei Arten von Preiseinfluss unterschieden werden: Sofortiger, vergänglicher und dauerhafter Preiseinfluss. Während bei der Suche nach optimalen Handelsstrategien im Hinblick auf sofortigen und dauerhaften Preiseinfluss Standardmethoden der Variationsrechnung ausreichen, führt die Berücksichtigung von vergänglichem Preiseinfluss zu komplexeren Optimierungsproblemen.

Diese Arbeit untersucht, wie vorgegebene Mengen im Hinblick auf vergänglichen Preiseinfluss optimal gehandelt werden sollten. Unter Annahme quadratischer Transaktionskosten werden Existenz und Eindeutigkeit von optimalen Handelsstrategien für einen einzelnen Investor sowie von Nash-Gleichgewichten für eine beliebige Anzahl an Investoren bewiesen. Für den Spezialfall exponentiell abklingenden Preiseinflusses werden Nash-Gleichgewichte in geschlossener Form hergeleitet, mit deren Hilfe eine ökonomische Bewertung opportunistischer Handelsstrategien, die vom Preiseinfluss anderer Strategien profitieren, vorgenommen wird.

Die enge Beziehung zwischen optimalen Handelsstrategien und Fredholm-Integralgleichungen ist ein weiterer Schwerpunkt dieser Arbeit. Mithilfe solcher Gleichungen lassen sich qualitative Eigenschaften optimaler Handelsstrategien, wie zum Beispiel Nichtnegativität und Konvexität, auch in Fällen beweisen, in denen eine explizite Lösung nicht verfügbar ist.

Acknowledgments

I wish to thank my PhD supervisor Alexander Schied. He went out of his way to ensure I had all the time and resources I needed to write my thesis. His generous support allowed me to meet the best researchers in my field and have a memorable and productive research stay in Waterloo, Canada.
I am also grateful to Martin Schlather for reviewing my thesis, to Bernd Lübcke, Wolfgang K. Seiler and Li Chen for completing my examination board, to Tao Zhang for the fruitful collaboration, to Deutsche Forschungsgesellschaft for financial support, and to my colleagues Susanne Spether and Alexander Kalinin.

Contents

Chapter 1

Introduction

The microstructure of financial markets has received much attention in recent years. Advances in economic and mathematical theory have made it possible to drop the assumption of "perfect" financial markets, where trading is frictionless and liquidity infinite. At the same time, the rise of high-speed algorithmic trading has made it increasingly important for investors and regulators to understand market microstructure and be aware of the influence that transaction costs and liquidity have on the profitability of trading strategies and on the stability of financial markets.
Liquidity is particularly relevant for investors who must buy or sell large amounts of a financial asset over a relatively short time period. Examples include:

- a proprietary trader quickly selling an asset after its price has dropped below his stop-loss threshold;
- a bank buying ten percent of a company's stock on behalf of a customer over the course of a week;
- a high frequency trader closing all positions at the end of a trading day;
- a manager of a physical index fund rebalancing his portfolio over the course of one trading day after the composition of the tracked index has changed.

In each case, the investor has a *liquidation constraint*, i.e., he must trade a fixed net amount over a fixed time horizon. The only *admissible* trading strategies are those that satisfy the liquidation constraint.
Trading large amounts in one piece can be prohibitively expensive or even impossible. For large US stocks, the total volume of orders available in the limit order book at any given time is only about one percent of traded daily volume (Bouchaud et al., 2009, p. 19). Hence a large order is typically split into a sequence of smaller so-called child orders that are executed over the given time period.
Every execution of an order changes the balance of supply and demand, and thus potentially impacts the market price. In addition, orders incur transaction costs such as transaction taxes, brokerage fees and slippage. An investor with a liquidation constraint therefore faces the *problem of optimal execution*: Among all admissible

strategies, find an *optimal strategy* that minimizes expected costs from price impact and transaction costs (or, more generally: that minimizes a given risk-return criterion).

Three types of *price impact* can be distinguished (see Figure 1.1): *Temporary*, *transient* and *permanent* price impact. Ho and Stoll (1981) explain temporary and transient price impact as a consequence of risk aversion among market makers. For the sake of illustration, consider an asset that is traded via a limit order book. A market maker places buy and sell offers in the limit order book. Once an investor submits a marketable sell order, it is executed against the market maker's buy offers with the highest price. These buy offers vanish from the limit order book; as a direct consequence, the price decreases. This effect is called *temporary price impact*. *Transient price impact* is more subtle. It does not affect the execution price of the order that caused it, but that of subsequent orders. When some of the market maker's buy offers are filled, he acquires a certain amount of the asset. This exposes him to inventory risk (from changes in the market price) and non-execution risk (from uncertainty about the arrival of marketable buy orders). A risk-averse market maker counteracts by submitting cheaper sell offers to attract buyers, and cheaper buy offers to repel sellers. On average, this leads to a surplus of buyers in the market, and the market maker's inventory is gradually reduced by incoming buy orders. As this happens, he again increases the price of his buy and sell offers until both have reached their initial level, again leaving the market maker with an empty inventory and the market in a state of balanced supply and demand.

An alternative to Ho and Stoll's explanation of price impact is information asymmetry (Glosten and Milgrom, 1985). Some marketable orders are submitted by investors because they have new information about the fundamental asset value. Consequently, large orders raise suspicion among market makers, who adjust the prices of their offers accordingly. If an order was indeed based on new information, the price adjustment persists. This effect is called *permanent price impact*.

Standard theories of arbitrage disregard price impact. Even if the "unaffected" price process (i.e., the price process that is realized if the investor does not trade) permits no risk-free gains, there might be a *price manipulation strategy* which makes money from its own price impact (Huberman and Stanzl, 2005). As a simple example, suppose there is no temporary and transient, but positive permanent price impact. In this case, an investor may buy the asset at the current unaffected market price of, say, €100. This purchase has a permanent price impact which increases the asset price to, say, €105. The investor may now sell the asset again to clear inventory and realize a profit from price manipulation of €5.

Price manipulation strategies are usually not arbitrage strategies in the classical sense because their profitability can be affected by random fluctuations in the asset price. But on average they earn excess returns and therefore belong to the larger class of *statistical arbitrage strategies*. For transient price impact in particular it is not obvious under what market conditions price manipulation is possible (Gatheral, 2010; Alfonsi et al., 2012).

Price impact can also be viewed as an externality on other investors. The price im-

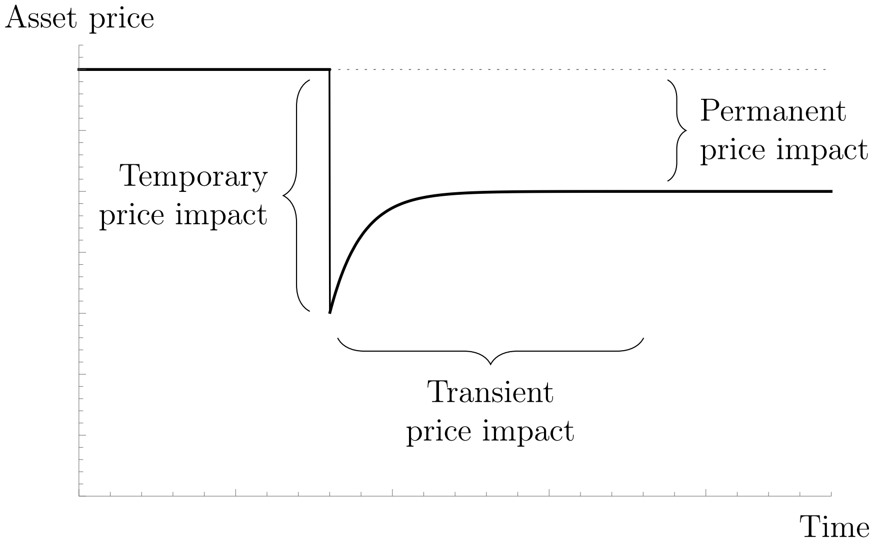

Figure 1.1: Idealized price impact of a sell order.

pact of a sell order constitutes a negative externality for every investor who is selling, and a positive externality for every investor who is buying. Opportunistic investors who are able to detect an execution strategy and predict its trading behavior can deliberately exploit this externality by pursuing an *order anticipation strategy*. Such a strategy often involves *predatory trading*, i.e., selling ahead of the execution strategy and buying back afterwards, to turn the price impact of the execution strategy into a predictable source of profit. Predatory trading can strongly increase an investor's costs from price impact. Consequently, execution strategies must adapt to the presence of other investors.

Studying the interaction between liquidating and opportunistic investors is also important from a regulatory point of view. If a large sell order is executed, the market price drops. Brunnermeier and Pedersen (2005) argue that the additional sell orders submitted by predatory traders must amplify this price drop, an effect known as *price overshooting*. This may lead to a domino effect: The amplified price drop triggers stop-loss thresholds from other investors, resulting in new large sell orders. These are again subject to predatory trading, creating an even larger price drop, etc., until the market breaks down. Transaction taxes have the potential to prevent predatory trading (Schied and Zhang, 2017) and therefore reduce the likelihood of market breakdowns.

Mathematical models of optimal execution

The key feature of any model of optimal execution is the *liquidation constraint*: An investor must execute a fixed net amount x^0 over a given time horizon $[0, T]$. Suppose the investor trades continuously and controls his instantaneous rate of trading $\alpha(t)\,\mathrm{d}t$ for all $t \in [0, T]$. Then α is an *admissible strategy* if it satisfies the liquidation constraint $\int_0^T \alpha(t)\,\mathrm{d}t = x^0$ (and some technical conditions). Let $S(t; \alpha)$ denote the asset price, including price impact, from the investor's strategy α. Integration by parts shows that the investor's total costs from price impact are $\int_0^T \alpha(t)S(t;\alpha)\,\mathrm{d}t$ plus a constant. Suppose further that a trading rate of $\alpha(t)\,\mathrm{d}t$ incurs additional transaction costs $c(\alpha(t))\,\mathrm{d}t$. Then the investor's total *expected costs* are given by

$$J[\alpha] = \mathbb{E}\Big[\int_0^T \big(c(\alpha(t)) + \alpha(t)S(t;\alpha)\big)\,\mathrm{d}t\Big].$$

The mathematical *problem of optimal execution* is to find a minimizer of J in the class of admissible strategies.

Bertsimas and Lo (1998) and Almgren and Chriss (2001) were among the first to study optimal execution under price impact. Their papers feature linear temporary and permanent, but no transient price impact. This corresponds to an asset price evolution of

$$S(t;\alpha) = S^0(t) + \frac{\gamma}{2}\alpha(t) + \lambda\int_0^t \alpha(s)\,\mathrm{d}s.$$

The process S^0 is the *unaffected price process* and is assumed to be a martingale. For risk-neutral investors, one can show that the optimal strategy does not depend on S^0, and that it is sufficient to consider deterministic strategies. The nonnegative constants γ and λ determine the respective size of temporary and permanent price impact. Notice that costs from temporary price impact contribute to expected costs in the same way as quadratic transaction costs $c(a) = \frac{\gamma}{2}a^2$.

One may also consider non-linear temporary and permanent price impact. Then

$$S(t;\alpha) = S^0(t) + \frac{\gamma}{2}f_1(\alpha(t)) + \lambda\int_0^t f_2(\alpha(s))\,\mathrm{d}s$$

for some functions f_1 and f_2. See for instance Almgren (2003), Huberman and Stanzl (2004) and Carmona and Yang (2011).

An empirical study by Bouchaud et al. (2004) suggests that price impact is typically transient, not permanent. Transient price impact can be modeled via a *decay kernel* G, which describes how the asset price "digests" orders over time. The asset price then evolves according to

$$S(t;\alpha) = S^0(t) + \int_0^t G(t-s)\alpha(s)\,\mathrm{d}s. \tag{1.1}$$

It is of course possible to add temporary and permanent price impact, the latter by adding a constant to G. The shape of G can be derived from microstructural models of limit order books (Alfonsi et al., 2010). From an economic point of view, it is sensible to consider decay kernels that are nonnegative and nonincreasing. Empirical studies suggest that transient price impact is described well by *power-law decay kernels* $G(t) = t^{-\rho}$ for $0 < \rho < 1$ (Bouchaud et al., 2004; Almgren et al., 2005). Theoretical models often consider *exponential decay kernels* $G(t) = e^{-\rho t}$ for $\rho > 0$ (Bouchaud, 2010; Lorenz and Schied, 2013; Obizhaeva and Wang, 2013;

Schied and Zhang, 2017). In this case, transient price impact must be linear to exclude price manipulation (Gatheral, 2010, Lemma 4.1). General decay kernels and their compatibility with absence of price manipulation are studied by Gatheral (2010), Gatheral et al. (2012) and Curato et al. (2017).
If more than one investor trades, the problem of optimal execution can be analyzed with tools from game theory. The search for an optimal strategy is extended to the search for a *Nash equilibrium*, i.e., a tuple of admissible strategies that minimize each investor's expected costs under the assumption that no other investor deviates from equilibrium. Multi-investor models of optimal execution include Carlin et al. (2007), Schöneborn and Schied (2009), Carmona and Yang (2011), Moallemi et al. (2012), Schied and Zhang (2015), Lachapelle et al. (2016), Cardaliaguet and Lehalle (2017) and Huang et al. (2017).
In models without transient price impact, optimal strategies can be obtained through classical methods from the calculus of variations. But new methods must be developed to deal with transient price impact. Gatheral et al. (2012), Alfonsi and Schied (2013) and Schied and Zhang (2017) map out the connection between optimal execution under transient price impact and different variants of *Fredholm integral equations*. For the single-investor case with transient price impact of the form (1.1) and quadratic trading costs $c(a) = \frac{\gamma}{2}a^2$, where $\gamma > 0$, it can be shown that an admissible strategy α^* is the unique optimal strategy if and only if there is a constant η such that

$$\gamma\alpha^*(t) + \int_0^T G(|t-s|)\alpha^*(s)\,\mathrm{d}s = \eta, \qquad t \in [0,T]. \tag{1.2}$$

This is a Fredholm integral equation of the second kind with constant free term η. Such equations are an interesting object of study in themselves. Fredholm integral equations have many other applications, for instance in electrostatics (Love, 1949), transport theory (Kaper and Kellogg, 1977) and quantum mechanics (Arfken and Weber, 2005, Example 16.1.1).

1.1 Statement of results

My thesis studies optimal execution under transient price impact for a single investor and for multiple investors. It focuses on qualitative features of optimal strategies and what these features imply from an economic point of view. Special consideration is given to the influence of transaction costs.

Chapter 2: Predatory trading in a game of optimal execution

Consider the discrete time model of optimal execution with two investors introduced by Zhang (2014). It can be viewed as a discrete time version of the model discussed above. Trading occurs at the equidistant time points $t_i := (i-1)T/n$ for $i = 1, 2, \dots, n+1$. At each time t_i, the first and second investor execute orders of size ξ_i and η_i. The first investor must trade a fixed net amount x^0. Consequently, any admissible strategy for the first investor $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_{n+1})$ must satisfy the liquidation constraint $\xi_1 + \xi_2 + \cdots + \xi_{n+1} = x^0$. Similarly, the second investor must trade a fixed net amount y^0, and any admissible strategy for the second investor $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_{n+1})$ must satisfy $\eta_1 + \eta_2 + \cdots + \eta_{n+1} = y^0$.

Price impact is transient and decays at an exponential rate. Given that the first and second investor pursue admissible strategies $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, the asset price evolves according to

$$S(t; \boldsymbol{\xi}, \boldsymbol{\eta}) := S^0(t) + \sum_{t_i < t} e^{-\rho(t - t_i)}(\xi_i + \eta_i), \qquad t \in [0, T],$$

where $\rho > 0$. The unaffected price process S^0 is assumed to be a martingale.

In addition to costs from price impact, each trade incurs quadratic transaction costs of size $\tilde{\gamma} \geq 0$. In total, the first investor's expected costs can be shown to be

$$J[\boldsymbol{\xi} \,|\, \boldsymbol{\eta}] := \mathbb{E}\Big[\sum_{i=1}^{n+1} \Big(\xi_i S(t_i; \boldsymbol{\xi}, \boldsymbol{\eta}) + \frac{1}{2}\xi_i^2 + \varepsilon_i\, \xi_i \eta_i + \tilde{\gamma}\, \xi_i^2\Big)\Big].$$

Here each ε_i is a Bernoulli(1/2)-distributed random variable modeling the order in which the trades ξ_i and η_i are executed. The second investor's expected costs $J[\boldsymbol{\eta} \,|\, \boldsymbol{\xi}]$ are defined similarly.

It is shown in Zhang's (2014) doctoral thesis, and also in Schied and Zhang (2017), that a unique Nash equilibrium $(\boldsymbol{\xi}^*, \boldsymbol{\eta}^*)$ exists for every pair of liquidation constraints (x^0, y^0). Both equilibrium strategies are deterministic and can be represented in terms of two matrix inverses. Furthermore, both investors engage in predatory trading if and only if the level of transaction costs $\tilde{\gamma}$ is smaller than 1/4. In this case, the investors' situation resembles a prisoner's dilemma: Both would benefit from agreeing to refrain from predatory trading, but either would have an incentive to deviate.

Transaction costs can come in the form of a transaction tax. In view of the results above, such a tax has the potential to prevent predatory trading. Numerical simulations by Schied and Zhang (2017) suggest that it can even reduce the expected costs of both investors. I make this observation precise in the first part of Chapter 2 by calculating both investors' equilibrium strategies and expected costs in closed form and analyzing under what conditions a transaction tax is advantageous or disadvantageous.

Although I obtain an explicit formula for the expected costs $J[\boldsymbol{\xi} \,|\, \boldsymbol{\eta}]$ for arbitrary n, it is too complicated to be of further use. The limit $n \to \infty$ of expected costs, on the other hand, is relatively simple.

Define three functions $c_+, c_0^1, c_0^2 \colon (0, \infty) \times \mathbb{R}^2 \to \mathbb{R}$ via

$$\begin{aligned} c_+(r, x, y) &:= \frac{(x + y)^2(36e^{6r}(8r + 13) - 60e^{3r} - 3)}{16(2e^{3r}(3r + 5) - 1)^2} + \frac{(x + y)(x - y)}{2(r + 1)} + \frac{(x - y)^2}{16(r + 1)^2}, \\ c_0^1(r, x, y) &:= \frac{(x + y)^2(6e^{6r} + 3)}{2(2e^{6r}(3r + 5) + e^{3r} + 3r + 7)} + \frac{(x + y)(x - y)}{2(e^{-r} + r + 1)} \end{aligned}$$

and

$$c_0^2(r,x,y) := \frac{(x+y)^2(6e^{6r}-3)}{2(2e^{6r}(3r+5)-3e^{3r}-3r-7)} + \frac{(x+y)(x-y)}{2(-e^{-r}+r+1)}.$$

Result 1 (Theorem 2.6).

(i) If $\tilde{\gamma} > 0$, then $J[\boldsymbol{\xi} \,|\, \boldsymbol{\eta}]$ converges to $c_+(\rho T, x^0, y^0)$ as $n \to \infty$.

(ii) If $\tilde{\gamma} = 0$, then $J[\boldsymbol{\xi} \,|\, \boldsymbol{\eta}]$ converges to $c_0^1(\rho T, x^0, y^0)$ for even n and $c_0^2(\rho T, x^0, y^0)$ for odd n as $n \to \infty$.

The limit of the second investor's expected costs $J[\boldsymbol{\eta} \,|\, \boldsymbol{\xi}]$ is obtained by interchanging x^0 and y^0. It is now easy to check numerically for which parameter combinations $(\rho T, x^0, y^0)$ positive transaction costs $\tilde{\gamma} > 0$ are advantageous or disadvantageous for both investors.

Observation 2. *In the limit $n \to \infty$, positive transaction costs $\tilde{\gamma} > 0$ are advantageous for both investors if ρT is sufficiently large and x^0 and y^0 are close. They are disadvantageous for both investors if x^0 and y^0 have different signs or the difference between x^0 and y^0 is large.*

For the special case $x^0 = y^0$, I show that positive transaction costs are advantageous if $\rho T > \log(4 + \sqrt{62}/3) \approx 0.69$.

The second part of Chapter 2 is concerned with the transition from discrete to continuous time and the convergence of Nash equilibria. It is based on the continuous time version of the model above, as introduced in Zhang's (2014) doctoral thesis. Assume again that two investors must trade fixed net amounts x^0 and y^0 until time T. Trading may now occur at any time $t \in [0, T]$. An admissible strategy for the first investor is a right-continuous, adapted and bounded process $X = (X(t))_{t \in [0-,T]}$ that has finite and $\mathbb{P}$-a.s. bounded total variation and satisfies the execution constraint $X(0-) = x^0$ and $X(T) = 0$ $\mathbb{P}$-a.s. The value $X(t)$ corresponds to the net order remaining at time t. An admissible strategy for the second investor Y is defined similarly.
If the investors pursue admissible strategies X and Y, the asset price evolves according to

$$S(t) = S(t; X, Y) := S^0(t) + \int_{[0,t)} e^{-\rho(t-s)}\, \mathrm{d}X(s) + \int_{[0,t)} e^{-\rho(t-s)}\, \mathrm{d}Y(s), \quad t \in [0, T].$$

Assume again that there are quadratic transaction costs of size $\tilde{\gamma} \geq 0$. Zhang (2014) shows that the the first investor's expected costs are

$$\begin{aligned} J[X \,|\, Y] := \mathbb{E}\Big[&\frac{1}{2} \int_{[0,T]} \int_{[0,T]} e^{-\rho|t-s|}\, \mathrm{d}X(s)\, \mathrm{d}X(t) + \int_{[0,T]} \int_{[0,t)} e^{-\rho(t-s)}\, \mathrm{d}Y(s)\, \mathrm{d}X(t) \\ &+ \frac{1}{2} \sum_{t \in [0,T]} \Delta X(t) \Delta Y(t) + \tilde{\gamma} \sum_{t \in [0,T]} \Delta X(t)^2 \Big]. \end{aligned}$$

The second investor's expected costs are $J[Y \mid X]$. The definition of J can be shown to follow naturally from the definition of expected costs in the corresponding discrete time model.
In his doctoral thesis, Zhang (2014) shows that the continuous time model admits a unique Nash equilibrium if and only if the level $\tilde{\gamma}$ of transaction costs equals $1/4$. In this case, the Nash equilibrium (X^*, Y^*) is deterministic and can be calculated explicitly. I show that expected costs in this equilibrium are equal to the limit of expected costs in the corresponding discrete time models.

Result 3 (Proposition 2.10). *If $\tilde{\gamma} = 1/4$, then the first investor's expected costs in equilibrium, $J[X^* \mid Y^*]$, equal $c_+(\rho T, x^0, y^0)$.*

Finally, I argue that non-existence of Nash equilibria for $\tilde{\gamma} \neq 1/4$ is an undesirable consequence of how transaction costs are implemented in the model: They apply to discrete trades $\Delta X(t)$ and can be avoided in continuous time by choosing a strategy which is absolutely continuous. I suggest that in continuous time models, transaction costs should apply to the instantaneous rate of trading $\mathrm{d}X(t)$ instead. This leads to a different model of optimal execution, which I study in the next chapter.

Chapter 3: A different approach to modeling transaction costs

Consider $n+1$ investors trading a financial asset. Each investor $i = 0, 1, \ldots, n$ must trade a fixed net amount x_i^0 and controls his instantaneous rate of trading $\alpha_i(t)\,\mathrm{d}t$ over the time horizon $[0, T]$. Consequently, a strategy α_i is admissible if it satisfies the liquidation constraint $\int_0^T \alpha_i(t)\,\mathrm{d}t = x_i^0$ (and some technical conditions).
Transient price impact is modeled via a decay kernel $G\colon [0,\infty) \to [0,\infty)$ that is assumed to be square-integrable. Given admissible strategies $\alpha \coloneqq (\alpha_0, \alpha_1, \ldots, \alpha_n)$, the asset price evolves according to

$$S(t;\alpha) \coloneqq S^0(t) + \int_0^t G(t-s) \sum_{i=0}^n \alpha_i(s)\,\mathrm{d}s, \qquad t \in [0,T].$$

This can be viewed as the extension of (1.1) to $n+1$ investors. Each investor faces quadratic transaction costs $\gamma_i/2 > 0$. In line with the argument above, these costs apply to the instantaneous rate of trading $\alpha_i(t)\,\mathrm{d}t$. In total, investor i's costs of execution are

$$J_i[\alpha_i \mid \alpha_{-i}] \coloneqq \int_0^T \big(\frac{\gamma_i}{2}\alpha_i(t)^2 + \alpha_i(t) S(t;\alpha)\big)\,\mathrm{d}t,$$

where $\alpha_{-i} \coloneqq (\alpha_0, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n)$. Assume that each investor minimizes expected costs of execution.
Absence of price manipulation demands that the decay kernel G be of positive type, i.e.,

$$\int_0^\tau \int_0^\tau G(|t-s|)\alpha(t)\alpha(s)\,\mathrm{d}s\,\mathrm{d}t \geq 0$$

for all strategies $\alpha \in L^2[0,\tau]$ and all $\tau > 0$. I show that this assumption is sufficient to guarantee existence and uniqueness of a Nash equilibrium.

Result 4 (Theorem 3.6). *There is a unique Nash equilibrium α^* in the class of admissible strategies. It is deterministic.*

Each investor obtains his optimal execution strategy by solving a Fredholm integral equation of the second kind.
I then consider the special case where only one strategic investor trades, i.e., $n = 0$. The investor's optimal strategy is available in closed form for some decay kernels, and can be shown to display a number of desirable features—such as differentiability, nonnegativity and convexity—for many others.
Given that the single investor pursues an admissible strategy $\alpha = \alpha_0$, the asset price evolves according to (1.1). Notice that in the single-investor case, transaction costs $\gamma = \gamma_0$ can also be interpreted as arising from (linear) temporary price impact as in Almgren and Chriss (2001). The investor's costs of execution are

$$J[\alpha] = J_0[\alpha] = \int_0^T \big(\frac{\gamma}{2}\alpha(t)^2 + \alpha(t)S(t;\alpha)\big)\,\mathrm{d}t.$$

As shown above, a unique optimal strategy α^* exists. It is the only admissible strategy which solves the Fredholm integral equation (1.2) for some $\eta \in \mathbb{R}$.
I present explicit solutions of (1.2) for specific decay kernels, including capped linear decay $G(t) = (1-t)^+$. Assume that $T = m$ is a natural number. For $i = 1, 2, \dots, m$, define

$$\lambda_i := 2\big(1 - \cos\big(\frac{i\pi}{m+1}\big)\big) \qquad \text{and} \qquad b_i := \sqrt{\lambda_i/\gamma}.$$

Denote by I the m-dimensional identity matrix. Define the m-dimensional square matrices

$$\begin{aligned} B &:= \operatorname{diag}(b_1, b_2, \dots, b_m), & E(t) &:= \operatorname{diag}\big(e^{b_1 t}, e^{b_2 t}, \dots, e^{b_m t}\big), \quad t \in [0,T], \\ \Sigma &:= \operatorname{diag}(1, -1, 1, \dots, \pm 1), & K &:= I + \big(\mathbb{1}_{\{j=m-i\}}\big)_{i,j=1,2,\dots,m}, \end{aligned}$$

where $\mathbb{1}$ denotes the indicator function, and

$$Q := \Big(\sin\big(\frac{ij\pi}{m+1}\big)\Big)_{i,j=1,2,\dots,m}.$$

Finally, define $a \in \mathbb{R}^m$ by

$$a := \Big(\gamma Q\big(E(1) + \Sigma\big) + KQ\big((E(1) - I)(\Sigma - I) + B(E(1) - \Sigma)\big)B^{-2}\Big)^{-1} \begin{pmatrix} \eta \\ \vdots \\ \eta \end{pmatrix}.$$

Result 5 (Proposition 3.8). *Suppose $T = m$ is a natural number and $G(t) = (1-t)^+$. Define $\psi = (\psi_1, \psi_2, \dots, \psi_m)\colon [0,1] \to \mathbb{R}^m$ via $\psi(\tau) = Q\big(E(\tau) + E(1-\tau)\Sigma\big)a$. Then the solution α^* of (3.8) satisfies $\alpha^*(\tau + i - 1) = \psi_i(\tau)$ for all $\tau \in [0,1]$ and all $i = 1, 2, \dots, m$.*

For many interesting decay kernels, no explicit solution α^* of (1.2) is known. But it is often still possible to deduce certain qualitative features of the solution. Two features that are particularly interesting from an economic perspective are nonnegativity and convexity. If α^* is interpreted as an optimal execution strategy, nonnegativity implies that transaction-triggered price manipulation strategies, which alternate between buy and sell orders instead of trading in one direction only, are not optimal. Optimal strategies that only trade in one direction are also desirable from a modeling perspective: They remain optimal if the asset price S is not specified exogenously by (1.1), but derived from a model of a (block-shaped) limit order book (Gatheral et al., 2012). Convexity of α^*, on the other hand, mirrors the empirically observed U-shape of the daily distribution of market liquidity.

The decay kernel G is said to be positive definite if $\sum_{i,j=1}^m G(|t_i - t_j|)z_i z_j \geq 0$ for all $m \in \mathbb{N}$, $t_1, t_2, \ldots, t_m \in [0, \infty)$ and $z_1, z_2, \ldots, z_m \in \mathbb{R}$. If G is continuous, then it is positive definite if and only if it is of positive type.

The decay kernel G is said to be completely monotone if it is smooth and $(-1)^n G^{(n)} \geq 0$ for all $n \in \mathbb{N}$.

Result 6 (Lemma 3.9, Proposition 3.10, Theorem 3.11 and Theorem 3.13). *Suppose $n = 0$. Let α^* denote the single investor's optimal strategy. The following statements are true:*

(i) α^* is symmetric around $T/2$, i.e., $\alpha^*(t) = \alpha^*(T - t)$ for every $t \in [0, T]$.

(ii) If G is continuous, then α^* is continuous. If G is m-times differentiable, then α^* is $(m + 1)$-times differentiable.

(iii) If G is positive definite and $\gamma \geq G(0)T$, or G is convex and nonincreasing, then $x_0^0 \alpha^*$ is nonnegative.

(iv) If G is completely monotone, then $x_0^0 \alpha^*$ is convex.

I prove the second part of (iii) as well as (iv) in a more general setting in Chapter 4.

Now consider the general case with $n + 1$ strategic investors. To make it tractable, I assume that price impact decays exponentially, i.e., $G(t) = e^{-\rho t}$ for $\rho > 0$. This allows an explicit representation of the unique Nash equilibrium. Define the $(n+2)$-dimensional square matrices

$$M := \begin{pmatrix} \rho & -\frac{1}{\gamma_0} & \cdots & -\frac{1}{\gamma_0} & \frac{2\rho}{\gamma_0} \\ -\frac{1}{\gamma_1} & \rho & \cdots & -\frac{1}{\gamma_1} & \frac{2\rho}{\gamma_1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{1}{\gamma_n} & -\frac{1}{\gamma_n} & \cdots & \rho & \frac{2\rho}{\gamma_n} \\ 1 & 1 & \cdots & 1 & -\rho \end{pmatrix}, \qquad N_1 := \begin{pmatrix} \rho\gamma_0 & 0 & \cdots & 0 & \rho \\ 0 & \rho\gamma_1 & \cdots & 0 & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \rho\gamma_n & \rho \\ \gamma_0 & \gamma_1 & \cdots & \gamma_n & n+1 \end{pmatrix}.$$

Define further the $(n+1)\times(n+2)$-dimensional matrix

$$W := \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

and the $(n+2)$-dimensional column vector $v := (0,0,\dots,0,1)$. Denote by I_{n+1} the $(n+1)$-dimensional identity matrix. Define the $(n+2)$-dimensional square block matrix

$$N_2 := \begin{bmatrix} W\big((M^{-1}+N_1T)e^{MT}-M^{-1}\big) \\ v^\top(I_{n+1}+N_1e^{MT}) \end{bmatrix}.$$

Result 7 (Theorem 3.14). *The matrix N_2 is invertible. Suppose $G(t)=e^{-\rho t}$ for $\rho>0$. Let $\alpha^*=(\alpha_0^*,\alpha_1^*,\dots,\alpha_{n+1}^*)$ be the unique Nash equilibrium and denote by $S^*:=S(\,\cdot\,;\alpha^*)$ the corresponding asset price. Then the function $\psi^*:=(\alpha_0^*,\alpha_1^*,\dots,\alpha_n^*,S^*)$ satisfies*

$$\psi^*(t)=(e^{Mt}+N_1e^{MT})N_2^{-1}\tilde{x}^0, \qquad t\in[0,T],$$

where $\tilde{x}^0:=(x_0^0,x_1^0,\dots,x_n^0,0)$.

This result makes it possible to study qualitative features of equilibrium strategies. I consider the case where one investor executes a large sell order, while n opportunistic investors pursue order anticipation strategies to benefit from the large sell order's price impact. My main objective is to test the claim by Brunnermeier and Pedersen (2005) that opportunistic investors cause price overshooting.

Observation 8. *In general, opportunistic investors do not cause price overshooting. On the contrary, they often reduce the price drop caused by the large sell order, in particular in models with a "short memory" (i.e., with large values of ρ).*

I discuss two possible explanations: Price overshooting occurs only in markets with long-lived or permanent price impact; or price overshooting is prevented by quadratic transaction costs.

Chapter 4: Completely monotone decay kernels

In the previous chapter, the decay kernel G was assumed to be square-integrable, and $G(0)$ was assumed to be finite. These assumptions are too rigid to study the class of power-law decay kernels $G(t)=t^{-\rho}$, where $0<\rho<1$, which seem to fit empirical observations of transient price impact.

Assume instead that $G\colon(0,\infty)\to[0,\infty)$ is nonconstant, continuous and satisfies $\int_0^\tau G(t)\,\mathrm{d}t<\infty$ for every $\tau>0$. Notice that G may have a weak singularity $\lim_{t\to0}G(t)=\infty$. As before, assume additionally that G is of positive type, i.e., $\int_0^\tau\int_0^\tau G(|t-s|)\alpha(t)\alpha(s)\,\mathrm{d}s\,\mathrm{d}t\geq0$ for every $\alpha\in L^2[0,\tau]$ for which the double integral is well-defined, and every $\tau>0$.

Suppose that a single investor must sell a fixed net amount x^0 over the time horizon $[0, T]$.
If there are no transaction costs, i.e., $\gamma = 0$, an admissible strategy is a function $X\colon [0-, T] \to \mathbb{R}$ which is right-continuous, of bounded total variation and satisfies the liquidation constraint $X(0-) = x^0$ and $X(T) = 0$. Finding an optimal execution strategy means minimizing the cost functional

$$J^0[X] := \frac{1}{2} \int_{[0,T]} \int_{[0,T]} G(|t-s|) \, \mathrm{d}X(s) \, \mathrm{d}X(t)$$

over admissible strategies X for which J^0 is well-defined. The value $X(t)$ specifies the size of the remaining net order that the investor must execute during $[t, T]$.
If there are positive quadratic transaction costs of size $\gamma > 0$, an admissible strategy is a function $\alpha\colon [0, T] \to \mathbb{R}$ which is square-integrable and satisfies the liquidation constraint $\int_0^T \alpha(t) \, \mathrm{d}t = x^0$. Finding an optimal execution strategy means minimizing the cost functional

$$J^\gamma[\alpha] := \frac{1}{2} \int_0^T \big(\gamma\alpha(t) + \int_0^T G(|t-s|)\alpha(t)\alpha(s) \, \mathrm{d}s\big) \, \mathrm{d}t$$

over admissible strategies α for which J^γ is well-defined.
The case $\gamma = 0$ must be treated separately because optimal strategies in this case are usually not absolutely continuous. Hence it is necessary to specify the remaining net amount $X(t)$ directly, as in the continuous time model of Chapter 2. Transaction costs $\gamma > 0$ enforce absolute continuity of the remaining net amount and it is more straightforward to optimize over the instantaneous rate of trading $\alpha(t) = -\frac{\mathrm{d}}{\mathrm{d}t} X(t)$.
Gatheral et al. (2012) show that if the decay kernel G is convex and nonincreasing, then J^0 admits a unique minimizer X^*, which is nonincreasing. I prove a parallel result for the case $\gamma > 0$.

Result 9 (Theorem 4.4). *If G is convex and nonincreasing, then J^γ admits a unique minimizer α^* in the class of admissible strategies for every $\gamma > 0$. In this case, α^* is nonnegative.*

If the decay kernel is additionally assumed to be completely monotone, this has strong implications for the minimizers X^* and α^*. Say that a function $f\colon [0, T] \to \mathbb{R}$ is symmetrically totally monotone if it is analytic on $(0, T)$ and there are nonnegative coefficients $(z_{2k})_{k\in\mathbb{N}}$ such that its power series development in $T/2$ is of the form $f(t) = \sum_{k=0}^{\infty} z_{2k}(t - T/2)^{2k}$.

Result 10 (Theorem 4.6). *Suppose G is completely monotone. Then the following statements are true:*

(i) For every $\gamma > 0$, the unique minimizer of J^γ is symmetrically totally monotone.

(ii) For $\gamma = 0$, let X^* be the unique minimizer of J^0. Then $-X^*$ admits a symmetrically totally monotone derivative on $(0, T)$.

Chapter 2

Predatory trading in a game of optimal execution

Building on Schöneborn's (2008) doctoral thesis, Schied and Zhang (2017) study how two investors interact in the discrete time model of optimal execution introduced by Obizhaeva and Wang (2013). Both investors must satisfy a liquidation constraint by trading at $n+1$ prespecified time points. Price impact is transient and decays at an exponential rate. Each trade incurs quadratic transaction costs.

An investor's price impact can be viewed as an externality on the other investor; a negative externality if both investors trade in the same direction, a positive externality otherwise. Schied and Zhang show that under certain conditions, the optimal reaction to the negative externality is *predatory trading*, i.e., trading aggressively to benefit from the other investor's price impact. At the same time, investors try to avoid being preyed upon. As a result, their optimal strategies quickly alternate between buy and sell orders. This is expensive and leads to a situation that resembles a prisoner's dilemma: Both investors would benefit from agreeing to pursue only non-predatory strategies, but either would have an incentive to deviate from such an agreement.

Introducing a sufficiently high transaction tax can make predatory trading unprofitable. Schied and Zhang show numerically that the cost savings from this may even exceed the additional costs from taxation. They conclude that in certain situations, transaction taxes are advantageous for both investors.

The main objective of this chapter is to corroborate Schied and Zhang's numerical observation. In Sections 2.1 and 2.2, I review the model and key results from Schied and Zhang (2017). I argue in what sense the investors' situation resembles a prisoner's dilemma and how a transaction tax can help. Section 2.3 makes this mathematically precise: For the limit case $n \to \infty$, I obtain closed-form expressions for both investors' expected costs in equilibrium. Using these expressions, I determine for which model parameters transaction taxes are advantageous or disadvantageous for both investors.

Section 2.4 is concerned with the transition from discrete to continuous time. I review the continuous time version of the model, which was first studied in Zhang's

(2014) doctoral thesis. Here, a Nash equilibrium fails to exist unless transaction costs are equal to a critical value. For this critical value, I prove that the limit of expected costs obtained in Section 2.3 coincides with the expected costs in the continuous time equilibrium. I argue that the nonexistence of Nash equilibria is an undesirable consequence of the way transaction costs are modeled in continuous time and suggest a different approach.
Section 2.5 contains the tedious derivation of the closed-form expressions for the limit case $n \to \infty$ from Section 2.3.
This chapter is a thoroughly revised and extended version of results published in Schied et al. (2017).

2.1 The model

Two investors trade a financial asset over the time period $[0, T]$. If neither investor trades, the price evolution of the financial asset is modeled as a right-continuous martingale $S^0 = (S^0(t))_{t\in[0,T]}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ satisfying the usual conditions. Assume that $\mathcal{F}_0$ is $\mathbb{P}$-trivial.
Both investors must trade a fixed net amount until time T. Let x^0 denote the first investor's net amount, and y^0 the second investor's. Trading occurs at the equidistant time points $t_i \coloneqq (i-1)T/n$ for $i = 1, 2, \ldots, n+1$. At each time t_i, the first investor executes an order of size ξ_i, and the second investor executes an order of size η_i. Consequently, a vector of bounded random variables $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_{n+1})$ is called an *(admissible) strategy for the first investor* if (i) every ξ_i is $\mathcal{F}_{t_i}$-measurable; and (ii) the execution constraint $\xi_1 + \xi_2 + \cdots + \xi_{n+1} = x^0$ is satisfied $\mathbb{P}$-a.s.
An *(admissible) strategy for the second investor* $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_{n+1})$ is defined in the same way. The execution constraint becomes $\eta_1 + \eta_2 + \cdots + \eta_{n+1} = y^0$.
To be consistent with subsequent chapters, a positive sign in these strategies indicates a buy order, a negative sign a sell order. This interpretation deviates from Schied and Zhang (2017), where a positive sign indicates a sell order. It follows Gatheral (2010) and Gatheral et al. (2012) instead. Remark 2.1 below explains why this is merely a matter of convention and does not affect the mathematical analysis.
Trading impacts the asset price. Assume that price impact is linear with exponential decay kernel $G(t) = e^{-\rho t}$, where $\rho > 0$. Given admissible execution strategies $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, the asset price evolves according to

$$S(t) = S(t; \boldsymbol{\xi}, \boldsymbol{\eta}) \coloneqq S^0(t) + \lambda \sum_{t_i < t} e^{-\rho(t-t_i)}(\xi_i + \eta_i), \qquad t \in [0, T],$$

where $\lambda > 0$. There is no loss of generality in letting $\lambda = 1$, since all other model parameters can be scaled accordingly.
Suppose the first investor executes an order of size ξ_i at time t_i. What does this cost him? Assume for the moment that the second investor does not trade at time t_i. The first investor's order moves the price from $S(t_i)$ to $S(t_i) + \lambda\,\xi_i = S(t_i) + \xi_i$. As Alfonsi et al. (2010) explain in detail, linear price impact corresponds to a flat order

book, hence the costs of execution are

$$\int_0^{\xi_i} (S(t_i) + x)\,\mathrm{d}x = \xi_i S(t_i) + \frac{1}{2}\xi_i^2.$$

Now assume instead that the second investor executes an order of size η_i at time t_i. Then the costs of execution depend on whose order is executed first. Suppose the second investor's order is executed before the first investor's. The price that the first investor now faces is not $S(t_i)$, but $S(t_i) + \eta_i$. His costs of execution become $\xi_i S(t_i) + \frac{1}{2}\xi_i^2 + \xi_i \eta_i$. If on the other hand the first investor's order is executed first, his costs of execution remain $\xi_i S(t_i) + \frac{1}{2}\xi_i^2$. The second investor's costs of execution, however, change by $\xi_i \eta_i$. Under the assumption that neither investor has a speed advantage, both orders are equally likely to be executed first. Let $(\varepsilon_i)_{i=1,2,\dots,n+1}$ be an independent sequence of Bernoulli(1/2)-distributed random variables that is independent of $\sigma(\bigcup_{t\in[0,T]} \mathcal{F}_t)$. Attribute the extra cost $\xi_i \eta_i$ to the first investor if $\varepsilon_i = 1$; attribute it to the second investor if $1 - \varepsilon_i = 1$.

Assume additionally that an order of size ξ_i incurs quadratic transaction costs $\tilde{\gamma}\,\xi_i^2$, where $\tilde{\gamma} \geq 0$. In single-investor models (including the model in Chapter 4), transaction costs of this form may be interpreted as costs arising from *temporary price impact*. It is tempting to follow Huang et al. (2017) in using the same interpretation for models with two or more investors. But this is incorrect: If an order generates temporary price impact, it affects the execution price of every order subsequently executed at the same time t_i. It becomes necessary to model the order in which trades arriving at the same time are executed (which is accomplished here by the random variables $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{n+1}$). One might also choose to apply the same costs from temporary price impact to all orders arriving at the same time, as in Carlin et al. (2007). But notice that the probability of being executed first (and thus the probability of being subject to temporary price impact from other investors) depends on the number of investors. This must be taken into account when comparing models with different numbers of investors. In any case, the transaction costs $\tilde{\gamma}\,\xi_i^2$ only affect the investor who caused them and cannot be viewed as costs from temporary price impact. They should be interpreted as general costs arising from market frictions (Gatheral, 2010, p. 751) or—as argued later—costs arising from a transaction tax. See Kissell et al. (2004) for a comprehensive overview of transaction costs on financial markets.

It would certainly be desirable to replace quadratic transaction costs $\tilde{\gamma}\,\xi_i^2$ with a more general cost function. But only quadratic transaction costs seem to allow a closed-form representation of optimal execution strategies. Consider also Proposition 2.6 in Schied and Zhang (2017): There is a piecewise linear cost function for which the optimal strategies derived under the assumption of quadratic transaction costs are also optimal. Notice however that this cost function depends on all model parameters. In particular, different values of n yield different cost functions.

In total, if the two investors pursue strategies $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, the *first investor's expected costs* add up to

$$J[\boldsymbol{\xi}\,|\,\boldsymbol{\eta}] := \mathbb{E}\Big[\sum_{i=1}^{n+1}\Big(\xi_i S(t_i;\boldsymbol{\xi},\boldsymbol{\eta}) + \frac{1}{2}\xi_i^2 + \varepsilon_i\,\xi_i\eta_i + \tilde{\gamma}\,\xi_i^2\Big)\Big]$$

and the *second investor's expected costs* add up to

$$J[\boldsymbol{\eta}\,|\,\boldsymbol{\xi}] := \mathbb{E}\Big[\sum_{i=1}^{n+1}\Big(\eta_i S(t_i;\boldsymbol{\xi},\boldsymbol{\eta}) + \frac{1}{2}\eta_i^2 + (1-\varepsilon_i)\,\xi_i\eta_i + \tilde{\gamma}\,\eta_i^2\Big)\Big].$$

No contradiction arises from denoting both the first and the second investor's expected costs by the same functional J. Indeed, by independence,

$$\mathbb{E}[\varepsilon_i\,\xi_i\eta_i] = \frac{1}{2}\mathbb{E}[\xi_i\eta_i] = \mathbb{E}[(1-\varepsilon_i)\,\xi_i\eta_i].$$

From an economic perspective, it would be more sensible to consider *expected execution shortfall* instead of expected costs. Execution shortfall is the difference between expected costs and the mark-to-market value $x^0 S(0)$ of the net order. Since the mark-to-market value is a constant, it is irrelevant to the task of finding a cost-minimizing strategy. Hence both quantities lead to the same optimal strategies, and considering expected costs instead of expected execution shortfall yields slightly shorter formulas.

Remark 2.1. *A positive sign of ξ_i indicates a sell order, not a buy order, in Schied and Zhang (2017) (and in Zhang, 2014, and Schied et al., 2017). The asset price $S(t;\boldsymbol{\xi},\boldsymbol{\eta})$ under this interpretation evolves according to*

$$S^0(t) - \sum_{t_i<t} e^{-\rho(t-t_i)}(\xi_i+\eta_i),$$

and the first investor's costs are

$$\begin{aligned}
&\sum_{i=1}^{n+1}\Big(-\,\xi_i S(t_i;\boldsymbol{\xi},\boldsymbol{\eta}) + \frac{1}{2}\xi_i^2 + \varepsilon_i\xi_i\eta_i + \tilde{\gamma}\,\xi_i^2\Big)\\
&= \sum_{i=1}^{n+1}\Big(-\,\xi_i S^0(t) + \xi_i\sum_{t_j<t_i} e^{-\rho(t_i-t_j)}(\xi_j+\eta_j) + \frac{1}{2}\xi_i^2 + \varepsilon_i\xi_i\eta_i + \tilde{\gamma}\,\xi_i^2\Big).
\end{aligned}$$

The only difference under this interpretation is that S^0 is replaced with $-S^0$. All results in this chapter are independent of S^0, and are therefore valid under either interpretation.

Assume that both investors are risk-neutral and want to minimize expected costs. A pair $(\boldsymbol{\xi}^*,\boldsymbol{\eta}^*)$ of admissible strategies for the first and the second investor is called

a *Nash equilibrium* if each investor's strategy minimizes costs, given that the other investor does not deviate from equilibrium, i.e., $J[\boldsymbol{\xi}^* \,|\, \boldsymbol{\eta}^*] \le J[\boldsymbol{\xi} \,|\, \boldsymbol{\eta}^*]$ for every admissible strategy for the first investor $\boldsymbol{\xi}$, and $J[\boldsymbol{\eta}^* \,|\, \boldsymbol{\xi}^*] \le J[\boldsymbol{\eta} \,|\, \boldsymbol{\xi}^*]$ for every admissible strategy for the second investor $\boldsymbol{\eta}$. In this case, $\boldsymbol{\xi}^*$ and $\boldsymbol{\eta}^*$ are called *optimal strategies* (for the first and the second investor, respectively).
This model can be viewed as a two-investor extension of the model in Obizhaeva and Wang (2013). The special case $\tilde{\gamma} = 0$ is analyzed in Schöneborn's (2008) doctoral thesis. The general case $\tilde{\gamma} \ge 0$ was first studied in 2013 in an early version of Schied and Zhang (2017), and subsequently in Zhang's (2014) doctoral thesis. The current version of Schied and Zhang (2017) extends the model to a large class of decay kernels G and arbitrary time grids $0 = t_1 < t_2 < \cdots < t_{n+1} = T$; the results in the next section are special cases of the results in this paper.

2.2 Predatory trading and transaction costs

The model admits a unique Nash equilibrium, which can be expressed in closed form. To this end, define the lower triangular, $(n+1)$-dimensional square matrix

$$\tilde{M} := \begin{pmatrix} 1/2 & 0 & 0 & \cdots & 0 & 0 \\ e^{-\rho(t_1-t_0)} & 1/2 & 0 & \cdots & 0 & 0 \\ e^{-\rho(t_2-t_0)} & e^{-\rho(t_2-t_1)} & 1/2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ e^{-\rho(t_{n-1}-t_0)} & e^{-\rho(t_{n-1}-t_1)} & e^{-\rho(t_{n-1}-t_2)} & \cdots & 1/2 & 0 \\ e^{-\rho(t_n-t_0)} & e^{-\rho(t_n-t_1)} & e^{-\rho(t_n-t_2)} & \cdots & e^{-\rho(t_n-t_{n-1})} & 1/2 \end{pmatrix}.$$

Let $M := \tilde{M} + \tilde{M}^\top$. (The matrix M is called a Kac-Murdock-Szegö matrix, in reference to Kac et al., 1953). Let I denote the $(n+1)$-dimensional identity matrix, and $\mathbf{1}$ the $(n+1)$-dimensional column vector containing only ones. Define the column vectors

$$\boldsymbol{v} := (M + \tilde{M} + 2\tilde{\gamma}I)^{-1}\mathbf{1} \qquad \text{and} \qquad \boldsymbol{w} := (M - \tilde{M} + 2\tilde{\gamma}I)^{-1}\mathbf{1}.$$

By Lemma 3.2 in Schied and Zhang (2017), both vectors are well-defined; furthermore, $\mathbf{1}^\top\boldsymbol{v} \neq 0$ and $\mathbf{1}^\top\boldsymbol{w} \neq 0$. The following result is due to Schied and Zhang (2017).

Theorem 2.2. (Theorem 2.5 in Schied and Zhang, 2017). *A unique Nash equilibrium $(\boldsymbol{\xi}^*, \boldsymbol{\eta}^*)$ exists. The optimal strategies are deterministic and given by*

$$\boldsymbol{\xi}^* = \frac{x^0 + y^0}{2(\mathbf{1}^\top\boldsymbol{v})}\boldsymbol{v} + \frac{x^0 - y^0}{2(\mathbf{1}^\top\boldsymbol{w})}\boldsymbol{w} \qquad \textit{and} \qquad \boldsymbol{\eta}^* = \frac{x^0 + y^0}{2(\mathbf{1}^\top\boldsymbol{v})}\boldsymbol{v} - \frac{x^0 - y^0}{2(\mathbf{1}^\top\boldsymbol{w})}\boldsymbol{w}.$$

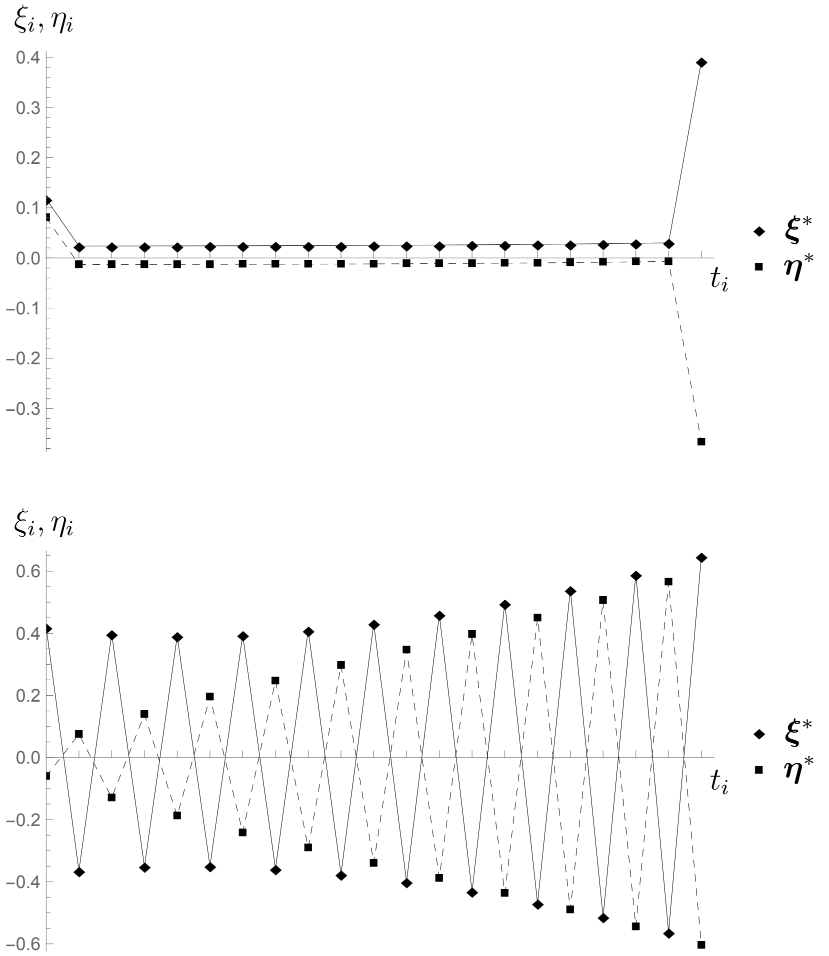

Figure 2.1: Optimal strategies $\boldsymbol{\xi}^*$ (solid line) and $\boldsymbol{\eta}^*$ (dashed line) for $\tilde{\gamma} = 1/4$ (top) and $\tilde{\gamma} = 0$ (bottom). Parameters: $n = 20$, $x^0 = 1$, $y^0 = -1/2$ and $\rho = T = 1$.

Remark 2.3. *A decay kernel G is said to be positive definite if*

$$\sum_{i,j=1}^{m} G(|t_i - t_j|) z_i z_j \geq 0$$

for all $m \in \mathbb{N}$, $t_1, t_2, \ldots, t_m \in [0, \infty)$ and $z_1, z_2, \ldots, z_m \in \mathbb{R}$. It is said to be strictly positive definite if equality holds only for $z_1 = z_2 = \cdots = z_m = 0$.
Theorem 2.2 is valid for all strictly positive definite decay kernels G and all time grids $0 = t_1 < t_2 < \cdots < t_{n+1} = T$, if one replaces each $e^{-\rho(t_i - t_j)}$ in the matrix $\tilde{M}$ with $G(|t_i - t_j|)$ (Schied and Zhang, 2017, Theorem 2.5).

Figure 2.1 shows $\boldsymbol{\xi}^*$ and $\boldsymbol{\eta}^*$ for different levels $\tilde{\gamma}$ of transaction costs. If $\tilde{\gamma}$ is large, there is little interaction between the two investors, because individual transaction costs dominate costs from price impact. The equilibrium strategies look roughly similar to the optimal strategy in the corresponding single-investor model (compare Figure 2 in Obizhaeva and Wang, 2013).

Things change if $\tilde{\gamma}$ is small (or zero): Both investors pile up trading volume by oscillating between buy and sell orders, throwing the asset back and forth as if it were a hot potato. Schied and Zhang (2017) provide an economic explanation:

> "The dominant form of interaction between two players is *predatory trading*, which consists in the exploitation of price impact generated by another agent. [...] Since predators prey on the drift created by the price impact of a large trade, protection against predatory trading requires the erasion of previously created price impact. Under transient price impact, the price impact of an earlier trade [...] can be erased by placing an order [...] of the opposite side. In this sense, oscillating strategies can be understood as a protection against predatory trading by opponents." (p. 10)

As mentioned above, oscillations only occur if $\tilde{\gamma}$ is small. This is made precise in the following theorem. It is due to Schied and Zhang (2017).

Theorem 2.4. (Theorem 2.7 in Schied and Zhang, 2017). *The following conditions are equivalent:*

(i) For every $n \in \mathbb{N}$ and $\rho > 0$, all components of $\boldsymbol{v}$ are nonnegative.

(ii) For every $n \in \mathbb{N}$ and $\rho > 0$, all components of $\boldsymbol{w}$ are nonnegative.

(iii) $\tilde{\gamma} \geq 1/4$.

Remark 2.5. *Theorem 2.4 remains valid if a positive constant is added to the kernel $G(t) = e^{-\rho t}$, and the equivalence of (ii) and (iii) even holds for all positive definite decay kernels G that are continuous, strictly positive and log-convex (Schied and Zhang, 2017, Theorem 2.7).*

Even if transaction costs are zero, oscillatory trading (as in the lower part of Figure 2.1) is expensive because of the price impact it generates. In fact, it can be so expensive that both investors would benefit if $\tilde{\gamma}$ were increased. For an illustration, let $n = 500$, $x^0 = y^0 = 10$ and $\rho = T = 1$. Let $(\boldsymbol{\xi}_0^*, \boldsymbol{\eta}_0^*)$ denote the optimal strategies for $\tilde{\gamma} = 0$; and let $(\boldsymbol{\xi}_{1/4}^*, \boldsymbol{\eta}_{1/4}^*)$ denote the optimal strategies for $\tilde{\gamma} = 1/4$.
Suppose actual transaction costs are zero, i.e., $\tilde{\gamma} = 0$. The only Nash equilibrium is $(\boldsymbol{\xi}_0^*, \boldsymbol{\eta}_0^*)$, but it is interesting to see what happens if one or both investors deviate. For the sake of the argument, assume that the first investor may only pursue $\boldsymbol{\xi}_0^*$ or $\boldsymbol{\xi}_{1/4}^*$, and the second investor may only pursue $\boldsymbol{\eta}_0^*$ or $\boldsymbol{\eta}_{1/4}^*$. Four scenarios emerge. The expected costs of both investors (under the assumption $S^0(0) = 0$) for each scenario are straightforward to compute. Expected costs in equilibrium are in italics.

	$\boldsymbol{\eta}_0^*$	$\boldsymbol{\eta}_{1/4}^*$
$\boldsymbol{\xi}_0^*$	*(75, 75)*	(65, 77)
$\boldsymbol{\xi}_{1/4}^*$	(77, 65)	(70, 70)

Expected costs $(J[\boldsymbol{\xi} \mid \boldsymbol{\eta}], J[\boldsymbol{\eta} \mid \boldsymbol{\xi}])$ if $\tilde{\gamma} = 0$.

This is a prisoner's dilemma. Both investors would benefit from agreeing to pursue $(\boldsymbol{\xi}^*_{1/4}, \boldsymbol{\eta}^*_{1/4})$. But both would have an incentive to deviate. If the second investor pursues $\boldsymbol{\eta}^*_{1/4}$, the first investor minimizes costs by pursuing $\boldsymbol{\xi}^*_0$. And if the first investor pursues $\boldsymbol{\xi}^*_{1/4}$, the second investor minimizes costs by pursuing $\boldsymbol{\eta}^*_0$. They end up with $(\boldsymbol{\xi}^*_0, \boldsymbol{\eta}^*_0)$ and are both worse off.
Now suppose instead that actual transaction costs are $\tilde{\gamma} = 1/4$. Recalculating expected costs shows that the prisoner's dilemma vanishes because predatory trading is now prohibitively expensive:

	$\boldsymbol{\eta}^*_0$	$\boldsymbol{\eta}^*_{1/4}$
$\boldsymbol{\xi}^*_0$	(1289, 1289)	(1280, 81)
$\boldsymbol{\xi}^*_{1/4}$	(81, 1280)	*(74, 74)*

Expected costs $(J[\boldsymbol{\xi} \,|\, \boldsymbol{\eta}], J[\boldsymbol{\eta} \,|\, \boldsymbol{\xi}])$ if $\tilde{\gamma} = 1/4$.

Notice that expected costs in equilibrium are lower for $\tilde{\gamma} = 1/4$ than for $\tilde{\gamma} = 0$. The reason is that investors no longer have to pursue oscillating strategies to protect against predatory trading. The cost savings from this more than outweigh the additional costs from transaction costs.
Transaction costs can come in the form of a financial market tax. In the current example, such a tax would reduce investors' expected costs (assuming that transaction costs without the tax are zero). It would have the additional benefits of raising tax income and of "calming the market" by eliminating orders that serve only as protection against predatory trading. This is the key economic insight of this chapter. The following section shows that it does not depend on the ad hoc assumption that only two strategies are available to each investor.

2.3 High frequency limit of expected costs

Both investors' expected costs depend on the model parameters n, x^0, y^0, ρ and T. To make the problem tractable, only the "high frequency" limit $n \to \infty$ will be analyzed.
Fix $x^0, y^0 \in \mathbb{R}$ and $\rho, T > 0$ and $n \in \{2, 3, \dots\}$. To make the dependence on n explicit, denote the expected costs functional as J_n, and the optimal strategies from Theorem 2.2 as $\boldsymbol{\xi}^*_{\boldsymbol{n}} := (\xi^n_1, \xi^n_2, \dots, \xi^n_{n+1})$ and $\boldsymbol{\eta}^*_{\boldsymbol{n}} := (\eta^n_1, \eta^n_2, \dots, \eta^n_{n+1})$.
Optimal strategies for different n are easier to compare after they have been converted to a step function on $[0, T]$: Define $m(t) := \lceil t\, n/T \rceil$ and

$$X_n(t) := x^0 - \sum_{i=1}^{m(t)} \xi^n_i \qquad \text{and} \qquad Y_n(t) := y^0 - \sum_{i=1}^{m(t)} \eta^n_i, \qquad t \in [0, T].$$

The value $X_n(t)$ is the first investor's net amount remaining at time t. Figure 2.2 shows X_n for the optimal strategies $\boldsymbol{\xi}^*_{\boldsymbol{n}}$ from Figure 2.1.

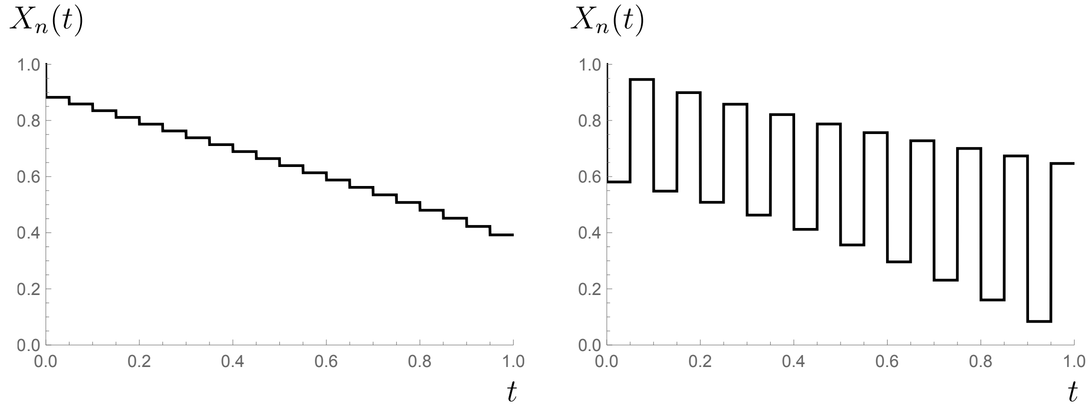

Figure 2.2: Remaining net amount X_n corresponding to the optimal strategy $\boldsymbol{\xi_n^*}$ for $\tilde{\gamma} = 1/4$ (left) and $\tilde{\gamma} = 0$ (right). All parameters are as in Figure 2.1.

It will be shown in Section 2.5 that X_n exhibits the following limit behavior as $n \to \infty$: For every $\tilde{\gamma} > 0$, the function X_n converges pointwise on $(0, T)$ to the same smooth function. For $\tilde{\gamma} = 0$, the function X_n does not converge on $(0, T)$, but oscillates between two smooth functions. The same is true for Y_n.
As suggested by the limit behavior of X_n and Y_n, expected costs converge to the same value for every $\tilde{\gamma} > 0$, and diverge with two cluster points for $\tilde{\gamma} = 0$. One reason why expected costs (in equilibrium) are identical for every $\tilde{\gamma} > 0$ is that transaction costs only apply to jumps ΔX_n and ΔY_n. In the limit, transaction costs of this form can be avoided by trading only infinitesimal amounts.
The limit behavior of the first investor's expected costs is made precise in the following theorem. Its tedious proof is the main mathematical contribution of this chapter and can be found in Section 2.5. Notice that the special case $x^0 = -y^0$ has already been proven in Zhang's (2014) doctoral thesis.
Define three functions $c_+, c_0^1, c_0^2 \colon (0, \infty) \times \mathbb{R}^2 \to \mathbb{R}$ via

$$\begin{aligned}
c_+(r, x, y) &\coloneqq \frac{(x+y)^2(36e^{6r}(8r+13) - 60e^{3r} - 3)}{16(2e^{3r}(3r+5) - 1)^2} + \frac{(x+y)(x-y)}{2(r+1)} + \frac{(x-y)^2}{16(r+1)^2}, \\
c_0^1(r, x, y) &\coloneqq \frac{(x+y)^2(6e^{6r} + 3)}{2(2e^{6r}(3r+5) + e^{3r} + 3r + 7)} + \frac{(x+y)(x-y)}{2(e^{-r} + r + 1)}, \\
c_0^2(r, x, y) &\coloneqq \frac{(x+y)^2(6e^{6r} - 3)}{2(2e^{6r}(3r+5) - 3e^{3r} - 3r - 7)} + \frac{(x+y)(x-y)}{2(-e^{-r} + r + 1)}.
\end{aligned}$$

Theorem 2.6.

(i) If $\tilde{\gamma} > 0$, then

$$\lim_{n \to \infty} J_n[\boldsymbol{\xi_n^*} \,|\, \boldsymbol{\eta_n^*}] = c_+(\rho T, x^0, y^0).$$

(ii) If $\tilde{\gamma} = 0$, then

$$\lim_{\substack{n \to \infty \\ n \,\text{even}}} J_n[\boldsymbol{\xi_n^*} \,|\, \boldsymbol{\eta_n^*}] = c_0^1(\rho T, x^0, y^0) \qquad \textit{and} \qquad \lim_{\substack{n \to \infty \\ n \,\text{odd}}} J_n[\boldsymbol{\xi_n^*} \,|\, \boldsymbol{\eta_n^*}] = c_0^2(\rho T, x^0, y^0).$$

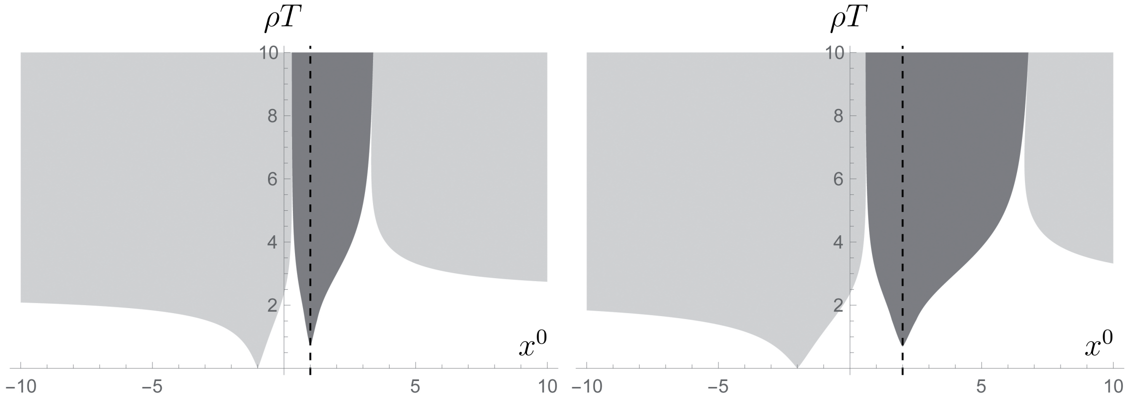

Figure 2.3: Given $y^0 = 1$ (left, dashed line) and $y^0 = 2$ (right, dashed line), transaction costs are advantageous if $(\rho T, x^0)$ lies in the dark gray area, and disadvantageous if $(\rho T, x^0)$ lies in the light gray area.

Since both investors are identical except for their execution constraint, the second investor's expected costs are obtained simply by interchanging x^0 and y^0 in the equations above.

Say that positive transaction costs $\tilde{\gamma} > 0$ are *advantageous* if both investors' expected costs for $\tilde{\gamma} > 0$ are smaller than the limit inferior of expected costs for $\tilde{\gamma} = 0$. In view of Theorem 2.6, this is equivalent to

$$\begin{aligned} c_+(\rho T, x^0, y^0) &< \min\{c_0^1(\rho T, x^0, y^0), c_0^2(\rho T, x^0, y^0)\} \quad \text{and} \\ c_+(\rho T, y^0, x^0) &< \min\{c_0^1(\rho T, y^0, x^0), c_0^2(\rho T, y^0, x^0)\}. \end{aligned}$$

Similarly, say that positive transaction costs $\tilde{\gamma} > 0$ are *disadvantageous* if both investors' expected costs for $\tilde{\gamma} > 0$ are larger than the limit superior of expected costs for $\tilde{\gamma} = 0$, i.e.,

$$\begin{aligned} c_+(\rho T, x^0, y^0) &> \max\{c_0^1(\rho T, x^0, y^0), c_0^2(\rho T, x^0, y^0)\} \quad \text{and} \\ c_+(\rho T, y^0, x^0) &> \max\{c_0^1(\rho T, y^0, x^0), c_0^2(\rho T, y^0, x^0)\}. \end{aligned}$$

It is easy to check numerically for given parameters $(\rho T, x^0, y^0)$ whether transaction costs are advantageous, disadvantageous or neither. Figure 2.3 illustrates some general observations: Transaction costs are advantageous if ρT is sufficiently large and x^0 and y^0 are close, and disadvantageous if x^0 and y^0 have different signs or if the difference between x^0 and y^0 is large. Transaction costs seem to address a very specific failure: the failure to coordinate if both investors face similar execution constraints. If x^0 and y^0 have different signs, investors can coordinate without transaction costs. If x^0 is much larger than y^0 or vice versa, the larger investor's influence is unavoidable and the smaller investor's influence is negligible. Introducing transaction costs does not change this.

Corroborating these numerical findings with analytic results is difficult. The following result considers the special case $x^0 = y^0$.

Corollary 2.7. *For every $x \in \mathbb{R} \setminus \{0\}$, if $r > \log(4 + \sqrt{62}/2)/3 \approx 0.69$, then*

$$c_+(r, x, x) < \min\{c_0^1(r, x, x), c_0^2(r, x, x)\}.$$

Proof. Let $x \in \mathbb{R} \setminus \{0\}$ and $r > 0$. It holds that

$$\begin{aligned} c_+(r, x, x) &= \frac{(2x)^2(36e^{6r}(8r + 13) - 60e^{3r} - 3)}{16(2e^{3r}(3r + 5) - 1)^2}, \\ c_0^1(r, x, x) &= \frac{(2x)^2(6e^{6r} + 3)}{2(2e^{6r}(3r + 5) + e^{3r} + 3r + 7)} \qquad \text{and} \\ c_0^2(r, x, y) &= \frac{(2x)^2(6e^{6r} - 3)}{2(2e^{6r}(3r + 5) - 3e^{3r} - 3r - 7)}. \end{aligned}$$

There is no loss of generality in letting $x = 1/2$. Consider the difference

$$\begin{aligned} &c_0^1(r, 1/2, 1/2) - c_0^2(r, 1/2, 1/2) \\ &= \frac{6e^{6r} + 3}{2(2e^{6r}(3r + 5) + e^{3r} + 3r + 7)} - \frac{6e^{6r} - 3}{2(2e^{6r}(3r + 5) - 3e^{3r} - 3r - 7)} \\ &= -\frac{3e^{3r}(2e^{3r} + 1)^2}{(2e^{6r}(3r + 5) + e^{3r} + 3r + 7)(2e^{6r}(3r + 5) - 3e^{3r} - 3r - 7)}. \end{aligned}$$

Rewrite the second factor in the denominator to see that the expression above is negative:

$$\begin{aligned} &2e^{6r}(3r + 5) - 3e^{3r} - 3r - 7 \\ &= (e^r - 1)(e^{2r} + e^r + 1)(10e^{3r} + 7) + 3r(2e^{6r} - 1) > 0. \end{aligned}$$

Hence $c_0^1(r, x, x) < c_0^2(r, x, x)$. It remains to show that if $r > \log(4 + \sqrt{62}/2)/3$, then

$$\frac{36e^{6r}(8r + 13) - 60e^{3r} - 3}{16(2e^{3r}(3r + 5) - 1)^2} < \frac{6e^{6r} + 3}{2(2e^{6r}(3r + 5) + e^{3r} + 3r + 7)}.$$

Multiply by both denominators to see that this is the case if and only if

$$\begin{aligned} 0 < {} &16(2e^{3r}(3r + 5) - 1)^2 \, (6e^{6r} + 3) \\ &- 2(2e^{6r}(3r + 5) + e^{3r} + 3r + 7) \, (36e^{6r}(8r + 13) - 60e^{3r} - 3). \end{aligned}$$

A tedious rearrangement of terms shows that the expression on the right hand side equals

$$6(2e^{3r} + 1)^2(3r + 5)\Big(2e^{6r} - e^{3r}\frac{48r + 79}{3r + 5} + \frac{3r + 15}{3r + 5}\Big).$$

Since $48r + 79 < 16(3r + 5)$ and $3r + 15 > 3r + 5$, this expression is larger than

$$6(2e^{3r} + 1)^2(3r + 5)(2e^{6r} - 16e^{3r} + 1).$$

The real-valued function $y \mapsto 2y^2 - 16y + 1$ has the two roots $y_1 := 4 - \sqrt{62}/2$

and $y_2 := 4+\sqrt{62}/2$. Since $y_1 < 1 < y_2$, the real-valued function $z \mapsto 2e^{6z} - 16e^{3z} + 1$ has exactly one positive root $\log(y_2)/3$.
Hence, for every $r > \log(y_2)/3 = \log(4+\sqrt{62}/2)/3$, it holds that

$$\begin{aligned}0 &< 6(2e^{3r}+1)^2(3r+5)(2e^{6r}-16e^{3r}+1)\\ &< 16(2e^{3r}(3r+5)-1)^2\,(6e^{6r}+3)\\ &\quad - 2(2e^{6r}(3r+5)+e^{3r}+3r+7)\,(36e^{6r}(8r+13)-60e^{3r}-3). \qquad \square\end{aligned}$$

Define the *half life* τ as the time it takes until half of the initial price impact $G(0)$ of an order has decayed. In the present model, τ satisfies the equality $e^{-\rho\tau} = 1/2$, hence $\tau = \log(2)/\rho \approx 0.69/\rho$.
The threshold $\rho T \approx 0.69$ in Corollary 2.7 therefore suggests the following rule of thumb, given that n is large and $x^0 \approx y^0$: Transaction costs are advantageous for both investors if the trading horizon T exceeds the half life τ.

2.4 From discrete to continuous time

Since the optimal strategies converge for every $\tilde{\gamma} > 0$ as $n \to \infty$, it seems worthwhile to explore the continuous time version of the model above, as in Section 3.3 of Zhang's (2014) doctoral thesis. One would assume that optimal strategies in continuous time are simply the limit of optimal strategies in the n-step models. This is correct in the corresponding single-investor model (Obizhaeva and Wang, 2013, Propositions 2 and 3), but turns out to be only partially correct here.
As before, suppose two investors trade over the time period $[0,T]$ a financial asset whose unaffected price process is a right-continuous martingale $S^0 = (S^0(t))_{t\in[0,T]}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ satisfying the usual conditions, with $\mathcal{F}_0$ assumed $\mathbb{P}$-trivial.
Both investors must trade fixed net amounts x^0 and y^0 until time T. Trading may now occur at any time $t \in [0,T]$.
In discrete time, a strategy was a list of trades $\boldsymbol{\xi_n} = (\xi_1^n, \xi_2^n, \dots, \xi_{n+1}^n)$, which was then converted to a step function X_n specifying the remaining net amount. In the present situation, a strategy X specifies the remaining net amount directly: A right-continuous, adapted and bounded process $X = (X(t))_{t\in[0-,T]}$ is called an *(admissible) strategy for the first investor* if it has finite and $\mathbb{P}$-a.s. bounded total variation and the execution constraint $X(0-) = x^0$ and $X(T) = 0$ is satisfied $\mathbb{P}$-a.s. Notice that X may jump immediately at time 0.
An *(admissible) strategy for the second investor* Y is defined similarly. The execution constraint becomes $Y(0-) = y^0$ and $Y(T) = 0$.
If the investors pursue admissible strategies X and Y, the asset price evolves according to

$$S(t) = S(t; X, Y) := S^0(t) + \int_{[0,t)} e^{-\rho(t-s)}\,\mathrm{d}X(s) + \int_{[0,t)} e^{-\rho(t-s)}\,\mathrm{d}Y(s), \quad t \in [0,T].$$

The *first investor's expected costs* are

$$J[X \,|\, Y] \coloneqq \mathbb{E}\Big[\frac{1}{2}\int_{[0,T]}\int_{[0,T]} e^{-\rho|t-s|}\,\mathrm{d}X(s)\,\mathrm{d}X(t) + \int_{[0,T]}\int_{[0,t)} e^{-\rho(t-s)}\,\mathrm{d}Y(s)\,\mathrm{d}X(t) \\ + \frac{1}{2}\sum_{t\in[0,T]}\Delta X(t)\Delta Y(t) + \tilde{\gamma}\sum_{t\in[0,T]}\Delta X(t)^2\Big].$$

The *second investor's expected costs* are $J[Y \,|\, X]$.

Remark 2.8. *The definition of expected costs follows naturally from its discrete time counterpart, as shown in Lemma 3.3.2 of Zhang's (2014) doctoral thesis. Choose admissible strategies X and Y. For $n \in \{2, 3, \dots\}$, let $t_i \coloneqq (i-1)T/n$ for $i = 1, 2, \dots, n+1$. Define vectors $\boldsymbol{\xi_n} \coloneqq (\xi_1^n, \xi_2^n, \dots, \xi_{n+1}^n)$ and $\boldsymbol{\eta_n} \coloneqq (\eta_1^n, \eta_2^n, \dots, \eta_{n+1}^n)$ via*

$$\begin{aligned} \xi_1^n &\coloneqq X(0) - X(0-) \quad \text{and} \quad \xi_i^n \coloneqq X(t_i) - X(t_{i-1}) \quad \text{for } i = 2, 3, \dots, n+1, \\ \eta_1^n &\coloneqq Y(0) - Y(0-) \quad \text{and} \quad \eta_i^n \coloneqq Y(t_i) - Y(t_{i-1}) \quad \text{for } i = 2, 3, \dots, n+1. \end{aligned}$$

Then $\boldsymbol{\xi_n}$ and $\boldsymbol{\eta_n}$ are admissible strategies for the first and the second investor in the n-step model of Section 2.1, and the expected costs $(J_n[\boldsymbol{\xi_n} \,|\, \boldsymbol{\eta_n}], J_n[\boldsymbol{\eta_n} \,|\, \boldsymbol{\xi_n}])$ converge to $(J[X \,|\, Y], J[Y \,|\, X])$ as $n \to \infty$.

A pair (X^*, Y^*) of admissible strategies for the first and the second investor is called a *Nash equilibrium* if $J[X^* \,|\, Y^*] \le J[X \,|\, Y^*]$ for every admissible strategy for the first investor X, and $J[Y^* \,|\, X^*] \le J[Y \,|\, X^*]$ for every admissible strategy for the second investor Y. In this case, X^* and Y^* are called *optimal strategies* (for the first and the second investor, respectively).
The following theorem provides a full characterization of Nash equilibria in this model. It was first published in Zhang's (2014) doctoral thesis. A revised proof is given by Schied et al. (2017, Theorem 4.5). Define two functions $V, W\colon [0,T) \to \mathbb{R}$,

$$V(t) \coloneqq \frac{e^{3\rho T}(6\rho(T-t)+4) - 4e^{3\rho t}}{2e^{3\rho T}(3\rho T+5) - 1} \qquad \text{and} \qquad W(t) \coloneqq \frac{\rho(T-t)+1}{\rho T+1}.$$

Theorem 2.9. (Theorem 3.3.6 in Zhang, 2014).

(i) If $\tilde{\gamma} = 1/4$, then a unique Nash equilibrium (X^*, Y^*) exists. The optimal strategies are deterministic and given by $X^*(0-) = x^0$, $Y^*(0-) = y^0$ and $X^*(T) = Y^*(T) = 0$ and, for every $t \in [0,T)$,

$$X^*(t) = \frac{x^0+y^0}{2}\,V(t) + \frac{x^0-y^0}{2}\,W(t)$$

and

$$Y^*(t) = \frac{x^0+y^0}{2}\,V(t) - \frac{x^0-y^0}{2}\,W(t).$$

(ii) If $\tilde{\gamma} \ne 1/4$, then a Nash equilibrium exists if and only if $x^0 = y^0 = 0$.

Notice that existence of a unique Nash equilibrium for the case $\tilde{\gamma} \neq 1/4$ and $x^0 = y^0 = 0$ is not claimed in Theorem 3.3.6 in Zhang's (2014) doctoral thesis, but follows easily from the proof.
The following result extends Zhang's (2014) analysis. It confirms that if $\tilde{\gamma} = 1/4$, then expected costs in the continuous time model coincide with the limit of expected costs in the n-step models.

Proposition 2.10. *If $\tilde{\gamma} = 1/4$, then the first investor's expected costs in equilibrium, $J[X^* \,|\, Y^*]$, equal $c_+(\rho T, x^0, y^0)$.*

Proof. Conclude from the definitions of V and W that X^* and Y^* are smooth on $[0, T)$, with

$$\mathrm{d}X^*(t) = -\rho\Big(\frac{3(x^0 + y^0)(e^{3\rho T} + 2e^{3\rho t})}{2e^{3\rho T}(3\rho T + 5) - 1} + \frac{x^0 - y^0}{2(\rho T + 1)}\Big)\,\mathrm{d}t$$

and

$$\mathrm{d}Y^*(t) = -\rho\Big(\frac{3(x^0 + y^0)(e^{3\rho T} + 2e^{3\rho t})}{2e^{3\rho T}(3\rho T + 5) - 1} - \frac{x^0 - y^0}{2(\rho T + 1)}\Big)\,\mathrm{d}t.$$

In addition,

$$\Delta X^*(0) = \Delta Y^*(0) = -\frac{3}{2}\,\frac{(x^0 + y^0)(2e^{3\rho T} + 1)}{2e^{3\rho T}(3\rho T + 5) - 1}$$

and

$$\Delta X^*(T) = -\Delta Y^*(T) = -\frac{x^0 - y^0}{2(\rho T + 1)}.$$

Remark 2.4 in Lorenz and Schied (2013) explains how to correctly treat the jumps at $t = 0-$ and $t = T$ in the integrals that follow. Obtain

$$\begin{aligned}
&\int_{[0,T]}\int_{[0,T]} e^{-\rho|t-s|}\,\mathrm{d}X^*(s)\,\mathrm{d}X^*(t)\\
&= \int_0^T\int_0^T e^{-\rho|t-s|}\,\mathrm{d}X^*(s)\,\mathrm{d}X^*(t) + 2\Delta X^*(0)\int_0^T e^{-\rho t}\,\mathrm{d}X^*(t)\\
&\quad + 2\Delta X^*(T)\int_0^T e^{-\rho(T-t)}\,\mathrm{d}X^*(t) + \Delta X^*(0)^2 + 2e^{-\rho T}\Delta X^*(0)\Delta X^*(T) + \Delta X^*(T)^2
\end{aligned}$$

and, using the fact that $\Delta Y^*(0) = \Delta X^*(0)$,

$$\begin{aligned}
&\int_{[0,T]}\int_{[0,t)} e^{-\rho(t-s)}\,\mathrm{d}Y^*(s)\,\mathrm{d}X^*(t)\\
&= \int_0^T\int_0^t e^{-\rho(t-s)}\,\mathrm{d}Y^*(s)\,\mathrm{d}X^*(t) + \Delta X^*(0)\int_0^T e^{-\rho t}\,\mathrm{d}X^*(t)
\end{aligned}$$

$$+\,\Delta X^*(T)\int_0^T e^{-\rho(T-t)}\,\mathrm{d}Y^*(t)+e^{-\rho T}\Delta X^*(0)\Delta X^*(T).$$

Plug into the definition of expected costs to obtain

$$\begin{aligned}J[X^*\,|\,Y^*]=\frac{1}{2}\int_0^T\int_0^T e^{-\rho|t-s|}\,\mathrm{d}X^*(s)\,\mathrm{d}X^*(t)+\int_0^T\int_0^t e^{-\rho(t-s)}\,\mathrm{d}Y^*(s)\,\mathrm{d}X^*(t)\\+2\Delta X^*(0)\int_0^T e^{-\rho t}\,\mathrm{d}X^*(t)\\+\Delta X^*(T)\Big(\int_0^T e^{-\rho(T-t)}\,\mathrm{d}X^*(t)+\int_0^T e^{-\rho(T-t)}\,\mathrm{d}Y^*(t)\Big)\\+\frac{5}{4}\Delta X^*(0)^2+2e^{-\rho T}\Delta X^*(0)\Delta X^*(T)+\frac{1}{4}\Delta X^*(T)^2.\end{aligned}$$

Lengthy computations yield

$$\begin{aligned}&\frac{1}{2}\int_0^T\int_0^T e^{-\rho|t-s|}\,\mathrm{d}X^*(s)\,\mathrm{d}X^*(t)\\&=\frac{1}{2}\Big(\int_0^T\int_0^t e^{-\rho(t-s)}\,\mathrm{d}X^*(s)\,\mathrm{d}X^*(t)+\int_0^T\int_t^T e^{-\rho(s-t)}\,\mathrm{d}X^*(s)\,\mathrm{d}X^*(t)\Big)\\&=3\big(3\rho Te^{6\rho T}-e^{3\rho T}+1\big)\Big(\frac{x^0+y^0}{2e^{3\rho T}(3\rho T+5)-1}\Big)^2\\&\quad+\frac{1}{2}\big((12\rho T-7)e^{3\rho T}+6e^{2\rho T}-2+3e^{-\rho T}\big)\Big(\frac{x^0+y^0}{2e^{3\rho T}(3\rho T+5)-1}\Big)\Big(\frac{x^0-y^0}{2(\rho T+1)}\Big)\\&\quad+\big(\rho T-1+e^{-\rho T}\big)\Big(\frac{x^0-y^0}{2(\rho T+1)}\Big)^2\end{aligned}$$

and

$$\begin{aligned}&\int_0^T\int_0^t e^{-\rho(t-s)}\,\mathrm{d}Y^*(s)\,\mathrm{d}X^*(t)\\&=3\big(3\rho Te^{6\rho T}-e^{3\rho T}+1\big)\Big(\frac{x^0+y^0}{2e^{3\rho T}(3\rho T+5)-1}\Big)^2\\&\quad+\frac{3}{2}\big(-e^{3\rho T}+2e^{2\rho T}-2+e^{-\rho T}\big)\Big(\frac{x^0+y^0}{2e^{3\rho T}(3\rho T+5)-1}\Big)\Big(\frac{x^0-y^0}{2(\rho T+1)}\Big)\\&\quad+\big(1-\rho T-e^{-\rho T}\big)\Big(\frac{x^0-y^0}{2(\rho T+1)}\Big)^2.\end{aligned}$$

Furthermore,

$$\begin{aligned}&2\Delta X^*(0)\int_0^T e^{-\rho t}\,\mathrm{d}X^*(t)\\&=9\big(2e^{6\rho T}-e^{3\rho T}-1\big)\Big(\frac{x^0+y^0}{2e^{3\rho T}(3\rho T+5)-1}\Big)^2\end{aligned}$$

$$+\,3\big(2e^{3\rho T}-2e^{2\rho T}+1-e^{-\rho T}\big)\Big(\frac{x^0+y^0}{2e^{3\rho T}(3\rho T+5)-1}\Big)\Big(\frac{x^0-y^0}{2(\rho T+1)}\Big)$$

and

$$\begin{aligned}&\Delta X^*(T)\Big(\int_0^T e^{-\rho(T-t)}\,\mathrm{d}X^*(t)+\int_0^T e^{-\rho(T-t)}\,\mathrm{d}Y^*(t)\Big)\\&=3\big(3e^{3\rho T}-2e^{2\rho T}-e^{-\rho T}\big)\Big(\frac{x^0+y^0}{2e^{3\rho T}(3\rho T+5)-1}\Big)\Big(\frac{x^0-y^0}{2(\rho T+1)}\Big).\end{aligned}$$

Finally,

$$\begin{aligned}&\frac{5}{4}\Delta X^*(0)^2+2e^{-\rho T}\Delta X^*(0)\Delta X^*(T)+\frac{1}{4}\Delta X^*(T)^2\\&=\frac{45}{16}\big(4e^{6\rho T}+4e^{3\rho T}+1\big)\Big(\frac{x^0+y^0}{2e^{3\rho T}(3\rho T+5)-1}\Big)^2\\&\quad+3\big(2e^{2\rho T}+e^{-\rho T}\big)\Big(\frac{x^0+y^0}{2e^{3\rho T}(3\rho T+5)-1}\Big)\Big(\frac{x^0-y^0}{2(\rho T+1)}\Big)\\&\quad+\frac{1}{4}\Big(\frac{x^0-y^0}{2(\rho T+1)}\Big)^2.\end{aligned}$$

Plug in and simplify to obtain

$$\begin{aligned}J[X^*\,|\,Y^*]&=\frac{(x^0+y^0)^2(36e^{6\rho T}(8\rho T+13)-60e^{3\rho T}-3)}{16(2e^{3\rho T}(3\rho T+5)-1)^2}\\&\quad+\frac{(x^0+y^0)(x^0-y^0)}{2(\rho T+1)}+\frac{(x^0-y^0)^2}{16(\rho T+1)^2},\end{aligned}$$

as desired. □

Notice that this also shows that the second investor's expected costs in equilibrium, $J[Y^*\,|\,X^*]$, equal the limit of expected costs, $c_+(\rho T, y^0, x^0)$, in the n-step models.

Remark 2.11. *A Nash equilibrium in the present model exists if and only if $\tilde\gamma = 1/4$. This is inconvenient from an economic point of view, since there seems to be no reason why the case $\tilde\gamma = 1/4$ should be special. To avoid this, I argue that transaction costs should be defined differently in continuous time models.*
Trading is discrete by nature: continuous trading is an idealization. In practice, every continuous time trading strategy $X(t), t \in [0,T]$, must be executed as a sequence of block trades

$$X(t_1)-X(t_0),\;X(t_2)-X(t_1),\;\ldots,\;X(t_n)-X(t_{n-1}).$$

Consequently, transaction costs on financial markets can only apply to discrete trades. It is reasonable to take the same approach in discrete time models of financial markets, as in Section 2.1. In continuous time, however, "discrete" transaction *costs on block trades form an awkward couple with continuous time trading strategies: An investor can completely avoid them by pursuing a trading strategy that is absolutely continuous.*

I therefore believe that it is more appropriate to idealize transaction costs in the same way as trading strategies. They should apply "continuously" to the instantaneous rate of trading $\mathrm{d}X_t$ *(as suggested for instance by Almgren and Chriss, 2001, and Gatheral, 2010), and not to block trades* $\Delta X(t)$. *Applying transaction costs to the instantaneous rate of trading implies that optimal strategies cannot jump if transaction costs are nonzero, a consequence criticized by Obizhaeva and Wang (2013) because "these modifications limit us to a subset of feasible strategies, which is in general sub-optimal" (p. 12). In my view, their argument loses sight of the fact that continuous time models are idealized approximations of reality, as Carmona and Yang (2011) point out: "Although the position of the trader is a piecewise constant function of time, an absolutely-continuous function can offer a reasonable approximation" (p. 3). Applying transaction costs to the instantaneous rate of trading ensures that these costs cannot be avoided in the model, just as transaction costs on discrete trades cannot be avoided in real markets.*

2.5 Proof of Theorem 2.6

Fix $x^0, y^0 \in \mathbb{R}$ and $\rho, T > 0$ and $n \in \{2, 3, \dots\}$. Recall that

$$m(t) \coloneqq \left\lceil t \, \frac{n}{T} \right\rceil,$$

and that $\boldsymbol{v} = (v_1, v_2, \dots, v_{n+1})$ and $\boldsymbol{w} = (w_1, w_2, \dots, w_{n+1})$ are defined as

$$\boldsymbol{v} \coloneqq (M + \tilde{M} + 2\tilde{\gamma} I)^{-1}\mathbf{1} \qquad \text{and} \qquad \boldsymbol{w} \coloneqq (M - \tilde{M} + 2\tilde{\gamma} I)^{-1}\mathbf{1}.$$

Notice that both vectors implicitly depend on n. Define

$$V_n(t) \coloneqq 1 - \frac{1}{\mathbf{1}^\top \boldsymbol{v}} \sum_{i=1}^{m(t)} v_i \qquad \text{and} \qquad W_n(t) \coloneqq 1 - \frac{1}{\mathbf{1}^\top \boldsymbol{w}} \sum_{i=1}^{m(t)} w_i, \qquad t \in [0, T].$$

Analyzing V_n and W_n instead of X_n and Y_n has the advantage of eliminating x^0 and y^0 from subsequent calculations. The step functions X_n^* and Y_n^* corresponding to $\boldsymbol{\xi_n^*} = (\xi_1^n, \xi_2^n, \dots, \xi_{n+1}^n)$ and $\boldsymbol{\eta_n^*} = (\eta_1^n, \eta_2^n, \dots, \eta_{n+1}^n)$ are easily recovered from V_n and W_n. Indeed,

$$\begin{aligned} X_n^*(t) = x^0 - \sum_{i=1}^{m(t)} \xi_i^n &= \frac{x^0 + y^0}{2}\Big(1 - \frac{1}{\mathbf{1}^\top \boldsymbol{v}} \sum_{i=1}^{m(t)} v_i\Big) + \frac{x^0 - y^0}{2}\Big(1 - \frac{1}{\mathbf{1}^\top \boldsymbol{w}} \sum_{i=1}^{m(t)} w_i\Big) \\ &= \frac{x^0 + y^0}{2} V_n(t) + \frac{x^0 - y^0}{2} W_n(t). \end{aligned}$$

To shorten subsequent calculations, let

$$a := e^{-\rho T/n} \qquad \text{and} \qquad \theta := 2\tilde{\gamma} + 1/2.$$

By definition, $e^{-\rho(t_i - t_j)} = a^{i-j}$. Notice that $\theta \geq 1/2$, with equality if $\tilde{\gamma} = 0$. The critical value $\tilde{\gamma} = 1/4$ in Theorem 2.4 corresponds to $\theta = 1$.
Theorem 2.6 has already been proved for the special case $x^0 = -y^0$ in Zhang's (2014) doctoral thesis. In this case, calculations are simplified by the fact that equilibrium strategies only depend on $\boldsymbol{w}$ and not on $\boldsymbol{v}$. The following lemma collects auxiliary results obtained for this special case. They follow from Equation (3.16), Lemma 3.1.7, Proposition 3.1.19, Lemma 3.1.20 and Proposition 3.1.21 in Zhang (2014). Define four functions $f_\pm^W, g_\pm^W : (0,T) \to \mathbb{R}$ via

$$f_\pm^W(t) := \frac{1 + \rho(T-t) \pm e^{-\rho(T-t)}}{1 + \rho T + e^{-\rho T}} \qquad \text{and} \qquad g_\pm^W(t) := \frac{1 + \rho(T-t) \pm e^{-\rho(T-t)}}{1 + \rho T - e^{-\rho T}}.$$

Lemma 2.12. (Zhang, 2014).

(i) For every $\theta \geq 1/2$, the components of $\boldsymbol{w}$ are given as

$$w_i = \frac{(1-a)\theta + a\left(\frac{a(\theta-1)}{\theta}\right)^{n+1-i}}{\theta(\theta - a(\theta-1))} \qquad \textit{for } i = 1, 2, \ldots, n+1.$$

(ii) Suppose $\theta > 1/2$. Then $\lim_{n\to\infty} \mathbf{1}^\top \boldsymbol{w} = \rho T + 1$ and

$$\lim_{n\to\infty} W_n(t) = \frac{\rho(T-t)+1}{\rho T + 1}, \qquad t \in (0,T).$$

(iii) Suppose $\theta = 1/2$. Then

$$\lim_{\substack{n\to\infty \\ n\,\text{even}}} \mathbf{1}^\top \boldsymbol{w} = e^{-\rho T} + \rho T + 1 \qquad \textit{and} \qquad \lim_{\substack{n\to\infty \\ n\,\text{odd}}} \mathbf{1}^\top \boldsymbol{w} = -e^{-\rho T} + \rho T + 1.$$

Furthermore, for every $t \in (0,T)$, the sequence $(W_n(t))_{n=2,4,6,\ldots}$ has exactly two cluster points $f_\pm^W(t)$ and the sequence $(W_n(t))_{n=1,3,5,\ldots}$ has exactly two cluster points $g_\pm^W(t)$.

(iv) For every $\theta \geq 1/2$ and all admissible strategies $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, it holds that

$$J_n[\boldsymbol{\xi} \,|\, \boldsymbol{\eta}] = \frac{1}{2}\,\mathbb{E}\big[\boldsymbol{\xi}^\top (M + 2\tilde{\gamma} I)\,\boldsymbol{\xi} + \boldsymbol{\xi}^\top \tilde{M} \boldsymbol{\eta}\big].$$

The first step in the proof of Theorem 2.6 is to obtain results similar to Lemma 2.12 (i)–(iii) for $\boldsymbol{v}$ and V_n.
Define the $(n+1)$-dimensional square matrix $N := (1-a^2)\big(I + M^{-1}(\tilde{M} + 2\tilde{\gamma} I)\big)$. The inverse of M has a simple tridiagonal structure, see for instance Section 7.2,

Problems 12 and 13, in Horn and Johnson (2013):

$$M^{-1} = \frac{1}{1-a^2}\begin{pmatrix} 1 & -a & 0 & \cdots & 0 & 0 \\ -a & 1+a^2 & -a & \cdots & 0 & 0 \\ 0 & -a & 1+a^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+a^2 & -a \\ 0 & 0 & 0 & \cdots & -a & 1 \end{pmatrix}. \tag{2.1}$$

Thus

$$\begin{aligned} N &= (1-a^2)I + (1-a^2)M^{-1}\begin{pmatrix} \theta & 0 & 0 & \cdots & 0 \\ a & \theta & 0 & \cdots & 0 \\ a^2 & a & \theta & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a^n & a^{n-1} & a^{n-2} & \cdots & \theta \end{pmatrix} \\ &= \begin{pmatrix} 1-2a^2+\theta & -a\theta & \cdots & 0 \\ -a(\theta-1) & 1+a^2(\theta-2)+\theta & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ddots & 1+a^2(\theta-2)+\theta & -a\theta \\ 0 & \cdots & -a(\theta-1) & 1-a^2+\theta \end{pmatrix}. \end{aligned}$$

Let $r \coloneqq \big(a^4(\theta-2)^2 - 2a^2(\theta^2-\theta+2) + (\theta+1)^2\big)^{1/2}$. Define

$$\begin{aligned} b_\pm &\coloneqq \frac{\pm(1-a^2(\theta+2)+\theta)+r}{2r}, \\ c_\pm &\coloneqq \frac{\pm(1+(1-a^2)\theta)+r}{2r} \qquad\qquad \text{and} \\ d_\pm &\coloneqq \frac{1+a^2(\theta-2)+\theta \pm r}{2}. \end{aligned}$$

Define further

$$\delta_i \coloneqq b_+d_+^i + b_-d_-^i \qquad\qquad \text{for } i = 0, 1, \ldots, n,$$

and

$$\varphi_i \coloneqq c_+d_+^{n+2-i} + c_-d_-^{n+2-i} \qquad \text{for } i = 2, 3, \ldots, n+2.$$

Notice that $\delta_0 = \varphi_{n+2} = 1$. Finally, let

$$\delta \coloneqq (1-a^2+\theta)(b_+d_+^n + b_-d_-^n) - a^2\theta(\theta-1)(b_+d_+^{n-1} + b_-d_-^{n-1}).$$

Lemma 2.13. *For $i = 1, 2, \ldots, n$, the i-th leading principal minor of N equals δ_i.*

Proof. For $i = 1, 2, \ldots, n$, denote by f_i the i-th principal minor of N. Straightforward calculations show

$$f_1 = \delta_1 \qquad \text{and} \qquad f_2 = \delta_2. \tag{2.2}$$

Because of the tridiagonal structure of N, the remaining principal minors satisfy the following homogeneous linear difference equation of second order (see for instance Theorem 2.1 in El-Mikkawy, 2004):

$$f_i = (1 + a^2(\theta - 2) + \theta)f_{i-1} - a^2\theta(\theta - 1)f_{i-2}, \qquad i = 3, 4, \ldots, n. \tag{2.3}$$

The characteristic equation of (2.3) is

$$0 = x^2 - (1 + a^2(\theta - 2) + \theta)x + a^2\theta(\theta - 1).$$

It has the two roots d_+ and d_-.
Now show that r is real to conclude that d_+ and d_- are real. Recall that $\theta \geq 1/2$ and $a \in (0, 1)$. Define the mapping $f\colon [1/2, \infty) \to \mathbb{R}$ via

$$f(x) = x^2(a^2 - 1)^2 - 2x(2a^4 - a^2 - 1) + (2a^2 - 1)^2.$$

Refactoring shows that $r = f(\theta)^{1/2}$. For all $x \geq 1/2$, it holds that

$$\begin{aligned} f'(x) &= 2\big(- 2a^4 + a^2 + 1 + x(a^2 - 1)^2\big) \\ &\geq 2\big(- 2a^4 + a^2 + 1 + \frac{1}{2}(a^2 - 1)^2\big) \\ &= 3(1 - a^4) \end{aligned}$$

is nonnegative. Conclude that $f(x) \geq f(1/2) = 9/4(a^2 - 1)^2 + a^2$, hence $f(x)$ is nonnegative for all $x \geq 1/2$ and r is real. The general solution to the difference equation (2.3) is therefore given by $z_+d_+^i + z_-d_-^i$ for $z_+, z_- \in \mathbb{R}$ (see for instance Theorem 3.7 in Kelley and Peterson, 1991). Requiring the initial conditions (2.2) proves the claim. $\square$

Lemma 2.14. *It holds that $\varphi_{n+2} = 1$, $\varphi_{n+1} = 1 - a^2 + \theta$ and*

$$\varphi_i = (1 + a^2(\theta - 2) + \theta)\varphi_{i+1} - a^2\theta(\theta - 1)\varphi_{i+2}, \qquad i = n, n-1, \ldots, 2.$$

Proof. Recall from the proof of Lemma 2.13 that the general solution to the homogeneous linear difference equation of second order

$$f_i = (1 + a^2(\theta - 2) + \theta)f_{i-1} - a^2\theta(\theta - 1)f_{i-2}$$

is $z_+d_+^i + z_-d_-^i$ for $z_+, z_- \in \mathbb{R}$. Requiring the initial conditions $f_0 = 1$ and $f_1 = 1-a^2+\theta$ yields $f_i = c_+d_+^i+c_-d_-^i$, and therefore $f_i = \varphi_{n+2-i}$ for all $i = 0, 1, \ldots, n$. $\square$

The inverse of N can now be expressed in terms of the constants δ_i, φ_i and δ.

Lemma 2.15. *The matrix N is nonsingular, and its inverse is given by*

$$\left(N^{-1}\right)_{ij} = \begin{cases} \delta^{-1}\left(a(\theta-1)\right)^{i-j}\delta_{j-1}\,\varphi_{i+1} & \text{if } j \le i, \\ \delta^{-1}\left(a\theta\right)^{j-i}\delta_{i-1}\,\varphi_{j+1} & \text{if } j \ge i, \end{cases} \qquad \text{for } i, j = 1, 2, \ldots, n+1.$$

Proof. By Lemma 3.2 in Schied and Zhang (2017), the matrix $M + \tilde{M} + 2\tilde{\gamma}I$ is nonsingular. Hence $N = (1-a^2)M^{-1}(M + \tilde{M} + 2\tilde{\gamma}I)$ is nonsingular.
Notice that δ is the determinant of N, hence $\delta \neq 0$. The claim now follows from Lemmas 2.13 and 2.14 with the help of Usmani's (1994a; 1994b) formula for the inversion of a tridiagonal Jacobi matrix. □

This yields an explicit representation of the components of $\boldsymbol{v}$.

Lemma 2.16. *The components of $\boldsymbol{v}$ are given as*

$$v_1 = \frac{1-a}{\delta}\Big(\varphi_2 + (1-a)\sum_{j=2}^{n}(a\theta)^{j-1}\varphi_{j+1} + (a\theta)^n\Big),$$

and, for $i = 2, 3, \ldots, n$,

$$\begin{aligned} v_i = \frac{1-a}{\delta}\Big(&(a(\theta-1))^{i-1}\varphi_{i+1} + (1-a)\sum_{j=2}^{i-1}(a(\theta-1))^{i-j}\delta_{j-1}\varphi_{i+1} \\ &+ (1-a)\sum_{j=i}^{n}(a\theta)^{j-i}\delta_{i-1}\varphi_{j+1} + (a\theta)^{n+1-i}\delta_{i-1}\Big) \end{aligned}$$

and

$$v_{n+1} = \frac{1-a}{\delta}\Big((a(\theta-1))^n + (1-a)\sum_{j=2}^{n}(a(\theta-1))^{n+1-j}\delta_{j-1} + \delta_n\Big).$$

Proof. Notice that

$$\boldsymbol{v} \coloneqq (M + \tilde{M} + 2\tilde{\gamma}I)^{-1}\mathbf{1} = \big(I + M^{-1}(\tilde{M} + 2\tilde{\gamma}I)\big)^{-1}M^{-1}\mathbf{1} = (1-a^2)N^{-1}M^{-1}\mathbf{1}.$$

The result now follows from Lemma 2.15 and the fact that

$$(1-a^2)M^{-1}\mathbf{1} = \big(1-a,\,(1-a)^2,\,(1-a)^2,\ldots,\,(1-a)^2,\,1-a\big),$$

which in turn follows from (2.1). □

To describe the limit behavior of V_n, define the four functions $f_\pm^V, g_\pm^V \colon (0,T) \to \mathbb{R}$ via

$$\begin{aligned} f_\pm^V(t) &:= \big(2e^{6\rho T}(3\rho T+5) + e^{3\rho T} + 3\rho T + 7\big)^{-1}\big(\pm 3e^{3\rho(T-t)} \pm 6e^{3\rho(2T-t)} + \\ &\qquad + e^{6\rho T}(6\rho(T-t)+4) + 3\rho(T-t) + 2e^{3\rho T} + 4e^{3\rho t} - 4e^{3\rho(T+t)} + 3\big), \\ g_\pm^V(t) &:= \big(2e^{6\rho T}(3\rho T+5) - 3e^{3\rho T} - 3\rho T - 7\big)^{-1}\big(\pm 3e^{3\rho(T-t)} \pm 6e^{3\rho(2T-t)} + \\ &\qquad + e^{6\rho T}(6\rho(T-t)+4) - 3\rho(T-t) - 2e^{3\rho T} - 4e^{3\rho t} - 4e^{3\rho(T+t)} - 3\big). \end{aligned}$$

Lemma 2.17.

(i) If $\theta > 1/2$, then

$$\lim_{n\to\infty} V_n(t) = \frac{e^{3\rho T}(6\rho(T-t)+4) - 4e^{3\rho t}}{2e^{3\rho T}(3\rho T+5) - 1}, \qquad t \in (0,T).$$

(ii) Suppose $\theta = 1/2$. For every $t \in (0,T)$, the sequence $(V_n(t))_{n=2,4,6,\dots}$ has exactly two cluster points $f_\pm^V(t)$ and the sequence $(V_n(t))_{n=1,3,5,\dots}$ has exactly two cluster points $g_\pm^V(t)$.

Proof. Continue to keep in mind that the dependence of variables on n is often implicit. For example, $\lim_{n\to\infty} a = 1$ because $a = e^{-\rho T/n}$.
The proof of Lemma 2.17 is split into three parts:

1. Limit behavior of V_n for $\theta = 1$
2. Limit behavior of V_n for $\theta > 1/2$ and $\theta \neq 1$
3. Limit behavior of V_n for $\theta = 1/2$

Throughout the proof, the formula for partial sums of the geometric series,

$$\sum_{i=1}^{m-1} x^i = \sum_{i=1}^{m-1} x^{m-i} = \frac{x - x^m}{1-x} \qquad \text{for } x \neq 1,$$

will be used frequently and without further reference.

Lemma 2.18. *If $\theta = 1$, then, for $m = 1, 2, \dots, n+1$,*

$$\sum_{i=1}^{m} v_i = \frac{(1-a)m + a}{2+a} + \frac{a(a^2-2)}{2(2+a)^2}\Big(\frac{a}{2-a^2}\Big)^{n+1} + \frac{a(1+a)}{(2+a)^2}\Big(\frac{a}{2-a^2}\Big)^{n+1-m}.$$

Proof. Let $\theta = 1$ and $m = 1, 2, \dots, n+1$. Plug into the definitions of r, $b_\pm$, $c_\pm$ and $d_\pm$ (noticing that $a < 1$) to show $\delta_i = 2(1-a^2)(2-a^2)^{i-1}$ for $i = 1, 2, \dots, n$, and $\varphi_i = (2-a^2)^{n+2-i}$ for $i = 2, 3, \dots, n+1$. Furthermore, $\delta = 2(1-a^2)(2-a^2)^n$. Conclude with Lemma 2.16 that

$$v_1 = \frac{1}{2+a}\Big(1 + \frac{2-a^2}{2}\Big(\frac{a}{2-a^2}\Big)^{n+1}\Big)$$

and, for $i = 2, 3, \dots, n+1$,

$$v_i = \frac{1}{2+a}\Big(1 - a + (1-a^2)\Big(\frac{a}{2-a^2}\Big)^{n+2-i}\Big).$$

Sum over $i = 1, 2, \dots, m$ to obtain the representation in the statement. □

Proof of Lemma 2.17, Part 1: Limit behavior of V_n for $\theta = 1$.
Let $\theta = 1$ and $t \in (0, T)$. Then, using L'Hôpital's rule, $\lim_{n\to\infty} (1-a)\, m(t) = \rho t$ and

$$\begin{aligned}\lim_{n\to\infty} (2-a^2)^{m(t)} &= \lim_{n\to\infty} \exp\big(m(t)\log(2-a^2)\big) \\ &= \lim_{n\to\infty} \exp\Big(\frac{2\rho t}{2a^{-2}-1}\Big) \\ &= e^{2\rho t}.\end{aligned}$$

Similarly, $\lim_{n\to\infty}(1-a)(n+1) = \rho T$ and $\lim_{n\to\infty} (2-a^2)^{n+1} = e^{2\rho T}$. Thus Lemma 2.18 shows

$$\lim_{n\to\infty} \sum_{i=1}^{m(t)} v_i = \frac{e^{-3\rho T}(4e^{3\rho t}-1) + 6(\rho t + 1)}{18} \qquad \text{and}$$
$$\lim_{n\to\infty} \sum_{i=1}^{n+1} v_i = \frac{-e^{-3\rho T} + 6\rho T + 10}{18}.$$

Applying these results to the definition of V_n concludes the proof of Part 1.
Define the following shorthand notation for $x \in \mathbb{R}$ and $k \in \mathbb{N}$:

$$[x]^k := \frac{1-a}{\delta} x^k.$$

Lemma 2.19. *Abbreviate*

$$C_1 := \frac{a(1+a)}{\theta + 1 - a(\theta-2)}.$$

If $\theta \geq 1/2$ and $\theta \neq 1$, then, for $m = 1, 2, \dots, n$,

$$\begin{aligned}\sum_{i=1}^{m} v_i = &\sum_{\sigma\in\{+,-\}} \frac{c_\sigma(d_\sigma - a^2\theta)}{d_\sigma - a\theta}[d_\sigma]^n \\ &+ (1-a)(m-1)\sum_{\sigma\in\{+,-\}} b_\sigma c_\sigma\Big(\frac{a(\theta-1)}{d_\sigma - a(\theta-1)} + \frac{d_\sigma}{d_\sigma - a\theta}\Big)[d_\sigma]^n \\ &+ C_1\left(1 + \sum_{\sigma\in\{+,-\}} \frac{b_\sigma d_\sigma\big(\big(\frac{d_\sigma}{a\theta}\big)^{m-1} - 1\big)}{d_\sigma - a\theta}\right)a^n[\theta]^n \\ &+ 2C_1 \sum_{\sigma\in\{+,-\}} \frac{c_\sigma d_\sigma\big(\frac{a(\theta-1)}{d_\sigma} - \big(\frac{a(\theta-1)}{d_\sigma}\big)^m\big)}{d_\sigma - a(\theta-1)}[d_\sigma]^n.\end{aligned}$$

Furthermore,

$$v_{n+1} = \sum_{\sigma\in\{+,-\}} \frac{b_\sigma(d_\sigma - a^2(\theta-1))}{d_\sigma - a(\theta-1)}[d_\sigma]^n + 2C_1 a^n[\theta-1]^n.$$

Proof. Let $\theta \geq 1/2$ and $\theta \neq 1$. For $i = 3, 4, \ldots, n$,

$$\begin{aligned}
&\sum_{j=2}^{i-1}(a(\theta-1))^{i-j}\delta_{j-1}\varphi_{i+1}\\
&= a(\theta-1)\sum_{\sigma\in\{+,-\}}\frac{b_\sigma c_\sigma d_\sigma^n}{d_\sigma - a(\theta-1)}\\
&\quad + \frac{a(\theta-1)b_+c_-d_-^{n+1}}{d_+(d_+ - a(\theta-1))}\Big(\frac{d_+}{d_-}\Big)^i + \frac{a(\theta-1)b_-c_+d_+^{n+1}}{d_-(d_- - a(\theta-1))}\Big(\frac{d_-}{d_+}\Big)^i\\
&\quad - a(\theta-1)\sum_{\sigma\in\{+,-\}}\frac{b_\sigma d_\sigma}{d_\sigma - a(\theta-1)}\sum_{\tau\in\{+,-\}}\frac{c_\tau d_\tau^{n+1}}{(a(\theta-1))^2}\Big(\frac{a(\theta-1)}{d_\tau}\Big)^i
\end{aligned}$$

and

$$\begin{aligned}
\sum_{j=i}^{n}(a\theta)^{j-i}\delta_{i-1}\varphi_{j+1} &= \sum_{\sigma\in\{+,-\}}\frac{b_\sigma c_\sigma}{d_\sigma - a\theta}d_\sigma^{n+1}\\
&\quad + \frac{b_+c_-d_-^{n+2}}{d_+(d_- - a\theta)}\Big(\frac{d_+}{d_-}\Big)^i + \frac{b_-c_+d_+^{n+2}}{d_-(d_+ - a\theta)}\Big(\frac{d_-}{d_+}\Big)^i\\
&\quad - \sum_{\sigma\in\{+,-\}}\frac{c_\sigma d_\sigma}{d_\sigma - a\theta}\sum_{\tau\in\{+,-\}}\frac{b_\tau(a\theta)^{n+1}}{d_\tau}\Big(\frac{d_\tau}{a\theta}\Big)^i.
\end{aligned}$$

Conclude from the definitions of r and $d_\pm$ that

$$\begin{aligned}
&a(\theta-1)(d_- - a\theta) + d_-(d_+ - a(\theta-1))\\
&= a(\theta-1)(d_+ - a\theta) + d_+(d_- - a(\theta-1))\\
&= d_+d_- - a^2\theta(\theta-1)\\
&= 0.
\end{aligned}$$

Hence four summands from the previous calculations cancel out:

$$\begin{aligned}
0 &= \frac{a(\theta-1)b_+c_-d_-^{n+1}}{d_+(d_+ - a(\theta-1))}\Big(\frac{d_+}{d_-}\Big)^i + \frac{a(\theta-1)b_-c_+d_+^{n+1}}{d_-(d_- - a(\theta-1))}\Big(\frac{d_-}{d_+}\Big)^i\\
&\quad + \frac{b_+c_-d_-^{n+2}}{d_+(d_- - a\theta)}\Big(\frac{d_+}{d_-}\Big)^i + \frac{b_-c_+d_+^{n+2}}{d_-(d_+ - a\theta)}\Big(\frac{d_-}{d_+}\Big)^i\\
&= \Big(\frac{a(\theta-1)}{d_+ - a(\theta-1)} + \frac{d_-}{d_- - a\theta}\Big)\frac{b_+c_-d_-^{n+1}}{d_+}\Big(\frac{d_+}{d_-}\Big)^i\\
&\quad + \Big(\frac{a(\theta-1)}{d_- - a(\theta-1)} + \frac{d_+}{d_+ - a\theta}\Big)\frac{b_-c_+d_+^{n+1}}{d_-}\Big(\frac{d_-}{d_+}\Big)^i.
\end{aligned}$$

For $i = 2, 3, \dots, n$, plug into the representation of v_i from Lemma 2.16 and simplify further to show

$$\begin{aligned} v_i = (1-a) \sum_{\sigma \in \{+,-\}} b_\sigma c_\sigma \Big(\frac{a(\theta-1)}{d_\sigma - a(\theta-1)} + \frac{d_\sigma}{d_\sigma - a\theta} \Big) [d_\sigma]^n \\ + 2C_1 \sum_{\sigma \in \{+,-\}} \frac{c_\sigma d_\sigma [d_\sigma]^n}{a(\theta-1)} \Big(\frac{a(\theta-1)}{d_\sigma} \Big)^i + C_1 \sum_{\sigma \in \{+,-\}} \frac{b_\sigma a^{n+1} \theta [\theta]^n}{d_\sigma} \Big(\frac{d_\sigma}{a\theta} \Big)^i . \end{aligned}$$

Observe with similar calculations that

$$v_1 = \sum_{\sigma \in \{+,-\}} \frac{c_\sigma (d_\sigma - a^2\theta)}{d_\sigma - a\theta} [d_\sigma]^n + C_1 a^n [\theta]^n$$

and

$$v_{n+1} = \sum_{\sigma \in \{+,-\}} \frac{b_\sigma (d_\sigma - a^2(\theta-1))}{d_\sigma - a(\theta-1)} [d_\sigma]^n + 2C_1 a^n [\theta-1]^n .$$

Sum over $i = 1, 2, \dots, m$ to conclude the proof. $\square$

Notice that if $t = T$, then $(-1)^{m(T)} = 1$ if n is even and $(-1)^{m(T)} = -1$ is n is odd.

If $t \in (0, T)$ then $0 < 2t/T < 2$. Recall that $m(t) = \lceil nt/T \rceil$. Define $k_n := \lceil 2nt/T \rceil$. If n increases by one, then k_n increases by zero, one or two. Clearly, there are infinitely many instances in which k_n increases either by one or two. Suppose there is an $n_0 \in \mathbb{N}$ such that k_n only increases by two for all $n \geq n_0$. Then there must be an $n_1 \geq n_0$ for which $k_{n_1} - 1 > 2n_1 t/T$. This contradicts the definition of the ceiling function. Conclude that there are infinitely many instances in which k_n increases by one. Hence $((-1)^{m(t)})_{n=2,4,6,\dots}$ oscillates between -1 and 1. The same is true for $((-1)^{m(t)})_{n=1,3,5,\dots}$

Introduce the following convention: For a sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ and a real number x, say that $\pm x$ is the limit of $(x_n)_{n\in\mathbb{N}}$ and write $\lim_{n\to\infty}(x_n)^{m(t)} = \pm x$ if

$$(x_n)^{m(t)} = (-1)^{m(t)} |x_n|^{m(t)} \qquad \text{and} \qquad \lim_{n\to\infty} |x_n|^{m(t)} = x.$$

To study the limit behavior of V_n and $J_n[\boldsymbol{\xi}_{\boldsymbol{n}}^* \,|\, \boldsymbol{\eta}_{\boldsymbol{n}}^*]$, they will be split into small parts that are easier to analyze. The following lemma contains the limits of these parts.

Lemma 2.20. *Let $\theta \geq 1/2$ and $\theta \neq 1$ and $t \in [0, T]$. For $n \to \infty$, the following limits hold:*

Expression	*Limit*	*Expression*	*Limit*	*Expression*	*Limit*
a	1	c_-	0	$\frac{b_+}{1-a^2}$	2θ
$a^{m(t)}$	$e^{-\rho t}$	d_+	θ	$\frac{c_-}{d_- -(\theta-1)}$	$-\frac{1}{2}$
r	1	d_-	$\theta - 1$	$\frac{c_-}{d_- - a(\theta-1)}$	$-\frac{2}{3}$
b_+	0	$\frac{b_+}{d_+ - \theta}$	2	$\frac{c_-}{d_- - a^2(\theta-1)}$	-1
b_-	1	$\frac{b_+}{d_+ - a\theta}$	$\frac{4}{3}$	$\frac{c_-}{1-a^2}$	$\theta - 1$
c_+	1	$\frac{b_+}{d_+ - a^2\theta}$	1	$(1-a)m(t)$	ρt

If additionally $\theta > 1/2$, then also the following limits hold:

Expression	*Limit*	*Expression*	*Limit*	*Expression*	*Limit*
$\left(\frac{\theta-1}{\theta}\right)^{m(t)}$	0	$\left(\frac{\theta-1}{d_-}\right)^{m(t)}$	$e^{4\rho t}$	$[\theta-1]^n$	0
$\left(\frac{d_+}{\theta}\right)^{m(t)}$	$e^{2\rho t}$	$[d_+]^n$	$\frac{1}{4\theta}$	$\frac{((\theta-1)/\theta)^n}{1-a^2}$	0
$\left(\frac{\theta-1}{d_+}\right)^{m(t)}$	0	$[d_-]^n$	0	$\frac{[d_-]^n}{1-a^2}$	0
$\left(\frac{d_-}{\theta}\right)^{m(t)}$	0	$[\theta]^n$	$\frac{e^{-2\rho T}}{4\theta}$	$\frac{[\theta-1]^n}{1-a^2}$	0

If, on the other hand, $\theta = 1/2$, then the limits in the second table are no longer valid. Instead, the following limits hold:

Expression	*Limit*	*Expression*	*Limit*	*Expression*	*Limit*
$\left(\frac{\theta-1}{\theta}\right)^{m(t)}$	± 1	$\left(\frac{\theta-1}{d_-}\right)^{m(t)}$	$e^{4\rho t}$	$\frac{b_+}{d_+(\theta-1)-d_-\theta}$	$-\frac{4}{3}$
$\left(\frac{d_+}{\theta}\right)^{m(t)}$	$e^{2\rho t}$	$\frac{d_+ + \theta - 1}{d_+ + a(\theta-1)}$	$\frac{2}{3}$	$\frac{c_-}{d_+(\theta-1)-d_-\theta}$	$\frac{2}{3}$
$\left(\frac{\theta-1}{d_+}\right)^{m(t)}$	$\pm e^{-2\rho t}$	$\frac{d_- + a^2\theta}{d_- + a\theta}$	$\frac{2}{3}$		
$\left(\frac{d_-}{\theta}\right)^{m(t)}$	$\pm e^{-4\rho t}$	$\frac{\theta + a(\theta-1)}{1-a^2}$	$\frac{1}{4}$		

Furthermore, if $\theta = 1/2$, then the following limits hold:

$$\lim_{\substack{n\to\infty \\ n\,\text{even}}} [d_+]^n = \frac{1}{e^{-6\rho T}+2}, \qquad \lim_{\substack{n\to\infty \\ n\,\text{odd}}} [d_+]^n = \frac{1}{-e^{-6\rho T}+2},$$

$$\lim_{\substack{n\to\infty \\ n\,\text{even}}} [d_-]^n = \frac{1}{2e^{6\rho T}+1}, \qquad \lim_{\substack{n\to\infty \\ n\,\text{odd}}} [d_-]^n = \frac{1}{-2e^{6\rho T}+1},$$

$$\lim_{\substack{n\to\infty \\ n\,\text{even}}} [\theta]^n = \frac{e^{4\rho T}}{2e^{6\rho T}+1}, \qquad \lim_{\substack{n\to\infty \\ n\,\text{odd}}} [\theta]^n = \frac{e^{4\rho T}}{2e^{6\rho T}-1},$$

$$\lim_{\substack{n\to\infty \\ n\,\text{even}}} [\theta-1]^n = \frac{e^{4\rho T}}{2e^{6\rho T}+1}, \qquad \lim_{\substack{n\to\infty \\ n\,\text{odd}}} [\theta-1]^n = \frac{e^{4\rho T}}{-2e^{6\rho T}+1}.$$

Proof. The limits in the first table are obvious or follow after an application of L'Hôpital's rule.

Consider the second table. The first limit follows directly from the fact that $\theta > 1/2$. To prove the second, write $(d_+/\theta)^n = \exp\big(n \log(d_+/\theta)\big)$ and apply L'Hôpital's rule. The third limit follows immediately, because

$$\frac{\theta - 1}{d_+} = \frac{\theta - 1}{\theta}\frac{\theta}{d_+}.$$

Prove the fourth and fifth limits in a similar fashion. Now recall that

$$\begin{aligned}&\frac{\delta}{1-a}\\ &= \frac{b_+\big(1 - a^2 + \theta - \frac{a^2\theta(\theta-1)}{d_+}\big)}{1-a} d_+^n + \frac{b_-((1-a^2+\theta)d_- - a^2\theta(\theta-1))}{d_-(1-a)} d_-^n.\end{aligned} \tag{2.4}$$

Applying L'Hôpital's rule shows

$$\lim_{n\to\infty} \frac{b_+\big(1 - a^2 + \theta - \frac{a^2\theta(\theta-1)}{d_+}\big)}{1-a} = 4\theta$$

and

$$\lim_{n\to\infty} \frac{b_-((1-a^2+\theta)d_- - a^2\theta(\theta-1))}{d_-(1-a)} = -2(\theta - 1).$$

Notice that this remains true if $\theta = 1/2$. By setting $t = T$, it follows from the second and fourth limits in the second table that

$$\lim_{n\to\infty} \left(\frac{d_-}{d_+}\right)^n = 0, \qquad \text{which implies} \qquad \lim_{n\to\infty} \left|\frac{d_+}{d_-}\right|^n = \infty.$$

Thus

$$\lim_{n\to\infty} [d_+]^n = \lim_{n\to\infty} \left(\frac{\delta}{(1-a)d_+^n}\right)^{-1} = \big(4\theta + 0\big)^{-1} \qquad \text{and} \qquad \lim_{n\to\infty} [d_-]^n = 0.$$

Similar considerations yield the limits of $[\theta]^n$ and $[\theta - 1]^n$. Further applications of L'Hôpital's rule show

$$\lim_{n\to\infty} \frac{((\theta-1)/\theta)^n}{1-a^2} = 0$$

as well as

$$\lim_{n\to\infty} \left|\frac{1-a^2}{(d_-/d_+)^n}\right| = \infty \qquad \text{and} \qquad \lim_{n\to\infty} \left|\frac{1-a^2}{((\theta-1)/d_+)^n}\right| = \infty.$$

Conclude with the fifth limit in the second table that

$$\lim_{n\to\infty} \frac{1-a^2}{((\theta-1)/d_-)^n} = 0.$$

Now proceed as above to obtain the last two limits in the second table.
Consider the third table. For $\theta = 1/2$, it holds that $d_- < 0 < d_+$ because

$$d_\pm = \frac{1}{2}\Big(\frac{3(1-a^2)}{2} \pm \sqrt{\frac{9(1-a^2)^2}{4} + a^2}\Big).$$

With this in mind, write $(\theta-1)^{m(t)} = (-1)^{m(t)}|\theta-1|^{m(t)}$ and $d_-^{m(t)} = (-1)^{m(t)}|d_-|^{m(t)}$. Then obtain the first five limits with similar arguments as before. For example,

$$\begin{aligned}\Big(\frac{d_-}{\theta}\Big)^{m(t)} &= (-1)^{m(t)}\left|\frac{3(1-a^2)}{2} - \sqrt{\frac{9(1-a^2)^2}{4} + a^2}\right|^{m(t)} \\ &= (-1)^{m(t)} \exp\Big(m(t)\log\Big(\sqrt{\frac{9(1-a^2)^2}{4} + a^2} - \frac{3(1-a^2)}{2}\Big)\Big).\end{aligned}$$

Apply L'Hôpital's rule to show

$$\lim_{n\to\infty} m(t)\log\Big(\sqrt{\frac{9(1-a^2)^2}{4} + a^2} - \frac{3(1-a^2)}{2}\Big) = -4\rho t.$$

The remaining five limits in the third table follow after an application of L'Hôpital's rule.
To find the limits of $[d_+]^n$, $[d_-]^n$, $[\theta]^n$ and $[\theta-1]^n$ for even and odd n, recall (2.4) and proceed as before. For example, conclude with similar arguments as for the second and fourth limit in the third table that $\lim_{n\to\infty}|d_-/d_+|^n = e^{-6\rho T}$, and obtain

$$\lim_{\substack{n\to\infty \\ n\,\text{even}}} [d_+]^n = \Big(2 + \lim_{n\to\infty}\Big|\frac{d_-}{d_+}\Big|\Big)^{-1} = \frac{1}{e^{-6\rho T} + 2}.$$

□

Proof of Lemma 2.17, Part 2: Limit behavior of V_n for $\theta > 1/2$ and $\theta \neq 1$. Let $\theta \neq 1$ and $t \in (0,T)$. With the help of Lemma 2.19 and the first table in Lemma 2.20, conclude that $\lim_{n\to\infty} C_1 = 2/3$ and thus

$$\begin{aligned}\lim_{n\to\infty}\sum_{i=1}^{m(t)} v_i = {}& \frac{4}{3}\lim_{n\to\infty}[d_+]^n + \rho t\Big(\frac{4\theta}{3}\lim_{n\to\infty}[d_+]^n - \frac{2(\theta-1)}{3}\lim_{n\to\infty}[d_-]^n\Big) \\ &+ \frac{2}{3}\Big(1 + \frac{4\theta}{3}\Big(e^{\rho t}\lim_{n\to\infty}\Big(\frac{d_+}{\theta}\Big)^{m(t)} - 1\Big) \\ &\qquad - (\theta-1)\Big(\frac{\theta e^{\rho t}}{\theta-1}\lim_{n\to\infty}\Big(\frac{d_-}{\theta}\Big)^{m(t)} - 1\Big)\Big)e^{-\rho T}\lim_{n\to\infty}[\theta]^n \\ &+ \frac{4}{3}\Big(\theta\Big(\frac{\theta-1}{\theta} - e^{-\rho t}\lim_{n\to\infty}\Big(\frac{\theta-1}{d_+}\Big)^{m(t)}\Big)\lim_{n\to\infty}[d_+]^n \\ &\qquad - \frac{2(\theta-1)}{3}\Big(1 - e^{-\rho t}\lim_{n\to\infty}\Big(\frac{\theta-1}{d_-}\Big)^{m(t)}\Big)\lim_{n\to\infty}[d_-]^n\Big).\end{aligned} \tag{2.5}$$

Notice that this requires writing

$$\frac{d_+ - a^2\theta}{d_+ - a\theta} = \Big(\frac{b_+}{d_+ - a^2\theta}\Big)^{-1} \frac{b_+}{d_+ - a\theta}$$

and

$$\frac{d_- - a^2(\theta-1)}{d_- - a(\theta-1)} = \Big(\frac{c_-}{d_- - a^2(\theta-1)}\Big)^{-1} \frac{c_-}{d_- - a(\theta-1)}.$$

Now let additionally $\theta > 1/2$. The remaining limits are in the second table in Lemma 2.20. In total, obtain for $t \in (0, T)$:

$$\lim_{n\to\infty} \sum_{i=1}^{m(t)} v_i = \frac{e^{-3\rho T}(4e^{3\rho t} - 1) + 6(\rho t + 1)}{18}$$

and, after showing $\lim_{n\to\infty} v_{n+1} = 0$,

$$\lim_{n\to\infty} \sum_{i=1}^{n+1} v_i = \frac{-e^{-3\rho T} + 6\rho T + 10}{18}.$$

Both limits are the same as in the case $\theta = 1$ (compare Part 1). Plug these results into the definition of $V_n(t)$ to conclude the proof of Part 2.

Proof of Lemma 2.17, Part 3: Limit behavior of V_n for $\theta = 1/2$.
Let $\theta = 1/2$ and $t \in (0, T)$. Equation (2.5) remains valid. The remaining limits can be found in Lemma 2.20. In total, obtain for $\sigma \in \{+, -\}$:

$$\begin{aligned}\lim_{\substack{n\to\infty\\ n\,\text{even}}} \sum_{i=1}^{m(t)} v_i = \big(18e^{6\rho T} + 9\big)^{-1}\big(&-\sigma 3e^{3\rho(T-t)} - \sigma 6e^{3\rho(2T-t)} \\ &+ 6e^{6\rho T}(\rho t + 1) + 3\rho t - e^{3\rho T} - 4e^{3\rho t} + 4e^{3\rho(T+t)} + 4\big)\end{aligned}$$

and

$$\begin{aligned}\lim_{\substack{n\to\infty\\ n\,\text{odd}}} \sum_{i=1}^{m(t)} v_i = \big(-18e^{6\rho T} + 9\big)^{-1}\big(&\sigma 3e^{3\rho(T-t)}\sigma 6e^{3\rho(2T-t)} \\ &- 6e^{6\rho T}(\rho t + 1) + 3\rho t + e^{3\rho T} - 4e^{3\rho t} - 4e^{3\rho(T+t)} + 4\big).\end{aligned}$$

Letting $\sigma = +$ and $\sigma = -$ yields the respective cluster points for even and odd $m(t)$. Notice that $m(T)$ is even if and only if n is even. Obtain

$$\lim_{\substack{n\to\infty\\ n\,\text{even}}} v_{n+1} = \frac{2(2e^{3\rho T} + 1)}{3(2e^{6\rho T} + 1)} \qquad \text{and} \qquad \lim_{\substack{n\to\infty\\ n\,\text{odd}}} v_{n+1} = \frac{2(2e^{3\rho T} + 1)}{3(-2e^{6\rho T} + 1)}.$$

Plug these results into the definition of $V_n(t)$ to conclude the proof of Part 3. $\square$

This completes the proof of the auxiliary Lemma 2.17. Part 3 also yields the following limits, which will be needed later: If $\theta = 1/2$, then

$$\begin{aligned}
\lim_{\substack{n\to\infty\\ n\,\mathrm{even}}} \mathbf{1}^\top \boldsymbol{v} &= \frac{2e^{6\rho T}(3\rho T+5)+e^{3\rho T}+3\rho T+7}{18e^{6\rho T}+9} \qquad \text{and}\\
\lim_{\substack{n\to\infty\\ n\,\mathrm{odd}}} \mathbf{1}^\top \boldsymbol{v} &= \frac{2e^{6\rho T}(3\rho T+5)-3e^{3\rho T}-3\rho T-7}{18e^{6\rho T}-9}.
\end{aligned} \tag{2.6}$$

The final step in the proof of Theorem 2.6 is to study the limit behavior of expected costs $J_n[\boldsymbol{\xi}_{\boldsymbol{n}}^* \,|\, \boldsymbol{\eta}_{\boldsymbol{n}}^*]$. The following lemma expresses them in terms of $\boldsymbol{v}$ and $\boldsymbol{w}$. It is valid for all $\theta \geq 1/2$.

Lemma 2.21. *It holds that*

$$\begin{aligned}
8\,J_n[\boldsymbol{\xi}_{\boldsymbol{n}}^* \,|\, \boldsymbol{\eta}_{\boldsymbol{n}}^*] = {}& \frac{(x^0+y^0)^2}{\mathbf{1}^\top \boldsymbol{v}} + \frac{(x^0+y^0)(x^0-y^0)}{\mathbf{1}^\top \boldsymbol{v}\,\mathbf{1}^\top \boldsymbol{w}}(\mathbf{1}^\top \boldsymbol{v}+\mathbf{1}^\top \boldsymbol{w}) + \frac{(x^0-y^0)^2}{\mathbf{1}^\top \boldsymbol{w}}\\
&+ \Big(\frac{x^0+y^0}{\mathbf{1}^\top \boldsymbol{v}}\Big)^2 \boldsymbol{v}^\top \tilde{M}\,\boldsymbol{v} + \frac{(x^0+y^0)(x^0-y^0)}{\mathbf{1}^\top \boldsymbol{v}\,\mathbf{1}^\top \boldsymbol{w}} \boldsymbol{w}^\top (\tilde{M}-\tilde{M}^\top)\boldsymbol{v}\\
&- \Big(\frac{x^0-y^0}{\mathbf{1}^\top \boldsymbol{w}}\Big)^2 \boldsymbol{w}^\top \tilde{M}\,\boldsymbol{w}.
\end{aligned}$$

Proof. Recall the representation of $J_n[\boldsymbol{\xi}_{\boldsymbol{n}} \,|\, \boldsymbol{\eta}_{\boldsymbol{n}}]$ in Lemma 2.12 (iv). Plug in for $\boldsymbol{\xi}_{\boldsymbol{n}}^*$ and $\boldsymbol{\eta}_{\boldsymbol{n}}^*$ from Theorem 2.2 and rearrange to see that the right-hand side equals

$$\begin{aligned}
\frac{1}{2}\bigg(&\Big(\frac{x^0+y^0}{2\,(\mathbf{1}^\top \boldsymbol{v})}\Big)^2 \boldsymbol{v}^\top (M+\tilde{M}+2\tilde{\gamma}I)\boldsymbol{v}\\
&+ \frac{(x^0+y^0)(x^0-y^0)}{4\,(\mathbf{1}^\top \boldsymbol{v})\,(\mathbf{1}^\top \boldsymbol{w})}\big(\boldsymbol{v}^\top (M-\tilde{M}+2\tilde{\gamma}I)\boldsymbol{w} + \boldsymbol{w}^\top (M+\tilde{M}+2\tilde{\gamma}I)\boldsymbol{v}\big)\\
&+ \Big(\frac{x^0-y^0}{2\,(\mathbf{1}^\top \boldsymbol{w})}\Big)^2 \boldsymbol{w}^\top (M-\tilde{M}+2\tilde{\gamma}I)\boldsymbol{w}\bigg) + \frac{1}{2}(\boldsymbol{\xi}_{\boldsymbol{n}}^*)^\top \tilde{M}\boldsymbol{\eta}_{\boldsymbol{n}}^*.
\end{aligned}$$

By definition, $(M+\tilde{M}+2\tilde{\gamma}I)\,\boldsymbol{v} = (M-\tilde{M}+2\tilde{\gamma}I)\,\boldsymbol{w} = \mathbf{1}$. Plug in for $(\boldsymbol{\xi}_{\boldsymbol{n}}^*)^\top \tilde{M}\boldsymbol{\eta}_{\boldsymbol{n}}^*$ and simplify further to conclude the proof. □

Lemma 2.22. *If $\theta > 1/2$, then*

$$\begin{aligned}
\lim_{n\to\infty} \boldsymbol{v}^\top \tilde{M}\,\boldsymbol{v} &= \frac{-e^{-6\rho T}-8e^{-3\rho T}+24\rho T+36}{216},\\
\lim_{n\to\infty} \boldsymbol{w}^\top (\tilde{M}-\tilde{M}^\top)\,\boldsymbol{v} &= \frac{-e^{-3\rho T}+4}{6} \qquad \textit{and}\\
\lim_{n\to\infty} \boldsymbol{w}^\top \tilde{M}\,\boldsymbol{w} &= \frac{2\rho T+1}{2}.
\end{aligned}$$

Proof. First, let $\theta = 1$. Notice from Lemma 2.12 (i) that

$$\boldsymbol{w} = (1-a, 1-a, \ldots, 1-a, 1).$$

Calculate

$$\begin{aligned}\boldsymbol{v}^\top \tilde{M}\,\boldsymbol{v} &= \frac{v_1^2}{2} + \frac{1}{2}\sum_{i=2}^{n+1} v_i^2 + v_1 \sum_{i=2}^{n+1} v_i a^{i-1} + \sum_{i=3}^{n+1}\sum_{j=2}^{i-1} v_i v_j a^{i-j} \\ &= \frac{1}{(2+a)^2}\Bigg(\frac{(1-a^2)n}{2} + \frac{-a^4+2a^2+4a+4}{2(4-a^2)} \\ &\qquad - \frac{a^2(a+1)}{2(a+2)}\Big(\frac{a}{2-a^2}\Big)^n - \frac{a^4}{8(4-a^2)}\Big(\frac{a}{2-a^2}\Big)^{2n}\Bigg),\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\,\boldsymbol{v} &= v_1\sum_{i=2}^{n+1} w_i a^{i-1} - v_{n+1}\sum_{i=1}^{n} w_i a^{n+1-i} - w_1 \sum_{j=2}^{n} v_j a^{j-1} \\ &\quad + w_{n+1}\sum_{j=2}^{n} v_j a^{n+1-j} + \sum_{i=3}^{n} w_i \sum_{j=2}^{i-1} v_j a^{i-j} - \sum_{i=2}^{n-1} w_i \sum_{j=i+1}^{n} v_j a^{j-i} \\ &= -\frac{a^2}{(2-a^2)(2+a)}\Big(\frac{2-a^2}{2}\Big(\frac{a}{2-a^2}\Big)^n + a^2 - 3\Big)\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{w}^\top \tilde{M}\,\boldsymbol{w} &= \frac{1}{2}\sum_{i=1}^{n}(1-a)^2 + \frac{1}{2} + \sum_{i=2}^{n}\sum_{j=1}^{i-1}(1-a)^2 a^{i-j} + \sum_{j=1}^{n}(1-a)a^{n+1-j} \\ &= \frac{n(1-a^2)+1}{2}.\end{aligned}$$

Take limits to obtain the results.

For the remainder of the proof, let $\theta \geq 1/2$ and $\theta \neq 1$. The case $\theta = 1/2$ is included for future reference. If a particular calculation does not hold for the case $\theta = 1/2$, this will be stated explicitly. Finding the limits of $\boldsymbol{v}^\top \tilde{M}\boldsymbol{v}$, $\boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\,\boldsymbol{v}$, and $\boldsymbol{w}^\top \tilde{M}\,\boldsymbol{w}$ is tedious. Begin by computing $\tilde{M}\boldsymbol{v}$. Let

$$C_1 \coloneqq a(1+a)/(1-a(\theta-2)+\theta)$$

as above, and define

$$C_2 \coloneqq \sum_{\sigma\in\{+,-\}} b_\sigma c_\sigma\Big(\frac{a(\theta-1)}{d_\sigma - a(\theta-1)} + \frac{d_\sigma}{d_\sigma - a\theta}\Big)[d_\sigma]^n$$

and

$$C_3 \coloneqq -C_2 + \sum_{\sigma\in\{+,-\}} c_\sigma\Big(\frac{d_\sigma - a^2\theta}{d_\sigma - a\theta} + \frac{2C_1(\theta-1)}{d_\sigma-(\theta-1)}\Big)[d_\sigma]^n.$$

Then

$$
\begin{aligned}
\big(\tilde{M}\boldsymbol{v}\big)_1 &= \frac{1}{2}\sum_{\sigma\in\{+,-\}} \frac{c_\sigma(d_\sigma - a^2\theta)}{d_\sigma - a\theta}[d_\sigma]^n + \frac{C_1 a^n}{2}[\theta]^n,\\
\big(\tilde{M}\boldsymbol{v}\big)_2 &= \sum_{\sigma\in\{+,-\}} c_\sigma\Big(\frac{a(d_\sigma - a^2\theta)}{d_\sigma - a\theta} + \frac{C_1 a(\theta-1)}{d_\sigma}\Big)[d_\sigma]^n\\
&\quad + \frac{C_2(1-a)}{2} + \frac{C_1(1+2a^2(\theta-1)+\theta)a^n}{2a\theta}[\theta]^n,
\end{aligned}
$$

and, for $i = 3, 4, \dots, n$,

$$
\begin{aligned}
\big(\tilde{M}\boldsymbol{v}\big)_i &= \frac{C_2(1+a)}{2} + \frac{C_1}{a(\theta-1)} \sum_{\sigma\in\{+,-\}} \frac{c_\sigma d_\sigma(1-\theta-d_\sigma)[d_\sigma]^n}{1-\theta+d_\sigma}\Big(\frac{a(\theta-1)}{d_\sigma}\Big)^i\\
&\quad + \frac{C_1 a^{n+1}\theta[\theta]^n}{2} \sum_{\sigma\in\{+,-\}} \frac{b_\sigma(d_\sigma + a^2\theta)}{d_\sigma(d_\sigma - a^2\theta)}\Big(\frac{d_\sigma}{a\theta}\Big)^i + C_3 a^{i-1},
\end{aligned}
$$

as well as

$$
\begin{aligned}
\big(\tilde{M}\boldsymbol{v}\big)_{n+1} &= \sum_{\sigma\in\{+,-\}} b_\sigma\Big(\frac{c_\sigma d_\sigma a}{d_\sigma - a\theta} + \frac{d_\sigma + (2c_\sigma - 1)a^2(\theta-1)}{2(d_\sigma - a(\theta-1))} + \frac{C_1 a^2\theta}{d_\sigma - a^2\theta}\Big)[d_\sigma]^n\\
&\quad + C_3 a^n - C_1 a^n\theta[\theta-1]^n.
\end{aligned}
$$

Denote by τ the opposite sign of $\sigma \in \{+,-\}$, i.e.,

$$
\tau = - \ \text{ if } \ \sigma = + \qquad \text{and} \qquad \tau = + \ \text{ if } \ \sigma = -.
$$

For $i = 3, 4, \dots, n$, compute each element $v_i\,(\tilde{M}\,\boldsymbol{v})_i$ of the vector product $\boldsymbol{v}\,(\tilde{M}\boldsymbol{v})$ with the help of the following decomposition: $v_i\,(\tilde{M}\,\boldsymbol{v})_i = D_i^1 + D_i^2 + D_i^3 + D_i^4$, where

$$
\begin{aligned}
D_i^1 &\coloneqq C_2(1+a)v_i/2,\\
\frac{1}{C_2(1-a)}D_i^2 &\coloneqq C_3 a^{i-1} + \frac{C_1}{a(\theta-1)} \sum_{\sigma\in\{+,-\}} \frac{c_\sigma d_\sigma(1-\theta-d_\sigma)[d_\sigma]^n}{1-\theta+d_\sigma}\Big(\frac{a(\theta-1)}{d_\sigma}\Big)^i\\
&\quad + \frac{C_1 a^{n+1}\theta[\theta]^n}{2} \sum_{\sigma\in\{+,-\}} \frac{b_\sigma(d_\sigma + a^2\theta)}{d_\sigma(d_\sigma - a^2\theta)}\Big(\frac{d_\sigma}{a\theta}\Big)^i,\\
\frac{a}{C_1C_3}D_i^3 &\coloneqq 2\sum_{\sigma\in\{+,-\}} \frac{c_\sigma d_\sigma[d_\sigma]^n}{a(\theta-1)}\Big(\frac{a^2(\theta-1)}{d_\sigma}\Big)^i + \sum_{\sigma\in\{+,-\}} \frac{b_\sigma a^{n+1}\theta[\theta]^n}{d_\sigma}\Big(\frac{d_\sigma}{\theta}\Big)^i
\end{aligned}
$$

and

$$
\frac{1}{C_1^2}D_i^4 \coloneqq \frac{2}{(a(\theta-1))^2}\Bigg(\sum_{\sigma\in\{+,-\}} \big(c_\sigma d_\sigma[d_\sigma]^n\big)^2 \frac{1-\theta-d_\sigma}{1-\theta+d_\sigma}\Big(\frac{a(\theta-1)}{d_\sigma}\Big)^{2i}
$$

$$
\begin{aligned}
&\qquad + c_+c_-d_+d_-[d_+]^n[d_-]^n\frac{(\theta-1)^2-d_+d_-}{(1-a^2)(1-\theta)}\Big(\frac{(a(\theta-1))^2}{d_+d_-}\Big)^i\Bigg) \\
&+\frac{(a^{n+1}\theta\,[\theta]^n)^2}{2}\Bigg(\sum_{\sigma\in\{+,-\}}\frac{b_\sigma^2(d_\sigma+a^2\theta)}{d_\sigma^2(d_\sigma-a^2\theta)}\Big(\frac{d_\sigma}{a\theta}\Big)^{2i} \\
&\qquad +\frac{b_+b_-((a^2\theta)^2-d_+d_-)}{d_+d_-a^2(1-a^2)\theta}\Big(\frac{d_+d_-}{(a\theta)^2}\Big)^i\Bigg) \\
&+\frac{a^n\theta\,[\theta]^n}{\theta-1}\Bigg(\frac{1-(1-a^2)\theta}{1-a^2}\sum_{\sigma\in\{+,-\}}b_\sigma c_\sigma[d_\sigma]^n\Big(\frac{\theta-1}{\theta}\Big)^i \\
&\qquad +\sum_{\sigma\in\{+,-\}}\frac{b_\tau c_\sigma d_\sigma[d_\sigma]^n}{d_\tau}\Big(\frac{1-\theta-d_\sigma}{1-\theta+d_\sigma}+\frac{d_\tau+a^2\theta}{d_\tau-a^2\theta}\Big)\Big(\frac{d_\tau(\theta-1)}{d_\sigma\theta}\Big)^i\Bigg).
\end{aligned}
$$

Sum over i to find that

$$
\begin{aligned}
\frac{2}{C_2(1+a)}\sum_{i=3}^n D_i^1 &= C_2(1-a)(n-2)+C_1\sum_{\sigma\in\{+,-\}}\frac{b_\sigma d_\sigma\big(\frac{a\theta}{d_\sigma}[d_\sigma]^n-\frac{d_\sigma}{a\theta}a^n[\theta]^n\big)}{d_\sigma-a\theta} \\
&\quad +2C_1\sum_{\sigma\in\{+,-\}}\frac{c_\sigma d_\sigma\Big(\big(\frac{a(\theta-1)}{d_\sigma}\big)^2[d_\sigma]^n-a^n[\theta-1]^n\Big)}{d_\sigma-a(\theta-1)}
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{i=3}^n D_i^2 &= C_2C_3(a^2-a^n) \\
&\quad +\frac{C_1C_2}{1+a}\sum_{\sigma\in\{+,-\}}\frac{(1-a^2)c_\sigma^2(1-\theta-d_\sigma)\big(d_\sigma^2a^n[\theta-1]^n-(a(\theta-1))^2[d_\sigma]^n\big)}{c_\sigma d_\sigma(d_\sigma-(\theta-1))(a(\theta-1)-d_\sigma)} \\
&\quad +\frac{C_1C_2}{2(1+a)}\sum_{\sigma\in\{+,-\}}\frac{(1-a^2)b_\sigma^2(d_\sigma+a^2\theta)\big((a\theta)^2[d_\sigma]^n-d_\sigma^2a^n[\theta]^n\big)}{b_\sigma a\theta(d_\sigma-a\theta)(d_\sigma-a^2\theta)}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{C_1C_3}\sum_{i=3}^n D_i^3 &= \sum_{\sigma\in\{+,-\}}\frac{2c_\sigma\big((d_\sigma a^n)^2[\theta-1]^n-a^n(a(\theta-1))^2[d_\sigma]^n\big)}{d_\sigma(a^2(\theta-1)-d_\sigma)} \\
&\quad +\sum_{\sigma\in\{+,-\}}\frac{b_\sigma a^n(\theta^2\,[d_\sigma]^n-d_\sigma^2[\theta]^n)}{\theta(d_\sigma-\theta)}
\end{aligned}
$$

and

$$\begin{aligned}
&\frac{1}{C_1^2}\sum_{i=3}^{n} D_i^4\\
&= 2\sum_{\sigma\in\{+,-\}} \frac{c_\sigma^2 d_\sigma^2(1-\theta-d_\sigma)}{(d_\sigma-(\theta-1))(a(\theta-1)-d_\sigma)(d_\sigma+a(\theta-1))}\big(a^n[\theta-1]^n\big)^2\\
&\quad-\frac{1}{2a^2\theta^2}\sum_{\sigma\in\{+,-\}} \frac{b_\sigma^2 d_\sigma^4(d_\sigma+a^2\theta)}{(d_\sigma-a\theta)(d_\sigma-a^2\theta)(d_\sigma+a\theta)}\big(a^n[\theta]^n\big)^2\\
&\quad+\sum_{\sigma\in\{+,-\}}\Bigg(\frac{b_\sigma^2(d_\sigma+a^2\theta)a^2\theta^2}{2(d_\sigma-a\theta)(d_\sigma-a^2\theta)(d_\sigma+a\theta)}\\
&\qquad\qquad-\frac{2c_\sigma^2 a^4(\theta-1)^4(1-\theta-d_\sigma)}{d_\sigma^2(d_\sigma-(\theta-1))(a(\theta-1)-d_\sigma)(d_\sigma+a(\theta-1))}\Bigg)\big([d_\sigma]^n\big)^2\\
&\quad+a^n\theta\sum_{\sigma\in\{+,-\}}\left(\frac{b_\sigma c_\tau d_\tau\big(\frac{1-\theta-d_\tau}{1-\theta+d_\tau}+\frac{d_\sigma+a^2\theta}{d_\sigma-a^2\theta}\big)}{d_\sigma(\theta-1)-d_\tau\theta}-\frac{b_\sigma c_\sigma(1-(1-a^2)\theta)}{1-a^2}\right)[d_\sigma]^n[\theta-1]^n\\
&\quad+\frac{a^n(\theta-1)^2}{\theta}\sum_{\sigma\in\{+,-\}}\left(\frac{b_\sigma c_\sigma(1-(1-a^2)\theta)}{1-a^2}-\frac{b_\tau c_\sigma d_\tau^2\big(\frac{1-\theta-d_\sigma}{1-\theta+d_\sigma}+\frac{d_\tau+a^2\theta}{d_\tau-a^2\theta}\big)}{d_\sigma(d_\tau(\theta-1)-d_\sigma\theta)}\right)[d_\sigma]^n[\theta]^n\\
&\quad+\frac{\theta((a^2-1)\theta+1)a^{2n}}{1-a^2}\left(\frac{b_+b_-(\theta-1)^2}{2\theta^2}\big([\theta]^n\big)^2-2c_+c_-\big([\theta-1]^n\big)^2\right)\\
&\quad+\frac{((1-a^2)\theta-1)(b_+b_-\theta^2-4c_+c_-(\theta-1)^2)}{2(1-a^2)\theta}[d_+]^n[d_-]^n.
\end{aligned}$$

To connect this representation with the limits found in Lemma 2.19, apply the substitution

$$\frac{b_+c_-d_-\big(\frac{1-\theta-d_-}{1-\theta+d_-}+\frac{d_++a^2\theta}{d_+-a^2\theta}\big)}{d_+(\theta-1)-d_-\theta}=\frac{\frac{c_-}{1-\theta+d_-}\,d_-(1-\theta-d_-)}{r\left(\frac{c_-}{1-a^2}\frac{1-a^2}{b_+}\,\theta+(\theta-1)\right)}+\frac{\frac{b_+}{d_+-a^2\theta}\,d_-(d_++a^2\theta)}{r\left(\theta+\frac{b_+}{1-a^2}\frac{1-a^2}{c_-}(\theta-1)\right)}.$$

Recall also the representations of v_1, v_2 and v_{n+1} from the proof of Lemma 2.19. The following results are only true if $\theta > 1/2$:

$$\lim_{n\to\infty} C_1 = 2/3, \qquad \lim_{n\to\infty} C_2 = 1/3, \qquad \lim_{n\to\infty} C_3 = 0$$

and

$$\begin{aligned}
&\lim_{n\to\infty}\boldsymbol{v}^\top\tilde{M}\,\boldsymbol{v}\\
&=\lim_{n\to\infty} v_1\lim_{n\to\infty}\big(\tilde{M}\,\boldsymbol{v}\big)_1+\lim_{n\to\infty} v_2\lim_{n\to\infty}\big(\tilde{M}\,\boldsymbol{v}\big)_2\\
&\quad+\lim_{n\to\infty}\sum_{k=1}^{4}\sum_{i=3}^{n} D_i^k+\lim_{n\to\infty} v_{n+1}\lim_{n\to\infty}\big(\tilde{M}\,\boldsymbol{v}\big)_{n+1}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{e^{-6\rho T}(2e^{3\rho T}+1)^2}{72\,\theta^2} + \frac{e^{-6\rho T}(2e^{3\rho T}+1)^2(\theta-1)(3\theta-1)}{72\,\theta^4} \\
&\quad + \frac{e^{-6\rho T}}{216\,\theta^4}\big(12e^{6\rho T}(\theta^4(2\rho T+3)-4\theta^2+4\theta-1) \\
&\qquad\qquad - 4e^{3\rho T}(2\theta^4+12\theta^2-12\theta+3)-\theta^4-12\theta^2+12\theta-3\Big)+0 \\
&= \frac{-e^{-6\rho T}-8e^{-3\rho T}+24\rho T+36}{216}.
\end{aligned}
$$

Now compute $\boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\,\boldsymbol{v}$, including again the case $\theta = 1/2$. Define

$$
C_4 \coloneqq \frac{(a^2(\theta-1)-\theta)-a\big(\frac{a(\theta-1)}{\theta}\big)^{n+1}}{(\theta-a(\theta-1))(a^2(\theta-1)-\theta)} \qquad \text{and}
$$

$$
C_5 \coloneqq \frac{a^2(\theta-1)(\theta+a(\theta-1))}{\theta^2(a^2(\theta-1)-\theta)}.
$$

It holds that

$$
\big(\boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\big)_1 = \frac{a}{\theta-a(\theta-1)}\Big(1-\Big(\frac{a(\theta-1)}{\theta}\Big)^n\Big)
$$

and, for $i = 2,3,\ldots,n$,

$$
\big(\boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\big)_i = C_4 a^i + C_5\Big(\frac{a(\theta-1)}{\theta}\Big)^{n-i},
$$

as well as

$$
\big(\boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\big)_{n+1} = \frac{a\big(a^n(\theta-a^2(\theta-1))\theta + a^2(\theta-1)\big(\theta-1+\big(\frac{a^2(\theta-1)}{\theta}\big)^n\big)-\theta^2\big)}{\theta(\theta-a(\theta-1))(\theta-a^2(\theta-1))}.
$$

For $i = 2,3,\ldots,n$, verify

$$
\begin{aligned}
&(\boldsymbol{w}_i^\top(\tilde{M}-\tilde{M}^\top))_i\, v_i \\
&= C_2(1-a)\Big(C_4 a^i + C_5\Big(\frac{a(\theta-1)}{\theta}\Big)^{n-i}\Big) \\
&\quad + 2C_1\sum_{\sigma\in\{+,-\}}\frac{c_\sigma d_\sigma}{a(\theta-1)}\Big(C_4\Big(\frac{a^2(\theta-1)}{d_\sigma}\Big)^i + C_5\Big(\frac{a(\theta-1)}{\theta}\Big)^n\Big(\frac{\theta}{d_\sigma}\Big)^i\Big)[d_\sigma]^n \\
&\quad + C_1\sum_{\sigma\in\{+,-\}}\frac{b_\sigma a^{n+1}\theta}{d_\sigma}\Big(C_4\Big(\frac{d_\sigma}{\theta}\Big)^i + C_5\Big(\frac{a(\theta-1)}{\theta}\Big)^n\Big(\frac{d_\sigma}{a^2(\theta-1)}\Big)^i\Big)[\theta]^n
\end{aligned}
$$

and then

$$
\begin{aligned}
&\sum_{i=2}^{n} (\boldsymbol{w}_i^\top(\tilde{M}-\tilde{M}^\top))_i\, v_i \\
&= C_2\Big(C_4(a^2-a^{n+1}) + C_5(1-a)\frac{a\theta(\theta-1) - a^n\theta^2\big(\frac{\theta-1}{\theta}\big)^n}{a(\theta-1)(\theta-a(\theta-1))}\Big) \\
&\quad + C_1 \sum_{\sigma\in\{+,-\}} \Big(C_4\Big(\frac{2c_\sigma a^3(\theta-1)}{d_\sigma - a^2(\theta-1)} - \frac{b_\sigma a^{n+1}\theta}{\theta-d_\sigma}\Big) \\
&\qquad\qquad + C_5\Big(\frac{2c_\sigma a^n\theta^2\big(\frac{\theta-1}{\theta}\big)^n}{a(\theta-1)(d_\sigma-\theta)} - \frac{b_\sigma a\theta}{a^2(\theta-1)-d_\sigma}\Big)\Big)[d_\sigma]^n \\
&\quad - C_1 a\theta\Big(\frac{C_4 a^n(\theta+1)}{\theta} + \frac{C_5(1-2a^2)\big(\frac{a^2(\theta-1)}{\theta}\big)^n}{a^2(1-a^2)(\theta-1)}\Big)[\theta]^n \\
&\quad - 2C_1\Big(C_4 a^{2n+1} + \frac{C_5\theta(1-(a^2-1)\theta)a^n}{a(1-a^2)(\theta-1)}\Big)[\theta-1]^n.
\end{aligned}
$$

To connect this representation with the limits found in Lemma 2.19, apply the substitutions

$$
\frac{\big(\frac{\theta-1}{\theta}\big)^n}{d_+-\theta} = \frac{\big(\frac{\theta-1}{\theta}\big)^n}{1-a^2}\,\frac{1-a^2}{b_+}\,\frac{b_+}{d_+-\theta} \qquad \text{and}
$$

$$
\frac{[d_-]^n}{d_- - a(\theta-1)} = \frac{[d_-]^n}{1-a^2}\,\frac{1-a^2}{c_-}\,\frac{c_-}{d_- - a(\theta-1)}.
$$

The following results are only true if $\theta > 1/2$:

$$
\lim_{n\to\infty} C_4 = 1, \qquad \lim_{n\to\infty} C_5 = \frac{(1-\theta)(2\theta-1)}{\theta^2}
$$

and

$$
\begin{aligned}
&\lim_{n\to\infty} \boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\,\boldsymbol{v} \\
&= \lim_{n\to\infty}\big(\boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\big)_1 \lim_{n\to\infty} v_1 + \lim_{n\to\infty}\sum_{i=2}^{n}(\boldsymbol{w}_i^\top(\tilde{M}-\tilde{M}^\top))_i\, v_i \\
&\quad + \lim_{n\to\infty}\big(\boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\big)_{n+1} \lim_{n\to\infty} v_{n+1} \\
&= \frac{e^{-3\rho T}+2}{6\theta} + \frac{2(2\theta-1)-(\theta+1)e^{-3\rho T}}{6\theta} + 0 \\
&= \frac{-e^{-3\rho T}+4}{6}.
\end{aligned}
$$

Finally, compute $\boldsymbol{w}^\top \tilde{M}\, \boldsymbol{w}$, including again the case $\tilde{\gamma} = 1/2$. With the help of Lemma 2.12 (i), this is easier:

$$\begin{aligned}
\boldsymbol{w}^\top \tilde{M}\, \boldsymbol{w} &= \frac{1}{2}\sum_{i=1}^{n} w_i^2 + \frac{w_{n+1}^2}{2} + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_i w_j a^{j-i} + w_{n+1}\sum_{i=1}^{n} w_i a^{n+1-i} \\
&= \frac{\theta^2 - 2a(1-a^2)\theta(\theta-1) - a^2(2-a^2)(\theta-1)^2}{2(\theta - a(\theta-1))^3(\theta + a(\theta-1))} \\
&\quad + \frac{(1-a^2)n}{2(\theta - a(\theta-1))^2} - \frac{a^{n+2}(1-(1-a)\theta)(\theta-1)\left(\frac{\theta-1}{\theta}\right)^n}{\theta(\theta - a(\theta-1))^3} \\
&\quad + \frac{a^{2(n+2)}(\theta-1)^2(2\theta-1)\left(\frac{\theta-1}{\theta}\right)^{2n}}{2(\theta - a(\theta-1))^3(\theta + a(\theta-1))\theta^2}.
\end{aligned}$$

Notice that $(1-a^2)n = (1+a)(1-a)m(T)$. Letting $\theta > 1/2$ and taking limits concludes the proof. □

Proof of Theorem 2.6 (i). Let $\theta > 1/2$. Recall from Lemma 2.12 (ii) that

$$\lim_{n\to\infty} \mathbf{1}^\top \boldsymbol{w} = \lim_{n\to\infty} \sum_{i=1}^{n+1} w_i = \rho T + 1.$$

The limit of $\mathbf{1}^\top \boldsymbol{v} = \sum_{i=1}^{n+1} v_i$ is obtained in Part 1 of the proof of Lemma 2.17 for the case $\theta = 1$, and (with the same result) in Part 2 for the case $\theta > 1/2$, $\theta \neq 1$. The limits of $\boldsymbol{v}^\top \tilde{M}\, \boldsymbol{v}, \boldsymbol{w}^\top(\tilde{M} - \tilde{M}^\top)\, \boldsymbol{v}$, and $\boldsymbol{w}^\top \tilde{M}\, \boldsymbol{w}$ can be found in Lemma 2.22. Plug into the representation of expected costs from Lemma 2.21 to conclude the proof.

Lemma 2.23. *If $\theta = 1/2$, then*

$$\begin{aligned}
\lim_{\substack{n\to\infty \\ n\,\text{even}}} \boldsymbol{v}^\top \tilde{M}\, \boldsymbol{v} &= \frac{2e^{6\rho T}(3\rho T + 5) + e^{3\rho T} + 3\rho T + 7}{54e^{6\rho T} + 27}, \\
\lim_{\substack{n\to\infty \\ n\,\text{even}}} \boldsymbol{w}^\top(\tilde{M} - \tilde{M}^\top)\, \boldsymbol{v} &= \frac{4e^{6\rho T} - 6e^{5\rho T} + e^{3\rho T} - 3e^{-\rho T} + 4}{6e^{6\rho T} + 3} \qquad \textit{and} \\
\lim_{\substack{n\to\infty \\ n\,\text{even}}} \boldsymbol{w}^\top \tilde{M}\, \boldsymbol{w} &= e^{-\rho T} + \rho T + 1.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\lim_{\substack{n\to\infty \\ n\,\text{odd}}} \boldsymbol{v}^\top \tilde{M}\, \boldsymbol{v} &= \frac{2e^{6\rho T}(3\rho T + 5) - 3e^{3\rho T} - 3\rho T - 7}{54e^{6\rho T} - 27}, \\
\lim_{\substack{n\to\infty \\ n\,\text{odd}}} \boldsymbol{w}^\top(\tilde{M} - \tilde{M}^\top)\, \boldsymbol{v} &= \frac{-4e^{6\rho T} - 6e^{5\rho T} + 3e^{3\rho T} + 3e^{-\rho T} + 4}{-6e^{6\rho T} + 3} \qquad \textit{and} \\
\lim_{\substack{n\to\infty \\ n\,\text{odd}}} \boldsymbol{w}^\top \tilde{M}\, \boldsymbol{w} &= -e^{-\rho T} + \rho T + 1.
\end{aligned}$$

Proof. The representations of $\boldsymbol{v}^\top \tilde{M}\,\boldsymbol{v}$, $\boldsymbol{w}^\top(\tilde{M}-\tilde{M}^\top)\,\boldsymbol{v}$, and $\boldsymbol{w}^\top \tilde{M}\,\boldsymbol{w}$ obtained in the proof of Lemma 2.22 (for the case $\theta \neq 1$) are also valid for $\theta = 1/2$. Observe

$$\begin{aligned}\frac{b_+c_-\big(\frac{1-\theta-d_-}{1-\theta+d_-}+\frac{d_++a^2\theta}{d_+-a^2\theta}\big)}{d_+(\theta-1)-d_-\theta} &= (1-\theta-d_-)\frac{c_-}{d_- -(\theta-1)}\frac{b_+}{d_+(\theta-1)-d_-\theta}\\ &\quad + (d_++a^2\theta)\frac{b_+}{d_+-a^2\theta}\frac{c_-}{d_+(\theta-1)-d_-\theta}\end{aligned}$$

and, by definition of d_+ and d_-,

$$\frac{\frac{1-\theta-d_+}{1-\theta+d_+}+\frac{d_-+a^2\theta}{d_--a^2\theta}}{d_-(\theta-1)-d_+\theta} = -\frac{16(3(1+a^2)-2r)}{3(15(1+a^4)-4r(1+r)-2a^2(2r+17))}.$$

Furthermore,

$$\begin{aligned}\frac{C_5}{1-a^2} &= \frac{a^2(\theta-1)}{\theta^2(a^2(\theta-1)-\theta)}\,\frac{\theta+a(\theta-1)}{1-a^2},\\ \frac{C_5}{d_+-\theta} &= \frac{C_5}{1-a^2}\,\frac{1-a^2}{b_+}\,\frac{b_+}{d_+-\theta} \qquad\text{and}\\ \frac{C_5}{d_--a^2(\theta-1)} &= \frac{C_5}{1-a^2}\,\frac{1-a^2}{c_-}\,\frac{c_-}{d_--a^2(\theta-1)}.\end{aligned}$$

The expression for $\boldsymbol{w}^\top \tilde{M}\,\boldsymbol{w}$ simplifies to

$$-\frac{2\big(a^2(1-(-1)^n2a^n)-2a-1-(1+a)(1-a)n\big)}{(1+a)^2}.$$

Plug in the limits from Lemma 2.20 to conclude the proof. □

Proof of Theorem 2.6 (ii). Recall from Lemma 2.12 (iii) that, for $\theta = 1/2$,

$$\lim_{\substack{n\to\infty\\ n\,\text{even}}} \mathbf{1}^\top\boldsymbol{w} = e^{-\rho T}+\rho T+1 \qquad\text{and}\qquad \lim_{\substack{n\to\infty\\ n\,\text{odd}}} \mathbf{1}^\top\boldsymbol{w} = -e^{-\rho T}+\rho T+1.$$

The cluster points of $\mathbf{1}^\top\boldsymbol{v}$ for $\theta = 1/2$ are stated in (2.6). With these and the results from Lemma 2.23, proceed as in the proof of (i). This concludes the proof of Theorem 2.6. □

Chapter 3

A different approach to modeling transaction costs

I argue at the end of Section 2.4 that in continuous time models of optimal execution, transaction costs should apply to the instantaneous rate of trading, not to block trades. Only the former approach ensures that investors are unable to avoid transaction costs by choosing a trading strategy that is absolutely continuous. It will be shown in this chapter that transaction costs of this form make the problem of optimal execution much more tractable. In particular, a relatively weak assumption on the absence of price manipulation strategies and some technical assumptions on the decay kernel are sufficient to ensure existence and uniqueness of a Nash equilibrium of optimal execution strategies for an arbitrary number of investors.

I present the general model in Section 3.1 and show existence and uniqueness of a Nash equilibrium. Each investor obtains his equilibrium strategy by solving a Fredholm integral equation of the second kind.

In Section 3.2, I consider the special case where only one strategic investor trades. In this case, the Fredholm integral equation characterizing the optimal execution strategy has a constant free term. I present closed-form representations of the optimal execution strategy for specific decay kernels, including capped linear decay, and briefly discuss numerical simulation methods. Since closed-form solutions are unknown for many interesting decay kernels, I also examine under which conditions qualitative features of the optimal execution strategy can be derived from information about the decay kernel without solving the Fredholm integral equation explicitly.

In Section 3.3, I derive a closed-form representation of equilibrium strategies for $n+1$ strategic investors under the assumption that transient price impact decays exponentially. The derivation develops ideas that will also be useful in Chapter 4: The system of Fredholm integral equations characterizing the Nash equilibrium is transformed into a system of ordinary differential equations. The latter is then solved in terms of a matrix exponential and an invertible matrix, and analyzed further by means of an eigendecomposition.

In Section 3.4, I provide an economic analysis of order anticipation strategies based on the closed-form representation obtained previously. One investor liquidates a sell order, and n opportunistic investors pursue order anticipation strategies to benefit from the liquidating investor's price impact. I study how the opportunistic investors affect the liquidating investor's optimal strategy and his expected costs. I further test the claim by Brunnermeier and Pedersen (2005) that opportunistic traders cause price overshooting. For many choices of parameters, it must be refuted. In fact, opportunistic investors often produce the opposite effect and reduce the price drop caused by a sell order. I provide two possible explanations: Price overshooting does not occur if price impact is transient and sufficiently short-lived; or price overshooting is prevented by quadratic transaction costs.

In Section 3.5, I propose an extension of the model in which opportunistic investors have additional time to build up and unwind positions before and after the liquidating investor trades.

Parts of this chapter are published as a working paper (Strehle, 2017). I wish to thank two anonymous referees for helpful remarks.

3.1 Existence of a Nash equilibrium

Consider a continuous time market for a single financial asset. The asset is traded by $n+1$ strategic investors over a time period $[0,T]$. In the absence of strategic trading, the asset price S^0 is modeled as a right-continuous martingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ satisfying the usual conditions. Assume that $\mathcal{F}_0$ is $\mathbb{P}$-trivial.

The strategic investors $i = 0, 1, \ldots, n$ control their instantaneous rate of trading $\alpha_i(t)\,\mathrm{d}t$, where a positive sign of $\alpha_i(t)$ corresponds to a buy order. Each investor i must trade a fixed net amount x_i^0 until time T. Consequently, $\alpha_i \in L^2[0,T]$ is called an *(admissible) strategy* (for investor i) if it is progressively measurable and satisfies the *liquidation constraint* $\int_0^T \alpha_i(t)\,\mathrm{d}t = x_i^0$.

Define the *remaining net amount* $X_i(t) \coloneqq x_i^0 - \int_0^t \alpha_i(s)\,\mathrm{d}s$ (this corresponds to the definition of an admissible strategy X in Section 2.4). In terms of X_i, the liquidation constraint reads $X_i(T) = 0$. Notice that x_i^0 and α_i together determine X_i and vice versa. Therefore, an absolutely continuous function $X_i\colon [0,T] \to \mathbb{R}$ will also be referred to as an admissible strategy if $X_i(0)$ equals x_i^0 and $\alpha_i \coloneqq -\frac{\mathrm{d}}{\mathrm{d}t}X_i$ is an admissible strategy.

Every strategic investor impacts the asset price. Price impact is assumed to be linear and transient. It is modeled via a square-integrable *decay kernel* $G\colon [0,\infty) \to [0,\infty)$. Suppose the strategic investors pursue strategies $\alpha \coloneqq (\alpha_0, \alpha_1, \ldots, \alpha_n)$. Then the asset price evolves according to

$$S(t) = S(t;\alpha) \coloneqq S^0(t) + \int_0^t G(t-s) \sum_{i=0}^n \alpha_i(s)\,\mathrm{d}s, \qquad t \in [0,T]. \tag{3.1}$$

Investor i's costs from price impact are $\int_0^T \alpha_i(t) S(t;\alpha)\,\mathrm{d}t$. In addition, each investor i

incurs quadratic transaction costs $\frac{\gamma_i}{2}\alpha_i(t)^2\,\mathrm{d}t$, where $\gamma_i \geq 0$ (see pp. 17–18 for a detailed discussion). Notice that the model explicitly allows for different levels of transaction costs for different investors.

In total, investor i has the following *costs of execution*:

$$J_i[\alpha_i \,|\, \alpha_{-i}] \coloneqq \int_0^T \big(\frac{\gamma_i}{2}\alpha_i(t)^2 + \alpha_i(t)S(t;\alpha)\big)\,\mathrm{d}t, \tag{3.2}$$

where $\alpha_{-i} \coloneqq (\alpha_0, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_n)$.

Assume that each investor is risk-neutral and therefore minimizes expected costs of execution. Assume further that all model parameters, including n and $x^0 \coloneqq (x_0^0, x_1^0, \dots, x_n^0)$, are known to each investor. Integration by parts shows that, for a given right-continuous martingale S^0, the term

$$\mathbb{E}\Big[\int_0^T \alpha_i(t)S^0(t)\,\mathrm{d}t\Big] = -x_i^0 S^0(0) - \mathbb{E}\Big[\int_0^T X_i(t)\,\mathrm{d}S^0(t)\Big] = -x_i^0 S^0(0)$$

is the same for all admissible strategies α_i. Hence there is no loss of generality in assuming that $S^0(t) = 0$ for all $t \in [0, T]$. Notice that this would be very different if traders were risk-averse (Almgren and Chriss, 2001) or the liquidation constraints were private information (Moallemi et al., 2012; Choi et al., 2015).

This model may be viewed as a multi-investor version of the continuous time model in Obizhaeva and Wang (2013). A similar model is also studied in Zhang's (2014) doctoral thesis, but with temporary and permanent price impact only.

The $(n+1)$-dimensional function $\alpha^* = (\alpha_0^*, \alpha_1^*, \dots, \alpha_n^*)$ is called a *Nash equilibrium (in the class of admissible strategies)* if for all $i = 0, 1, \dots, n$, the strategy α_i^* is admissible, and $\mathbb{E}[J_i[\alpha_i^* \,|\, \alpha_{-i}^*]] \leq \mathbb{E}[J_i[\alpha_i \,|\, \alpha_{-i}^*]]$ for every admissible strategy α_i for investor i. In this case, α_i^* is called an *optimal strategy* (for investor i). Furthermore, α^* is called a *Nash equilibrium in the class of deterministic strategies* if each strategy α_i^* is deterministic, and $J_i[\alpha_i^* \,|\, \alpha_{-i}^*] \leq J_i[\alpha_i \,|\, \alpha_{-i}^*]$ for every deterministic admissible strategy α_i for investor i.

Remark 3.1. *The current analysis limits itself to (stochastic) open-loop strategies $\alpha_i(\omega, t)$, instead of closed-loop strategies*

$$\alpha_i\big(\omega, t, \alpha_0(\omega, t), \dots, \alpha_{i-1}(\omega, t), \alpha_{i+1}(\omega, t), \dots, \alpha_n(\omega, t)\big).$$

In closed-loop Nash equilibria, investors still react optimally if another investor departs from equilibrium. In open-loop Nash equilibria, this is typically not the case: Each investor implicitly assumes that all other investors will pursue their respective equilibrium strategies. Carmona and Yang (2011) show that this affects the equilibrium itself: An open-loop Nash equilibrium need not be a closed-loop Nash equilibrium and vice versa. Closed-loop Nash equilibria are an appealing concept, but notoriously difficult to find. See nonetheless the aforementioned paper for numerical simulations of open-loop and closed-loop Nash equilibria in a model of optimal execution under temporary price impact. Closed-loop equilibria for the model in Section 2.1 are stud-

ied in Section 3.2.3 of Zhang's (2014) doctoral thesis. For a detailed discussion of open-loop and closed-loop equilibria in the context of stochastic differential games, see Section 2.2 in Yeung and Petrosjan (2006).

Not every decay kernel G is sensible from an economic point of view. Suppose there is only one investor $i = 0$ and assume for simplicity that $\gamma_0 = 0$. The investor's costs from price impact are

$$\begin{aligned}\int_0^T \alpha_0(t) S(t;\alpha_0)\,\mathrm{d}t &= \int_0^T \int_0^t G(t-s)\alpha_0(t)\alpha_0(s)\,\mathrm{d}s\,\mathrm{d}t \\ &= \frac{1}{2}\int_0^T \int_0^T G(|t-s|)\alpha_0(t)\alpha_0(s)\,\mathrm{d}s\,\mathrm{d}t.\end{aligned} \tag{3.3}$$

If there is a strategy α_0 that makes (3.3) negative, this means that the investor can exploit his own price impact to generate arbitrarily large expected profits: the decay kernel G admits *price manipulation* in the sense of Huberman and Stanzl (2004). Gatheral (2010) points out that price manipulation strategies do not constitute classical arbitrage, because their profitability is affected by random fluctuations in the asset price S^0. Instead, they belong to the larger class of statistical arbitrage strategies. Notice also the difference between price manipulation and order anticipation: A price manipulation strategy generates profits from its own price impact, an order anticipation strategy generates profits from another investor's price impact.

If (3.3) is nonnegative for every $\alpha_0 \in L^2[0,T]$ and every $T > 0$, then G is said to be of *positive type* (Mercer, 1909). Assume from now on that G is of positive type.

In single-investor models of optimal execution under transient price impact (Gatheral et al., 2012; Obizhaeva and Wang, 2013) and the two-investor model in Section 2.4, impulse trades—i.e., jumps in X_i—are optimal if transaction costs are zero. But in the current model, such jumps are inadmissible. This suggests that no Nash equilibrium exists as soon as $\gamma_i = 0$ for some i. Assume from now on that $\gamma_i > 0$ for all $i = 0, 1, \dots, n$.

Under these assumptions on G and γ_i, uniqueness of Nash equilibria is a simple consequence of the convexity of the cost functionals J_i. The following results are easily adapted from Proposition 4.8 and Lemma 4.9 in Schied et al. (2017), or from Lemmas 2.1.3 and 3.3.12 in Zhang's (2014) doctoral thesis. Notice that uniqueness should be understood as uniqueness $\mathcal{B}([0,T]) \otimes \mathbb{P}$-almost everywhere.

Lemma 3.2.

(i) There is at most one Nash equilibrium in the class of admissible strategies.

(ii) A Nash equilibrium in the class of deterministic strategies is also a Nash equilibrium in the class of admissible strategies.

Proof. (i) Suppose $\alpha^0 = (\alpha_0^0, \alpha_1^0, \dots, \alpha_n^0)$ and $\alpha^1 = (\alpha_0^1, \alpha_1^1, \dots, \alpha_n^1)$ are Nash equilibria in the class of admissible strategies.

For $z \in [0,1]$, define $\alpha^z := (1-z)\alpha^0 + z\alpha^1$ and

$$f(z) := \sum_{i=0}^{n} \big(\mathbb{E}\big[J_i[\alpha_i^z \,|\, \alpha_{-i}^0]\big] + \mathbb{E}\big[J_i[\alpha_i^{1-z} \,|\, \alpha_{-i}^1]\big]\big).$$

For every $i = 0, 1, \dots, n$, the strategies α_i^0 and α_i^1 are optimal reactions to α_{-i}^0 and α_{-i}^1, respectively. Hence

$$\frac{\mathrm{d}}{\mathrm{d}z} f(z)\bigg|_{z=0} \geq 0.$$

But interchanging differentiation and integration shows

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}z} f(z)\bigg|_{z=0} \\
&= -\sum_{i=0}^{n} \mathbb{E}\Big[\int_0^T \Big(\gamma_i\big(\alpha_i^0(t) - \alpha_i^1(t)\big)^2 \\
&\qquad + \big(\alpha_i^0(t) - \alpha_i^1(t)\big) \int_0^t G(t-s)\big(\alpha_i^0(s) - \alpha_i^1(s)\big)\,\mathrm{d}s \\
&\qquad + \big(\alpha_i^0(t) - \alpha_i^1(t)\big) \int_0^t G(t-s) \sum_{j=0}^{n} \big(\alpha_j^0(s) - \alpha_j^1(s)\big)\,\mathrm{d}s\Big)\,\mathrm{d}t\Big] \\
&= -\sum_{i=0}^{n} \gamma_i\, \mathbb{E}\Big[\int_0^T \big(\alpha_i^0(t) - \alpha_i^1(t)\big)^2\,\mathrm{d}t\Big] \\
&\quad - \frac{1}{2}\sum_{i=0}^{n} \mathbb{E}\Big[\int_0^T \int_0^T G(|t-s|)\big(\alpha_i^0(t) - \alpha_i^1(t)\big)\big(\alpha_i^0(s) - \alpha_i^1(s)\big)\,\mathrm{d}s\,\mathrm{d}t\Big] \\
&\quad - \frac{1}{2}\mathbb{E}\Big[\int_0^T \int_0^T G(|t-s|) \sum_{i=0}^{n} \big(\alpha_i^0(t) - \alpha_i^1(t)\big) \sum_{i=0}^{n} \big(\alpha_i^0(s) - \alpha_i^1(s)\big)\,\mathrm{d}s\,\mathrm{d}t\Big].
\end{aligned}$$

Recall that $\gamma_i > 0$ for all $i = 0, 1, \dots, n$, and that G is of positive type. Hence the last expression can only be nonnegative if α^0 and α^1 are identical $\mathcal{B}([0,T]) \otimes \mathbb{P}$-almost everywhere.

(ii) Suppose $\alpha^* = (\alpha_0^*, \alpha_1^*, \dots, \alpha_n^*)$ is a Nash equilibrium in the class of deterministic strategies. Let $i = 0, 1, \dots, n$. For every admissible strategy α_i for investor i, it must be true that $J_i[\alpha_i^* \,|\, \alpha_{-i}^*] \leq J_i[\alpha_i(\omega) \,|\, \alpha_{-i}^*]$ for almost all $\omega \in \Omega$. This implies $J_i[\alpha_i^* \,|\, \alpha_{-i}^*] \leq \mathbb{E}\big[J_i[\alpha_i \,|\, \alpha_{-i}^*]\big]$, hence α_i^* is also optimal in the class of admissible strategies. □

The next step is to show that Nash equilibria are characterized by $n+1$ Fredholm integral equations of the second kind. Existence of a Nash equilibrium then follows from the invertibility of the corresponding integral operator.

For $\eta \in \mathbb{R}^{n+1}$, let $\boldsymbol{\eta}$ denote the $(n+1)$-dimensional constant function $\boldsymbol{\eta}(t) = \eta$. With slight abuse of notation, let $\mathbf{0}$ and $\mathbf{1}$ denote the $(n+1)$-dimensional constant functions $\mathbf{0}(t) = (0, 0, \dots, 0)$ and $\mathbf{1}(t) = (1, 1, \dots, 1)$. Define $\Gamma := \operatorname{diag}(\gamma_0, \gamma_1, \dots, \gamma_n)$, and a linear operator F on $L^2([0,T];\mathbb{R}^{n+1})$ via

$$\begin{aligned}(F\alpha)(t) \coloneqq \Gamma\alpha(t) + \Big(\int_0^t G(t-s)\alpha(s)^\top \mathbf{1}(s)\,\mathrm{d}s\Big)\,\mathbf{1}(t) \\ + \int_t^T G(s-t)\alpha(s)\,\mathrm{d}s.\end{aligned} \tag{3.4}$$

The connection between Nash equilibria and the operator F is based on the fundamental lemma of the calculus of variations, which is reproduced here for the sake of completeness.

Lemma 3.3. (Fundamental lemma of the calculus of variations). *Let $f \in L^2[0,T]$. Suppose $\int_0^T f(t)g(t)\,\mathrm{d}t = 0$ for every $g \in L^2[0,T]$ satisfying $\int_0^T g(t)\,\mathrm{d}t = 0$. Then there is a constant $z \in \mathbb{R}$ such that $f(t) = z$ for almost all $t \in [0,T]$.*

Proof. The proof is taken from Gelfand and Fomin (1963, Lemma 2).
Let $z \coloneqq \frac{1}{T}\int_0^T f(t)\,\mathrm{d}t$ and define $g \in L^2[0,T]$ via $g(t) \coloneqq f(t) - z$. It follows immediately from the definition of g that $\int_0^T g(t)\,\mathrm{d}t = 0$. Since

$$\int_0^T (f(t)-z)^2\,\mathrm{d}t = \int_0^T f(t)g(t)\,\mathrm{d}t - z\int_0^T g(t)\,\mathrm{d}t = 0,$$

conclude that $f(t) = z$ for almost all $t \in [0,T]$. □

The following lemma connects Nash equilibria with the operator F.

Lemma 3.4. *The function $\alpha^* = (\alpha_0^*, \alpha_1^*, \ldots, \alpha_n^*) \in L^2([0,T];\mathbb{R}^{n+1})$ is a Nash equilibrium in the class of deterministic strategies if and only if*

(i) $\int_0^T \alpha_i^*(t)\,\mathrm{d}t = x_i^0$ for every $i = 0,1,\ldots,n$, and

(ii) there is an $\eta = (\eta_0, \eta_1, \ldots, \eta_n) \in \mathbb{R}^{n+1}$ such that $(F\alpha^*)(t) = \eta$ for almost all $t \in [0,T]$.

In this case, $\eta_i x_i^0 \geq J_i[\alpha_i^* \,|\, \alpha_{-i}^*]$ for every $i = 0,1,\ldots,n$.

Proof. Define the linear subspace $B \coloneqq \{\beta \in L^2[0,T] \mid \int_0^T \beta(t)\,\mathrm{d}t = 0\}$.
Necessity: Suppose $\alpha^* = (\alpha_0^*, \alpha_1^*, \ldots, \alpha_n^*)$ is a Nash equilibrium in the class of deterministic strategies. Let $i = 0,1,\ldots,n$. For every $y \in \mathbb{R}$ and $\beta \in B$, the function $\alpha_i^* + y\beta$ is a deterministic admissible strategy for investor i. It follows that a necessary condition for the optimality of α_i^* is

$$\begin{aligned}0 &= \frac{\mathrm{d}}{\mathrm{d}y} J_i[\alpha_i^* + y\beta \,|\, \alpha_{-i}^*]\bigg|_{y=0} \\ &= \int_0^T \Big(\beta(t)\big(\gamma_i\alpha_i^*(t) + S(t;\alpha^*)\big) + \int_0^t G(t-s)\alpha_i^*(t)\beta(s)\,\mathrm{d}s\Big)\,\mathrm{d}t.\end{aligned}$$

Conclude with Fubini's theorem that

$$\int_0^T \int_0^t G(t-s)\alpha_i^*(t)\beta(s)\,\mathrm{d}s\,\mathrm{d}t = \int_0^T \int_t^T G(s-t)\beta(t)\alpha_i^*(s)\,\mathrm{d}s\,\mathrm{d}t.$$

Hence

$$\begin{aligned} 0 &= \int_0^T \beta(t)\Big(\gamma_i\alpha_i^*(t) + \int_0^t G(t-s)\sum_{i=0}^n \alpha_i^*(s)\,\mathrm{d}s + \int_t^T G(s-t)\alpha_i^*(s)\Big)\,\mathrm{d}s \\ &= \int_0^T \beta(t)(F\alpha^*)_i(t)\,\mathrm{d}t. \end{aligned}$$

The fundamental lemma of the calculus of variations (Lemma 3.3) implies that $(F\alpha^*)_i$ is constant for almost all $t \in [0,T]$.
Sufficiency: Let $\alpha^* = (\alpha_0^*, \alpha_1^*, \ldots, \alpha_n^*)$ be such that for every $i = 0, 1, \ldots, n$, the liquidation constraint $\int_0^T \alpha_i^*(t)\,\mathrm{d}t = x_i^0$ is satisfied and $(F\alpha^*)_i(t) = \eta_i$ for almost all $t \in [0,T]$ for some $\eta = (\eta_0, \eta_1, \ldots, \eta_n) \in \mathbb{R}^{n+1}$.
Let $i = 0, 1, \ldots, n$. The liquidation constraint implies that any deterministic admissible strategy α_i for investor i can be written as $\alpha_i = \alpha_i^* + \beta$ for some $\beta \in B$. Conclude with Fubini's theorem that

$$\begin{aligned} &J_i[\alpha_i \,|\, \alpha_{-i}^*] \\ &= \int_0^T \Big(\frac{\gamma_i}{2}\alpha_i^*(t)^2 + \alpha_i^*(t)\int_0^t G(t-s)\sum_{j=0}^n \alpha_j^*(s)\Big)\,\mathrm{d}s\,\mathrm{d}t \\ &\quad + \int_0^T \beta(t)\Big(\gamma_i\alpha_i^*(t) + \int_0^t G(t-s)\sum_{j=0}^n \alpha_j^*(s)\,\mathrm{d}s + \int_t^T G(s-t)\alpha_i^*(s)\,\mathrm{d}s\Big)\,\mathrm{d}t \\ &\quad + \int_0^T \Big(\frac{\gamma_i}{2}\beta(t)^2 + \beta(t)\int_0^t G(t-s)\beta(s)\,\mathrm{d}s\Big)\,\mathrm{d}t \\ &= J[\alpha_i^* \,|\, \alpha_{-i}^*] + \eta_i \int_0^T \beta(t)\,\mathrm{d}t + \frac{1}{2}\int_0^T \Big(\gamma_i\beta(t)^2 + \int_0^T G(|t-s|)\beta(t)\beta(s)\,\mathrm{d}s\Big)\,\mathrm{d}t \\ &\geq J[\alpha_i^* \,|\, \alpha_{-i}^*]. \end{aligned}$$

Hence α_i^* is the optimal strategy for investor i, given that the other investors pursue α_{-i}^*. This is true for all i, showing that α^* is a Nash equilibrium.
Furthermore, for all $i = 0, 1, \ldots, n$,

$$\begin{aligned} &J_i[\alpha_i^* \,|\, \alpha_{-i}^*] \\ &= \int_0^T \Big(\frac{\gamma_i}{2}\alpha_i^*(t)^2 + \alpha_i^*(t)\int_0^t G(t-s)\sum_{j=0}^n \alpha_j^*(s)\,\mathrm{d}s\Big)\,\mathrm{d}t \\ &= \int_0^T \alpha_i^*(t)(F\alpha^*)_i(t)\,\mathrm{d}t - \frac{1}{2}\int_0^T \alpha_i^*(t)\Big(\gamma_i\alpha_i^*(t) + \int_0^T G(|t-s|)\alpha_i^*(s)\,\mathrm{d}s\Big)\,\mathrm{d}t \\ &\leq \eta_i x_i^0. \end{aligned}$$

□

Remark 3.5. *The optimality condition* $(F\alpha^*)(t) = \eta$*,* $t \in [0,T]$*, can be written as*

$$\gamma_i \alpha_i^*(t) + \int_0^T G(|t-s|)\alpha_i^*(s)\,\mathrm{d}s = \eta_i - \int_0^t G(t-s)\sum_{j\neq i}\alpha_j^*(s)\,\mathrm{d}s, \quad t\in[0,T], \quad (3.5)$$

for $i = 0, 1, \ldots, n$*. For fixed* $(\alpha_0^*, \ldots, \alpha_{i-1}^*, \alpha_{i+1}^*, \ldots, \alpha_n^*)$*, Equation* (3.5) *is a one-dimensional Fredholm integral equation of the second kind. The single-investor version of* (3.5) *will be the analyzed in detail in Section 3.2. Fredholm integral equations are connected with the more prevalent Euler-Lagrange equations in the following way: Consider the constant decay kernel* $G(t) = 1$*. Then, for admissible strategies* $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_n)$*, the asset price evolves according to*

$$S(t) = \int_0^t \big(\alpha_i(s) + \sum_{j\neq i}\alpha_j(s)\big)\,\mathrm{d}s.$$

This turns the minimization of expected costs $J_i[\alpha_i \,|\, \alpha_{-i}]$ *into a classical problem in the calculus of variations. The corresponding Euler-Lagrange equation characterizing the optimal strategy* α_i^* *is*

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\big[\gamma_i\alpha_i^* + S\big] - \alpha_i^*.$$

A straightforward calculation shows that this is the t*-derivative of* (3.5)*. But as soon as the decay kernel is not constant, the nonlocal term* $\int_t^T G(s-t)\alpha_i^*(s)\,\mathrm{d}s$ *prevents the derivation of a "proper" (i.e., local) Euler-Lagrange equation from* (3.5)*. One might suspect that the minimization of expected costs can still be performed with the help of a Euler-Lagrange equation by considering the two-dimensional process* (α_i^*, S) *instead of* α_i^**. But* S *is a function of* α_i^* *and a "chain rule" applies. The additional nonlocal term introduced into the Euler-Lagrange equation by this chain rule is just* $\int_t^T G(|t-s|)\alpha_i^*(s)\,\mathrm{d}s$ *(Avron, 2003). Hence Fredholm integral equations, not Euler-Lagrange equations, are the appropriate tool for solving problems of optimal execution under transient price impact.*

Existence of a Nash equilibrium in the class of deterministic strategies is now shown by proving that F is invertible and invoking the uniqueness result from Lemma 3.2 (i).

Theorem 3.6. *There is a unique Nash equilibrium* α^* *in the class of admissible strategies. It is deterministic.*

Proof. Uniqueness has been shown in Lemma 3.2 (i).
Define F as in (3.4). Let $\langle\cdot,\cdot\rangle$ denote the L^2-inner product on $[0,T]$, and $\|\cdot\|$ its induced norm. Without loss of generality, let $\gamma_0 \le \gamma_i \le \gamma_n$ for all $i = 0, 1, \ldots, n$.
Conclude with the Cauchy-Schwarz inequality and Jensen's inequality that F is bounded.

Indeed, for every $\alpha \in L^2([0,T];\mathbb{R}^{n+1})$,

$$
\begin{aligned}
&\|F\alpha\| \\
&\le \|\Gamma\alpha\| + \Big(\int_0^T \Big(\int_0^t G(t-s)\sum_{i=0}^n \alpha_i(s)\,\mathrm{d}s\Big)^2 (n+1)\,\mathrm{d}t\Big)^{1/2} \\
&\quad + \Big(\sum_{i=0}^n \int_0^T \Big(\int_t^T G(s-t)\alpha_i(s)\,\mathrm{d}s\Big)^2 \mathrm{d}t\Big)^{1/2} \\
&\le \gamma_n\|\alpha\| \\
&\quad + \Big(\int_0^T\int_0^T G(|t-s|)^2\,\mathrm{d}s\,\mathrm{d}t\Big)^{1/2}\Big(\sqrt{n+1}\Big(\int_0^T \Big(\sum_{i=0}^n \alpha_i(s)\Big)^2 \mathrm{d}s\Big)^{1/2} + \|\alpha\|\Big) \\
&\le \gamma_n\|\alpha\| + (n+2)\Big(\int_0^T\int_0^T G(|t-s|)^2\,\mathrm{d}s\,\mathrm{d}t\Big)^{1/2}\|\alpha\|.
\end{aligned}
$$

Recall that G is square-integrable. Apply Fubini's theorem and recall that G is of positive type to see that

$$
\begin{aligned}
\langle F\alpha,\alpha\rangle &= \sum_{i=0}^n \gamma_i \int_0^T \alpha_i(t)^2\,\mathrm{d}t + \frac{1}{2}\int_0^T\int_0^T G(|t-s|)\sum_{i=0}^n \alpha_i(t)\sum_{i=0}^n \alpha_i(s)\,\mathrm{d}s\,\mathrm{d}t \\
&\quad + \frac{1}{2}\sum_{i=0}^n \int_0^T\int_0^T G(|t-s|)\alpha_i(t)\alpha_i(s)\,\mathrm{d}s\,\mathrm{d}t \\
&\ge \gamma_0\|\alpha\|^2.
\end{aligned}
$$

Hence $\|F\alpha\|\,\|\alpha\| \ge \gamma_0\|\alpha\|^2$, showing that F is also bounded from below.
The adjoint F^* of F is given by

$$
(F^*\alpha)(t) = \Gamma\alpha(t) + \Big(\int_t^T G(s-t)\alpha(s)^\top \mathbf{1}(s)\,\mathrm{d}s\Big)\,\mathbf{1}(t) + \int_0^t G(t-s)\alpha(s)\,\mathrm{d}s.
$$

The same arguments show that F^* is bounded from above and below. Hence F is invertible.
Now define a linear operator $A\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ via

$$
A\eta \coloneqq \int_0^T (F^{-1}\boldsymbol{\eta})(t)\,\mathrm{d}t.
$$

A is invertible: Suppose $\eta^0 \in \mathbb{R}^{n+1}$ is such that $\eta^0 \in \ker(A)$, i.e., $A\eta^0 = (0,0,\dots,0)$. Define $\alpha^0 \coloneqq F^{-1}\boldsymbol{\eta}^0 \in L^2([0,T];\mathbb{R}^{n+1})$. Then

$$
\int_0^T \alpha^0(t)\,\mathrm{d}t = \int_0^T (F^{-1}\boldsymbol{\eta}^0)(t)\,\mathrm{d}t = A\eta^0 = (0,0,\dots,0).
$$

Furthermore, $F\alpha^0 = \boldsymbol{\eta}^0$. Conclude with Lemma 3.4 (i) that α^0 is a Nash equilibrium if (and only if) all liquidation constraints $x^0 = (x_0^0, x_1^0, \dots, x_n^0)$ are zero. But it is easy to check that in this case, the constant function $\mathbf{0}$ is also a Nash equilibrium. Hence, by Lemma 3.2 (i), it must be true that $\alpha^0 = \mathbf{0}$. Because $\ker(F) = \{\mathbf{0}\}$, it follows from $F^{-1}\boldsymbol{\eta}^0 = \alpha^0 = \mathbf{0}$ that $\eta^0 = (0, 0, \ldots, 0)$. Hence $\ker(A) = \{(0, 0, \ldots, 0)\}$, and A is invertible.

For given liquidation constraints x^0, define $\eta := A^{-1}x^0$ and $\alpha^* := F^{-1}\boldsymbol{\eta}$. Then α^* is a Nash equilibrium in the class of deterministic strategies by Lemma 3.4: First, $\int_0^T \alpha^*(t)\,\mathrm{d}t = A\eta = x^0$; second, $F\alpha^* = \boldsymbol{\eta}$. By Lemma 3.2, it is the unique Nash equilibrium in the class of admissible strategies. □

3.2 Optimal execution for a single investor

Before studying the general case with an arbitrary number of investors, consider the case where only one strategic investor trades, i.e., $n = 0$. The investor's optimal strategy can be obtained in closed form for some decay kernels, and displays a number of desirable features—such as differentiability, nonnegativity and convexity—for many others.

Given that the investor pursues an admissible strategy α_0, the asset price evolves according to

$$S(t; \alpha_0) = \int_0^t G(t-s)\alpha_0(s)\,\mathrm{d}s, \qquad t \in [0, T]. \tag{3.6}$$

Notice that in the single-investor case, transaction costs γ_0 can also be interpreted as arising from (linear) *temporary price impact* as in Almgren and Chriss (2001). In this case, the asset price evolves according to $\tilde{S}(t; \alpha_0) = \frac{\gamma_0}{2}\alpha_0(t) + \int_0^t G(t-s)\alpha_0(s)\,\mathrm{d}s$. Compare also the remarks on transaction costs and costs from temporary price impact on p. 17. Under either interpretation, the investor's *costs of execution* are

$$J_0[\alpha_0] = \int_0^T \big(\frac{\gamma_0}{2}\alpha_0(t)^2 + \alpha_0(t)S(t; \alpha_0)\big)\,\mathrm{d}t = \int_0^T \alpha_0(t)\tilde{S}(t; \alpha_0)\,\mathrm{d}t.$$

For convenience, write α, γ and J instead of α_0, γ_0 and J_0 throughout this section. The operator $F\colon L^2[0, T] \to L^2[0, T]$ becomes

$$(F\alpha)(t) = \gamma\alpha(t) + \int_0^T G(|t-s|)\alpha(s)\,\mathrm{d}s. \tag{3.7}$$

Notice that, in contrast to the multi-investor case, F is self-adjoint. The optimality condition from Lemma 3.4 (ii) characterizing optimal strategies α^* reads

$$\gamma\alpha^*(t) + \int_0^T G(|t-s|)\alpha^*(s)\,\mathrm{d}s = \eta, \qquad t \in [0, T]. \tag{3.8}$$

Remark 3.7. *The problem of optimal execution for a single investor is equivalent to a problem of constrained norm minimization. To see this, denote by* $\langle \cdot, \cdot \rangle$ *the*

L^2-inner product on $[0,T]$ and by $\|\cdot\|$ its induced norm. Define for $\alpha, \beta \in L^2[0,T]$:

$$\langle \alpha, \beta \rangle_F \coloneqq \langle F\alpha, \beta \rangle = \int_0^T (F\alpha)(t)\, \beta(t)\, \mathrm{d}t.$$

Since F is linear and self-adjoint, similar arguments as in the proof of Theorem 3.6 show that $\langle \cdot, \cdot \rangle_F$ is a vector product, and that its induced norm $\|\cdot\|_F$ is equivalent to $\|\cdot\|$.

For $\eta \in \mathbb{R}$, denote by $\boldsymbol{\eta}$ the constant function $\boldsymbol{\eta}(t) = \eta$. Since $J[\alpha] = \|\alpha\|_F^2/2$, minimizing J over admissible strategies is equivalent to

$$\textit{Minimize } \|\alpha\|_F \textit{ over } \alpha \in L^2[0,T] \textit{ satisfying } \langle \alpha, \mathbf{1} \rangle = x^0. \tag{3.9}$$

This representation leads to an alternative proof of the fact that an admissible strategy α^* solves the problem of optimal execution if and only if it solves the Fredholm integral equation (3.8) *for some $\eta \in \mathbb{R}$.*

Define the linear subspace $B \coloneqq \{\beta \in L^2[0,T] \mid \langle \beta, \mathbf{1} \rangle = 0\}$. It is closed under $\|\cdot\|$. Indeed, if $\beta_1, \beta_2, \ldots \in B$ converge to $\beta \in L^2[0,T]$ under $\|\cdot\|$, then

$$\Big| \int_0^T \beta(t)\, \mathrm{d}t \Big| = \Big| \int_0^T \big(\beta(t) - \beta_n(t)\big)\, \mathrm{d}t \Big| \le T\, \|\beta - \beta_n\| \to 0 \qquad \textit{as } n \to \infty.$$

Suppose first that $\alpha^* \in L^2[0,T]$ is an optimal strategy and therefore solves (3.9)*. By equivalence of norms, B is a closed linear subspace under $\|\cdot\|_F$, and one may decompose $\alpha^* = \alpha^B + \alpha^\perp$, where $\alpha^B \in B$ and $\langle \alpha^\perp, \beta \rangle_F = 0$ for every $\beta \in B$. Hence*

$$\|\alpha^*\|_F^2 = \|\alpha^B\|_F^2 + \|\alpha^\perp\|_F^2 \ge \|\alpha^\perp\|_F^2.$$

But $\langle \alpha^\perp, \mathbf{1} \rangle = \langle \alpha^*, \mathbf{1} \rangle = x^0$. Since α^* is an optimal strategy, $\|\alpha^*\|_F^2 \le \|\alpha^\perp\|_F^2$. Conclude that $0 = \|\alpha^*\|_F^2 - \|\alpha^\perp\|_F^2 = \|\alpha^B\|_F^2$ and thus $\alpha^* = \alpha^\perp$. Consequently,

$$0 = \langle \alpha^\perp, \beta \rangle_F = \langle \alpha^*, \beta \rangle_F = \int_0^T (F\alpha^*)(t)\, \beta(t)\, \mathrm{d}t \qquad \textit{for every } \beta \in B.$$

By the fundamental lemma of the calculus of variations (Lemma 3.3), $F\alpha^*$ is constant almost everywhere on $[0,T]$.

Suppose on the other hand that α^* satisfies $\langle \alpha^*, \mathbf{1} \rangle = x^0$ and that there is an $\eta \in \mathbb{R}$ such that $(F\alpha^*)(t) = \eta$ for almost all $t \in [0,T]$. Every $\alpha \in L^2[0,T]$ with $\langle \alpha, \mathbf{1} \rangle = x^0$ can be written as $\alpha = \alpha^B + \alpha^*$ for some $\alpha^B \in B$, so α^* indeed solves (3.9) *and therefore the problem of optimal execution:*

$$\|\alpha\|_F^2 = \|\alpha^B\|_F^2 + 2\eta \,\langle \alpha^B, \mathbf{1} \rangle + \|\alpha^*\|_F^2 = \|\alpha^B\|_F^2 + \|\alpha^*\|_F^2 \ge \|\alpha^*\|_F^2.$$

A similar proof will show in Chapter 4 that (3.8) *still characterizes the optimal execution strategy if the decay kernel is no longer square-integrable but continuous and integrable, possibly with a weak singularity $G(0) = \lim_{t \to 0} G(t) = \infty$ (Lemma 4.3).*

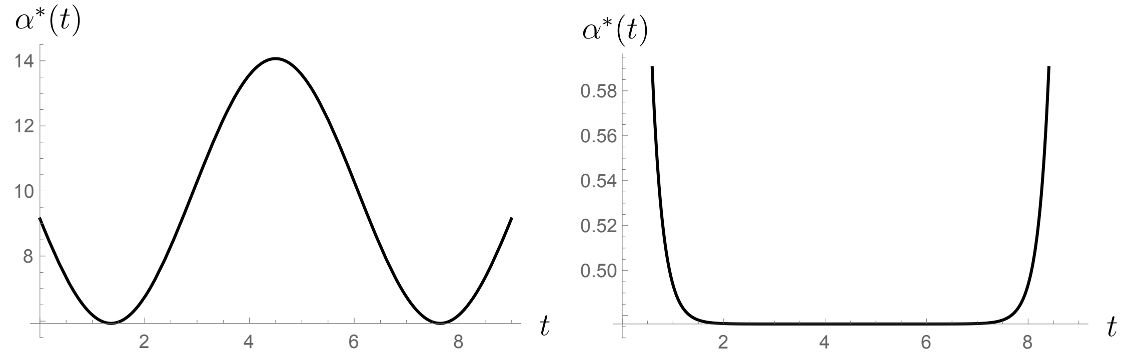

Figure 3.1: Solution $\alpha^*(t)$ of (3.8) for the trigonometric decay kernel $G(t) = \cos(t)$ (left) and the exponential decay kernel $G(t) = e^{-t}$ (right). Parameters: $T = 9$, $\gamma = 0.1$ and $\eta = 1$.

Examples and numerical simulation

The following explicit solutions of the Fredholm integral equation (3.8) are straightforward to verify. To obtain the respective optimal strategy α_0^* for a given liquidation constraint x^0, determine η via the equality $\int_0^T \alpha^*(t)\,\mathrm{d}t = x^0$. Let $\rho > 0$ and $\eta \in \mathbb{R}$.

- *Constant decay kernel:* If $G(t) = \rho$, then

$$\alpha^*(t) = \frac{\eta}{\gamma + \rho T}.$$

- *Trigonometric decay kernel:* If $G(t) = \cos(\rho t)$, then

$$\alpha^*(t) = \frac{\eta}{\gamma}\Big(1 - \frac{2\tan(\rho T/2)\big(\cos(\rho t) + \cos(\rho(T-t))\big)}{\rho(2\gamma + T) + \sin(\rho T)}\Big).$$

- *Linear decay kernel:* For $G(t) = 1 - \rho t$, define $a := 2\rho/\gamma$. Then

$$\alpha^*(t) = \frac{\eta\sqrt{a}(e^{\sqrt{a}t} + e^{\sqrt{a}(T-t)})}{e^{\sqrt{a}T}(\sqrt{a}\gamma + 2 - \rho T) + \sqrt{a}\gamma - 2 + \rho T}.$$

- *Exponential decay kernel:* For $G(t) = e^{-\rho t}$, define $a := \rho^2 + 2\rho/\gamma$. Then

$$\alpha^*(t) = \frac{\eta\rho^2}{\gamma a}\Big(1 + \frac{2(e^{\sqrt{a}t} + e^{\sqrt{a}(T-t)})}{\gamma(e^{\sqrt{a}T}(\rho + \sqrt{a}) + \rho - \sqrt{a})}\Big).$$

Figure 3.1 plots α^* for a trigonometric and an exponential decay kernel.
Now consider the *capped linear decay kernel* $G(t) = (1-t)^+$. Assume that $T = m$ is a natural number. For $i = 1, 2, \ldots, m$, define

$$\lambda_i := 2\big(1 - \cos\big(\frac{i\pi}{m+1}\big)\big) \qquad \text{and} \qquad b_i := \sqrt{\lambda_i/\gamma}.$$

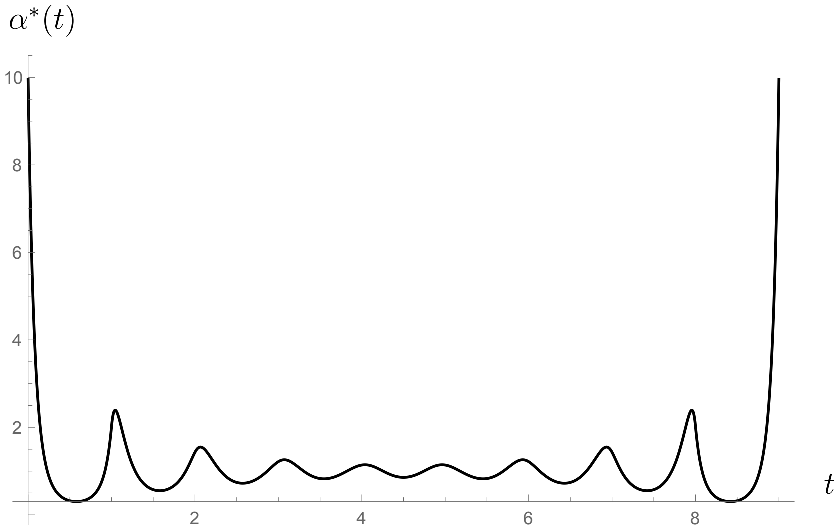

Figure 3.2: Solution $\alpha^*(t)$ of (3.8) for the capped linear decay kernel $G(t) = (1-t)^+$. Parameters: $T = 13$, $\gamma = 0.1$ and $\eta = 1$.

Denote by I the m-dimensional identity matrix. Define the m-dimensional square matrices

$$\begin{aligned} B &:= \operatorname{diag}(b_1, b_2, \dots, b_m), & E(t) &:= \operatorname{diag}\big(e^{b_1 t}, e^{b_2 t}, \dots, e^{b_m t}\big), \quad t \in [0, T], \\ \Sigma &:= \operatorname{diag}(1, -1, 1, \dots, \pm 1), & K &:= I + \big(\mathbb{1}_{\{j=m-i\}}\big)_{i,j=1,2,\dots,m}, \end{aligned}$$

where $\mathbb{1}$ denotes the indicator function, and

$$Q := \Big(\sin\big(\frac{ij\pi}{m+1}\big)\Big)_{i,j=1,2,\dots,m.}$$

Finally, define $a \in \mathbb{R}^m$ by

$$a := \Big(\gamma Q\big(E(1) + \Sigma\big) + KQ\big((E(1) - I)(\Sigma - I) + B(E(1) - \Sigma)\big)B^{-2}\Big)^{-1} \begin{pmatrix} \eta \\ \eta \\ \vdots \\ \eta \end{pmatrix}.$$

The key to calculating the solution $\alpha^*(t)$ for $t \in [0, m]$ is to consider each interval $[0, 1], [1, 2], \dots, [m-1, m]$ separately.

Proposition 3.8. *Suppose $T = m \in \mathbb{N}$ and $G(t) = (1-t)^+$ for all $t \in [0, m]$. Define the function $\psi = (\psi_1, \psi_2, \dots, \psi_m)\colon [0, 1] \to \mathbb{R}^m$ via*

$$\psi(\tau) = Q\big(E(\tau) + E(1-\tau)\Sigma\big)a.$$

Then the solution α^* of (3.8) satisfies $\alpha^*(\tau + i - 1) = \psi_i(\tau)$ for all $\tau \in [0, 1]$ and all $i = 1, 2, \dots, m$.

Proof. 1. Let α^* denote the unique solution of (3.8). It holds that

$$\begin{aligned}
&\int_0^m \big(1-|t-s|\big)^+\alpha^*(s)\,\mathrm{d}s \\
&= \begin{cases} \int_0^t (1-t+s)\alpha^*(s)\,\mathrm{d}s + \int_t^{t+1}(1+t-s)\alpha^*(s)\,\mathrm{d}s, & t\in[0,1], \\ \int_{t-1}^t (1-t+s)\alpha^*(s)\,\mathrm{d}s + \int_t^{t+1}(1+t-s)\alpha^*(s)\,\mathrm{d}s, & t\in[1,m-1], \\ \int_{t-1}^t (1-t+s)\alpha^*(s)\,\mathrm{d}s + \int_t^{m}(1+t-s)\alpha^*(s)\,\mathrm{d}s, & t\in[m-1,m]. \end{cases}
\end{aligned}$$

Define $\psi=(\psi_1,\psi_2,\dots,\psi_m)\colon[0,1]\to\mathbb{R}^m$ via $\psi_i(\tau):=\alpha^*(\tau+i-1)$ for $i=1,2,\dots,m$. Differentiating twice and replacing α^* with $\psi_1,\psi_2,\dots,\psi_m$ yields that, for $\tau\in[0,1]$,

$$\begin{aligned}
\gamma\psi_1''(\tau) &= 2\psi_1(\tau)-\psi_2(\tau), \\
\gamma\psi_i''(\tau) &= 2\psi_i(\tau)-\psi_{i-1}(\tau)-\psi_{i+1}(\tau), \qquad i=2,3,\dots,m-1, \\
\gamma\psi_m''(\tau) &= 2\psi_m(\tau)-\psi_{m-1}(\tau).
\end{aligned}$$

Hence ψ solves the following m-dimensional system of ordinary differential equations on $[0,1]$:

$$\psi''=\frac{1}{\gamma}\begin{pmatrix} 2 & -1 & \dots & 0 & 0 \\ -1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & \dots & -1 & 2 \end{pmatrix}\psi.$$

Denote the matrix in this equation by M. It is a tridiagonal Toeplitz matrix. Its eigenvalues are $\lambda_1,\lambda_2,\dots,\lambda_m$, as defined above, and the columns of Q contain corresponding eigenvectors. Eigendecomposition of M shows

$$\psi(\tau)=Q\big(E(\tau)z+E(1-\tau)\tilde{z}\big), \qquad \tau\in[0,1],$$

for some vectors $z,\tilde{z}\in\mathbb{R}^m$.

2. Define $k:=\lceil m/2\rceil$. Let $I_k, J_k, 0_k$ denote the k-dimensional identity matrix, reverse identity matrix and zero matrix, respectively. Let $\tau\in[0,1]$. It is shown in Lemma 3.9 that α^* is symmetric, i.e., $\alpha^*(t)=\alpha^*(T-t)$ for all $t\in[0,m]$. This implies that $\psi_i(\tau)=\psi_{m+1-i}(1-\tau)$ for all $i=1,2,\dots,m$. In terms of the matrices defined above, this reads

$$\big[I_k\ 0_k\big]Q\big(E(\tau)z+E(1-\tau)\tilde{z}\big)=\big[0_k\ J_k\big]Q\big(E(1-\tau)z+E(\tau)\tilde{z}\big). \tag{3.10}$$

For all $i\in\{1,2,\dots,k\}$ and $j\in\{1,2,\dots,m\}$, it holds that

$$\begin{aligned}
\sin\Big(\frac{(m+1-i)j\pi}{m+1}\Big) &= \sin(j\pi)\cos\Big(\frac{ij\pi}{m+1}\Big)-\cos(j\pi)\sin\Big(\frac{ij\pi}{m+1}\Big) \\
&= (-1)^{j+1}\sin\Big(\frac{ij\pi}{m+1}\Big).
\end{aligned}$$

Hence $\big[0_k\ J_k\big]Q = \big[I_k\ 0_k\big]Q\Sigma$. Notice that $\Sigma^{-1} = \Sigma$. Conclude that (3.10) holds for all $\tau \in [0,1]$ if and only if $\tilde{z} = \Sigma z$.
3. Let $t = i \in \{1, 2, \ldots, m-1\}$. Then (3.8) and symmetry of α^* imply that

$$\begin{aligned}
\eta &= \gamma\alpha^*(i) + \int_{i-1}^{i} (1 - i + s)\,\alpha^*(s)\,\mathrm{d}s + \int_{i}^{i+1} (1 + i - s)\,\alpha^*(s)\,\mathrm{d}s \\
&= \gamma\psi_i(1) + \int_{i-1}^{i} (1 - i + s)\,\psi_i(1 - i + s)\,\mathrm{d}s + \int_{i}^{i+1} (1 + i - s)\,\psi_{m-i}(1 + i - s)\,\mathrm{d}s \\
&= \gamma\psi_i(1) + \int_0^1 s\big(\psi_i(s) + \psi_{m-i}(s)\big)\,\mathrm{d}s.
\end{aligned}$$

Similar arguments yield $\eta = \gamma\psi_m(1) + \int_0^1 s\psi_m(s)\,\mathrm{d}s$.
A straightforward calculation shows

$$\int_0^1 s\psi(s)\,\mathrm{d}s = Q\big((E(1) - I)(\Sigma - I) + B(E(1) - \Sigma)\big)B^{-2}z.$$

In total, this leads to the m-dimensional system of equations

$$(\eta, \eta, \ldots, \eta) = \big(\gamma Q\big(E(1) + \Sigma\big) + KQ\big((E(1) - I)(\Sigma - I) + B(E(1) - \Sigma)\big)B^{-2}\big)z.$$

It remains to solve for z. □

Figure 3.2 plots α^* for the capped linear decay kernel $G(t) = (1-t)^+$. Proposition 3.8 also characterizes solutions for the more general decay kernels $G(t) = (1-\rho t)^+$, where $\rho > 0$. Indeed, assume that $m := \rho T$ is a natural number. Let $\beta^* \in L^2[0, m]$ solve

$$\rho\gamma\beta^*(t) + \int_0^m \big(1 - |t - s|\big)^+ \beta^*(s)\,\mathrm{d}s = \rho\eta, \qquad t \in [0, m].$$

This equation falls in the domain of Proposition 3.8, so β^* is characterized by a function $\psi\colon [0,1] \to \mathbb{R}^m$ as described above.
Define $\alpha^*(t) \in L^2[0, T]$ via $\alpha^*(t) := \beta^*(\rho t)$. Integration by substitution shows that for $t \in [0, T]$,

$$\begin{aligned}
&\rho\Big(\gamma\alpha^*(t) + \int_0^T \big(1 - \rho|t - s|\big)^+ \alpha^*(s)\,\mathrm{d}s\Big) \\
&= \rho\gamma\beta^*(\rho t) + \int_0^{\rho T} \big(1 - |\rho t - s|\big)^+ \beta^*(s)\,\mathrm{d}s \\
&= \rho\eta.
\end{aligned}$$

Hence α^* solves (3.8) for $G(t) = (1 - \rho t)^+$.

For many other decay kernels, explicit solutions of (3.8) are unknown. Examples include $G(t) = 1/(1 + t^2)$ and *Gaussian decay kernels* $G(t) = e^{-\rho t^2}$, where $\rho > 0$. But numerical solutions are readily available. The Nyström method (1930) with trapezoidal quadrature, as described in Example 11.4.5 in Atkinson and Han (2001),

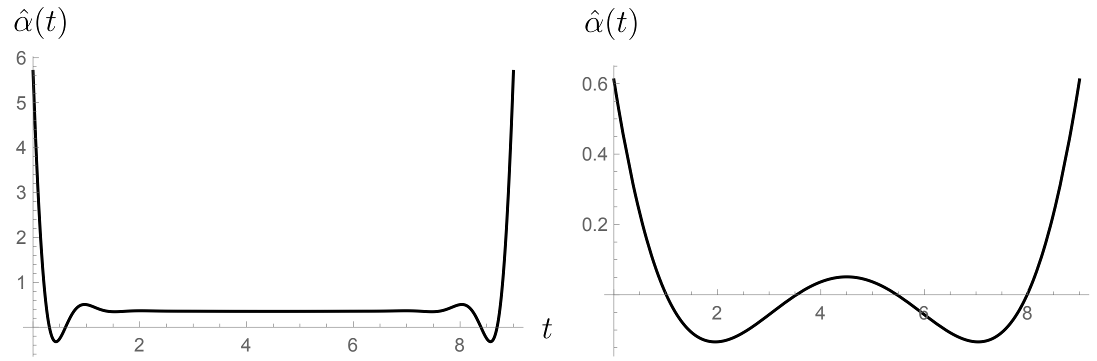

Figure 3.3: Numerical solution $\hat{\alpha}(t)$ of (3.8) for $G(t) = 1/(1+t^2)$ (left) and the Gaussian decay kernel $G(t) = 10e^{-0.01t^2}$ (right). Parameters: $m = 150$, $T = 9$, $\gamma = 0.1$ and $\eta = 1$.

is a simple approach for non-singular decay kernels. More sophisticated methods can be found in Chapter 11 of Aktinson and Han's book, in Atkinson (1997) and in Chapter 4 of Wazwaz (2011).

Pick $m \in \mathbb{N}$ and define equidistant time steps $t_i := iT/m$ for $i = 0, 1, \ldots, m$. Then solve the following linear system for $z = (z_0, z_1, \ldots, z_m)$:

$$\gamma z_i + \frac{T}{2n} \sum_{j=1}^{m} \big(G(|t_i - t_{j-1}|) z_{j-1} + G(|t_i - t_j|) z_j \big) = 1, \qquad i = 0, 1, \ldots, m.$$

Approximate the solution of (3.8) by

$$\hat{\alpha}(t) := \frac{\eta}{\gamma} \Big(1 - \frac{T}{2n} \sum_{j=1}^{m} \big(G(|t - t_{j-1}|) z_{j-1} + G(|t - t_j|) z_j \big) \Big), \qquad t \in [0, T].$$

Figure 3.3 shows some approximations obtained with this method.

To simulate the optimal strategy for a given liquidation constraint x^0, proceed as follows:

(i) Use a numerical method (e.g., the Nyström method) to find an approximate solution $\hat{\beta} \in L^2[0, T]$ of the Fredholm integral equation (3.8) for $\eta = 1$.

(ii) Approximate the optimal strategy α^* for a given liquidation constraint x^0 by

$$\hat{\alpha} := \frac{x^0}{\int_0^T \hat{\beta}(t) \, \mathrm{d}t} \hat{\beta}.$$

This construction ensures that $\hat{\alpha}$ is an admissible strategy. It therefore can be used to find an upper bound $J[\hat{\alpha}]$ on the minimal costs of execution $J[\alpha^*]$.

Qualitative features of solutions

It seems worthwhile to investigate whether some general qualitative statements about α^* can be derived without finding an explicit solution. What conditions on G are sufficient to ensure that α^* is continuous, differentiable, nonnegative or convex?
All qualitative features should be understood as holding almost everywhere on $[0, T]$.
A first simple example of a qualitative statement about α^* is its symmetry around the midpoint $T/2$.

Lemma 3.9. *Suppose α^* solves (3.8). Then $\alpha^*(t) = \alpha^*(T - t)$ for all $t \in [0, T]$.*

Proof. Define $\overline{\alpha}(t) := \alpha^*(T-t)$, $t \in [0, T]$. Integration by substitution shows that $\overline{\alpha}$ solves (3.8) as well. Indeed, for all $t \in [0, T]$,

$$
\begin{aligned}
&\gamma\overline{\alpha}(t) + \int_0^T G(|t-s|)\overline{\alpha}(s)\,\mathrm{d}s \\
&= \gamma\alpha^*(T-t) + \int_0^T G(|(T-t)-(T-s)|)\alpha^*(T-s)\,\mathrm{d}s \\
&= \gamma\alpha^*(T-t) + \int_0^T G(|(T-t)-s|)\alpha^*(s)\,\mathrm{d}s \\
&= \eta.
\end{aligned}
$$

By Lemma 3.2, the solution of (3.8) is unique. Hence $\alpha^* = \overline{\alpha}$. □

The following proposition shows that α^* inherits continuity and differentiability from the decay kernel. Furthermore, if α^* is differentiable, then each of its derivatives solves a Fredholm integral equation of the second kind with nonconstant free term. The regularity of solutions of Fredholm integral equations of the second kind is well-studied, and similar results have been derived for general free terms and under weaker assumptions on the decay kernel (see for instance Kahane, 1965, Pitkäranta, 1980 and Vainikko, 2006).

Proposition 3.10. *Suppose α^* solves (3.8). If G is continuous on $[0, T]$, then α^* is continuous. If G is m-times continuously differentiable on $[0, T]$, then α^* is $(m+1)$-times continuously differentiable. In this case, there are $f_1, f_2, \ldots, f_{m+1} \in L^2[0, T]$ such that, for $k = 1, 2, \ldots, m+1$, the function f_k is continuous and*

$$
\gamma\alpha^{*,(k)}(t) + (-1)^k \int_0^T G(|t-s|)\alpha^{*,(k)}(s)\,\mathrm{d}s = f_k(t), \qquad t \in [0, T], \tag{3.11}
$$

where $\alpha^{*,(k)}$ denotes the k-th derivative of α^*. In particular, α^* is smooth if G is smooth.

Proof. It is clear from (3.8) that α^* is continuous if G is continuous. Now suppose G is m-times continuously differentiable for some $m \in \mathbb{N}$. Show by induction that for $k = 1, 2, \dots, m+1$, the derivative $\alpha^{*,(k)}$ exists and (3.11) holds for

$$f_k(t) \coloneqq \sum_{i=1}^{k} \big((-1)^k \alpha^{*,(i-1)}(T)\, G^{(k-i)}(T-t) + (-1)^{i+1} \alpha^{*,(i-1)}(0)\, G^{(k-i)}(t)\big),$$

for $t \in [0,T]$. Indeed, integration by parts yields that, for every $t \in [0,T]$,

$$\begin{aligned} \gamma \alpha^{*,(1)}(t) &= -\Big(\int_0^t G^{(1)}(t-s) \alpha^*(s)\, \mathrm{d}s + \int_t^T \big(- G^{(1)}(s-t)\big) \alpha^*(s)\, \mathrm{d}s \Big) \\ &= \int_0^T G(|t-s|) \alpha^{*,(1)}(s)\, \mathrm{d}s - \alpha^*(T)\, G(T-t) + \alpha^*(0)\, G(t). \end{aligned}$$

The first equality shows that $\alpha^{*,(1)}$ exists. For $k = 1, 2, \dots, m$, induction shows that

$$\begin{aligned} &\gamma \alpha^{*,(k+1)}(t) \\ &= (-1)^{k+1} \Big(\int_0^t G^{(1)}(t-s) \alpha^{*,(k)}(s)\, \mathrm{d}s + \int_t^T \big(- G^{(1)}(s-t)\big) \alpha^{*,(k)}(s)\, \mathrm{d}s \Big) \\ &\quad + \sum_{i=1}^{k} \big((-1)^{k+1} \alpha^{*,(i-1)}(T)\, G^{(k+1-i)}(T-t) + (-1)^{i+1} \alpha^{*,(i-1)}(0)\, G^{(k+1-i)}(t)\big) \\ &= (-1)^{k+2} \int_0^T G(|t-s|) \alpha^{*,(k+1)}(s)\, \mathrm{d}s \\ &\quad + \sum_{i=1}^{k+1} \big((-1)^{k+1} \alpha^{*,(i-1)}(T)\, G^{(k+1-i)}(T-t) + (-1)^{i+1} \alpha^{*,(i-1)}(0)\, G^{(k+1-i)}(t)\big) \end{aligned}$$

for every $t \in [0,T]$. □

Assuming G to be of positive type ensures that profits from simple price manipulation strategies are impossible, i.e., $J[\alpha] \geq 0$ for every $\alpha \in L^2[0,T]$ (as discussed on p. 56). But Alfonsi et al. (2012) argue that a more subtle type of price manipulation may still be optimal. The price increase caused by a buy order improves the execution price of all subsequent sell orders, and vice versa. Therefore, it may be optimal for an investor who must trade a net amount $x^0 \neq 0$ to engage in *transaction-triggered price manipulation* by submitting both buy and sell orders, instead of only trading in one direction. These strategies "look similar to usual price manipulation strategies but occur only when triggered by a given transaction" (Alfonsi et al., 2012, p. 512).

Optimal strategies that only trade in one direction are also desirable from a modeling perspective. They remain optimal if the asset price is not specified exogenously by (3.6), but derived from a model of a (block-shaped) limit order book (Gatheral et al., 2012, Remark 2.4).

Notice that if transaction-triggered price manipulation is optimal for some liquidation constraint $x^0 \neq 0$, then it is optimal for all liquidation constraints $x^0 \neq 0$, because optimal strategies are linear in x^0.

Since x^0 and η have the same sign, transaction-triggered price manipulation is optimal if and only if $\eta\alpha^*(t) < 0$ for some $t \in [0, T]$. The following theorem provides two different sufficient conditions that ensure the nonnegativity of $\eta\alpha^*$.
The first condition guarantees that the decay term $(1/\gamma)\int_0^T G(|t-s|)\alpha^*(s)\,\mathrm{d}s$ is negligible. In this case, nonnegativity of $\eta\alpha^*(t)$ is a simple consequence of the fact that $\alpha^*(t) \approx \eta$. The second condition is more subtle and holds for all $\gamma > 0$. It is also sufficient for absence of transaction-triggered price manipulation in the corresponding discrete time model without transaction costs (Alfonsi et al., 2012, Theorem 1).
The decay kernel G is said to be *positive definite* (in the sense of Bochner, 1932) if

$$\sum_{i,j=1}^{m} G(|t_i - t_j|) z_i z_j \geq 0$$

for all $m \in \mathbb{N}$, $t_1, t_2, \dots, t_m \in [0, \infty)$ and $z_1, z_2, \dots, z_m \in \mathbb{R}$. Clearly, every positive definite decay kernel is of positive type. If G is continuous, the two concepts are equivalent (Mercer, 1909, §9 and §10). If G is convex, nonincreasing and nonnegative on $[0, \infty)$, then it is positive definite (see Young, 1913, or Gatheral et al., 2012, for a more recent reference).

Theorem 3.11. *Suppose α^* solves* (3.8) *and one of the following two conditions is satisfied:*

(i) G is positive definite and $\gamma \geq G(0)T$.

(ii) G is convex, nonincreasing, nonnegative and continuous.

Then $\eta\alpha^*(t) \geq 0$ for all $t \in [0, T]$.

Proof. (i) First, prove the following auxiliary result:

Lemma 3.12. *Suppose G is positive definite. Let $t_0 \in [0, T]$. Then*

$$\int_0^T \int_0^T H(t,s)\alpha(t)\alpha(s)\,\mathrm{d}s\,\mathrm{d}t \geq 0 \qquad \textit{for all } \alpha \in L^2[0,T],$$

where

$$H(t,s) := G(|t-s|) - 2G(|t-t_0|) + G(0), \qquad s,t \in [0,T].$$

Proof. It is enough to show $\sum_{i,j=1}^n H(t_i,t_j)z_i z_j \geq 0$ for all $n \in \mathbb{N}$, $z_1, z_2, \dots, z_n \in \mathbb{R}$ and $t_1, t_2, \dots, t_n \in [0, T]$.
Define $z_0 := -\sum_{i=1}^n z_i$. Since G is positive definite,

$$\begin{aligned} &0 \\ \leq &\sum_{i,j=0}^{n} G(|t_i - t_j|) z_i z_j \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i,j=1}^{n} G(|t_i - t_j|) z_i z_j - 2\Big(\sum_{i=1}^{n} G(|t_i - t_0|) z_i\Big)\Big(\sum_{i=1}^{n} z_i\Big) + G(0)\Big(\sum_{i=1}^{n} z_i\Big)^2 \\
&= \sum_{i,j=1}^{n} H(t_i, t_j) z_i z_j,
\end{aligned}
$$

as desired. $\square$

It is enough to consider the case $\eta = 1$. Let α^* solve (3.8). Define the linear operator F as in (3.7) and let $x^0 := \int_0^T \alpha^*(t)\,\mathrm{d}t$. For all $t_0 \in [0,T]$, obtain both

$$
\int_0^T G(|t-t_0|)\alpha^*(t)\,\mathrm{d}t = (F\alpha^*)(t_0) - \gamma\alpha^*(t_0) = 1 - \gamma\alpha^*(t_0)
$$

and

$$
\begin{aligned}
\int_0^T \int_0^T G(|t-s|)\alpha^*(t)\alpha^*(s)\,\mathrm{d}s\,\mathrm{d}t &= \langle F\alpha^*, \alpha^*\rangle - \gamma \int_0^T \alpha^*(t)^2\,\mathrm{d}t \\
&= x^0 - \gamma \int_0^T \alpha^*(t)^2\,\mathrm{d}t.
\end{aligned}
$$

Notice that

$$
\int_0^T \alpha^*(t)^2\,\mathrm{d}t \ge \min_{\alpha\colon \int_0^T \alpha(t)\,\mathrm{d}t = x^0} \int_0^T \alpha(t)^2\,\mathrm{d}t = \int_0^T \Big(\frac{x^0}{T}\Big)^2\,\mathrm{d}t = \frac{(x^0)^2}{T}.
$$

From Lemma 3.12, obtain

$$
\begin{aligned}
0 \ge &-\int_0^T \int_0^T G(|t-s|)\alpha^*(t)\alpha^*(s)\,\mathrm{d}s\,\mathrm{d}t \\
&+ 2\Big(\int_0^T \alpha^*(t)\,\mathrm{d}t\Big)\Big(\int_0^T G(|t-t_0|)\alpha^*(t)\,\mathrm{d}t\Big) - G(0)\bigg(\int_0^T \alpha^*(t)\,\mathrm{d}t\bigg)^2 \\
= &\,\gamma \int_0^T \alpha^*(t)^2\,\mathrm{d}t - x^0 + 2x^0\big(1 - \gamma\alpha^*(t_0)\big) - G(0)(x^0)^2 \\
\ge &\,\frac{\gamma (x^0)^2}{T} + x^0 - 2\gamma x^0 \alpha^*(t_0) - G(0)(x^0)^2,
\end{aligned}
$$

or equivalently,

$$
\alpha^*(t_0) \ge \frac{1}{2\gamma}\Big(x^0\big(\frac{\gamma}{T} - G(0)\big) + 1\Big).
$$

Notice that $0 < J[\alpha^*] = \int_0^T (F\alpha^*)(t)\alpha^*(t)\,\mathrm{d}t = \eta x^0 = x^0$. Conclude that $\alpha^*(t_0) \ge 0$.
(ii) This will be shown in more generality in Theorem 4.4. $\square$

Numerical simulations suggest that $\eta\alpha^*$ is convex with a minimum at $T/2$ for many, but not all, convex, nonincreasing and nonnegative decay kernels G. In the context of optimal execution, convexity of α^* mirrors the "empirically observed U-shape of the daily distribution of market liquidity. That is, if [α^* is convex with a minimum at $T/2$] and the liquidation horizon is one trading day, as it is often the case, then the optimal liquidation strategy [α^*] involves fast trading toward the beginning and end of the trading day when liquidity is high and slower trading when liquidity is low" (Schied and Strehle, 2017, p. 2).

It will be shown that complete monotonicity of G is sufficient to ensure convexity of α^*. The decay kernel G is said to be *completely monotone* if it is smooth and

$$(-1)^m G^{(m)}(t) \geq 0 \qquad \text{for all } t \in (0, \infty) \text{ and all } m \in \mathbb{N}.$$

Clearly, every completely monotone decay kernel is convex, nonincreasing, and nonnegative on $(0, T)$. In particular, if G is completely monotone on $(0, \infty)$, then it is positive definite and thus of positive type.

In fact, complete monotonicity of G implies that $\eta\alpha^*$ is *symmetrically totally monotone*, i.e., it is analytic on $(0, T)$ and there are nonnegative coefficients $(z_{2k})_{k\in\mathbb{N}}$ such that its power series development in $T/2$ is of the form

$$\eta\alpha^*(t) = \sum_{k=0}^{\infty} z_{2k}(t - T/2)^{2k}, \qquad t \in (0, T).$$

The following theorem is proven in more generality as Theorem 4.6 in the next section.

Theorem 3.13. *Suppose α^* solves (3.8). If G is completely monotone, then $\eta\alpha^*$ is symmetrically totally monotone. In particular, it is analytic on $(0, T)$ and convex.*

3.3 Exponential price impact

Now consider again the general case with $n+1$ strategic investors. Unsurprisingly, it is more complicated. Equilibrium strategies do not display the regularity and strong dependence on the decay kernel G that was observed in the previous section.

To make the problem tractable, assume that transient price impact decays at an exponential rate, i.e., $G(t) = e^{-\rho t}$, as in Obizhaeva and Wang (2013) and in Chapter 2. The parameter $\rho > 0$ determines the size and persistence of price impact. A small ρ implies large impact and slow recovery (see Figure 3.4). The limit $\rho = 0$ corresponds to permanent price impact as in Almgren and Chriss (2001).

In searching for a closed-form representation of the Nash equilibrium, it will turn out that the $(n + 2)$-dimensional function

$$\psi \coloneqq (\alpha_0, \alpha_1, \ldots, \alpha_n, S)$$

is a more natural object of study than $\alpha \coloneqq (\alpha_0, \alpha_1, \ldots, \alpha_n)$. Clearly, ψ determines α and vice versa. With a slight abuse of terminology, $\psi^* = (\alpha_0^*, \alpha_1^*, \ldots, \alpha_n^*, S^*)$ will

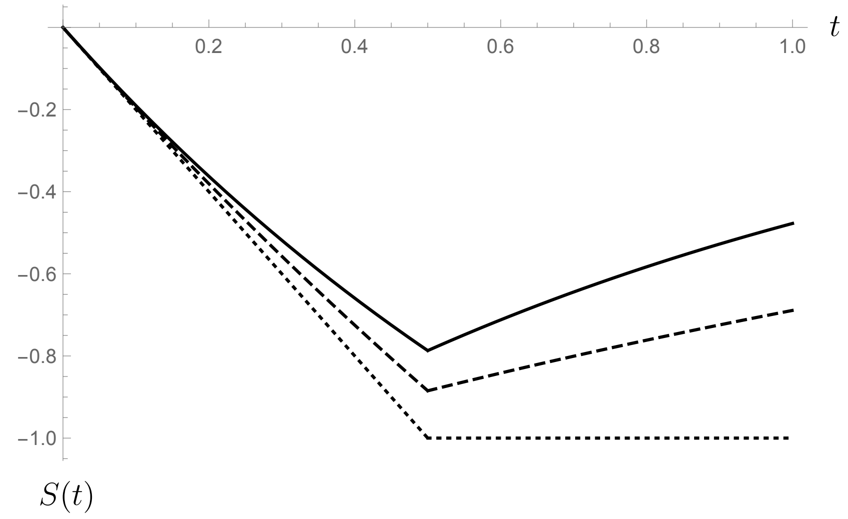

Figure 3.4: Illustration of transient price impact with exponential decay $G(t) = e^{-\rho t}$. Asset price $S(t)$ for $\rho = 1$ (solid line), $\rho = 0.5$ (dashed line) and for permanent price impact $\rho = 0$ (dotted line). Here, a single strategic investor trades at a constant rate $\alpha_0(t) = -2$ while $t \leq 1/2$ and $\alpha_0(t) = 0$ while $t > T/2$. Parameters: $n = 0$, $T = 1$ and $x_0^0 = -1$. Notice that α_0 is admissible but not optimal.

also be called a Nash equilibrium if $\alpha^* := (\alpha_0^*, \alpha_1^*, \ldots, \alpha_n^*)$ is a Nash equilibrium and $S^* = S(\cdot\,; \alpha^*)$.

Notice that G is of positive type. Assume again that $\gamma_i > 0$ for all $i = 0, 1, \ldots, n$. Let $\alpha^* = (\alpha_0^*, \alpha_1^*, \ldots, \alpha_n^*)$ be the unique Nash equilibrium (compare Lemma 3.2 and Theorem 3.6). It is deterministic and continuous. Let $\eta = (\eta_0, \eta_1, \ldots, \eta_n)$ denote the corresponding vector for which $F\alpha^* = \boldsymbol{\eta}$ (compare Lemma 3.4). Finally, let $S^* := S(\,\cdot\,; \alpha^*)$.

The optimality conditions in Lemma 3.4 may be written as a system of integral equations:

$$\gamma_i \alpha_i^*(t) + S^*(t) + \int_t^T e^{\rho(t-s)} \alpha_i^*(s)\,\mathrm{d}s = \eta_i, \qquad t \in [0, T], \tag{3.12}$$

for $i = 0, 1, \ldots, n$.

If all investors had homogeneous transaction costs $\gamma_0 = \gamma_1 = \cdots = \gamma_n$, one could sum (3.12) over i to obtain a two-dimensional system of differential equations characterizing $\sum_{i=0}^n \alpha_i^*$ and S^*. Once this system were solved, (3.12) would reduce to $n+1$ identical one-dimensional ordinary differential equations. The model in Schied and Zhang (2015) allows for this approach. But if transaction costs are heterogeneous, all functions $\alpha_0^*, \alpha_1^*, \ldots, \alpha_n^*$ and S^* must be computed simultaneously.

Let $i = 0, 1, \ldots, n$. It is clear from (3.12) that α_i^* is differentiable in t. Differentiating and plugging in from (3.12) yields the ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\alpha_i^* = \rho\alpha_i^* - \frac{1}{\gamma_i}\sum_{j \neq i} \alpha_j^* + \frac{2\rho}{\gamma_i} S^* - \frac{\rho\eta_i}{\gamma_i}. \tag{3.13}$$

Furthermore, (3.1) shows that S^* satisfies the ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} S^* = \sum_{i=0}^{n} \alpha_i^* - \rho S^*. \tag{3.14}$$

Combine (3.13) and (3.14) to conclude that the Nash equilibrium

$$\psi^* := (\alpha_0^*, \alpha_1^*, \ldots, \alpha_n^*, S^*)$$

solves a system of differential equations of the form $\frac{\mathrm{d}}{\mathrm{d}t}\psi^* = M\psi^* + m$, where M is a square matrix and m is a vector.
Let e^{Mt} denote the matrix exponential of Mt. If M is invertible, then ψ^* must be of the form $\psi^*(t) = e^{Mt}z - M^{-1}m$, $t \in [0, T]$, for some $z \in \mathbb{R}^{n+2}$. Notice that the investors' liquidation constraints translate into unusual boundary conditions for ψ^*: They apply to $\int_0^T \alpha_i^*(t)\,\mathrm{d}t$, not α_i^*. This complicates the calculation of z.
Building on these considerations, the next theorem derives a closed-form representation of ψ^* (and therefore α^*). Define the $(n+2)$-dimensional square matrices

$$M := \begin{pmatrix} \rho & -\frac{1}{\gamma_0} & \cdots & -\frac{1}{\gamma_0} & \frac{2\rho}{\gamma_0} \\ -\frac{1}{\gamma_1} & \rho & \cdots & -\frac{1}{\gamma_1} & \frac{2\rho}{\gamma_1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{1}{\gamma_n} & -\frac{1}{\gamma_n} & \cdots & \rho & \frac{2\rho}{\gamma_n} \\ 1 & 1 & \cdots & 1 & -\rho \end{pmatrix}, \qquad N_1 := \begin{pmatrix} \rho\gamma_0 & 0 & \cdots & 0 & \rho \\ 0 & \rho\gamma_1 & \cdots & 0 & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \rho\gamma_n & \rho \\ \gamma_0 & \gamma_1 & \cdots & \gamma_n & n+1 \end{pmatrix}.$$

Define further the $(n+1) \times (n+2)$-dimensional matrix

$$W := \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

and the $(n+2)$-dimensional column vector $v := (0, 0, \ldots, 0, 1)$. For $m \in \mathbb{N} \setminus \{0\}$, denote by I_m the m-dimensional identity matrix. Define the $(n+2)$-dimensional square block matrix

$$N_2 := \begin{bmatrix} W\big((M^{-1} + N_1 T)e^{MT} - M^{-1}\big) \\ v^\top (I_{n+1} + N_1 e^{MT}) \end{bmatrix}.$$

Theorem 3.14. *The matrix N_2 is invertible and it holds that*

$$\psi^*(t) = (e^{Mt} + N_1 e^{MT}) N_2^{-1} \tilde{x}^0, \qquad t \in [0, T], \tag{3.15}$$

where $\tilde{x}^0 := (x_0^0, x_1^0, \ldots, x_n^0, 0)$.

Proof. A function $\psi = (\alpha_0, \alpha_1, \ldots, \alpha_n, S) \in L^2([0, T]; \mathbb{R}^{n+2})$ shall be called *regular* if for each $i = 0, 1, \ldots, n$, the function α_i is a deterministic admissible strategy for investor i, and $S = S(\cdot\,; \alpha)$ is the asset price corresponding to $\alpha \coloneqq (\alpha_0, \alpha_1, \ldots, \alpha_n)$. Define the $(n+1) \times (n+2)$-dimensional matrices

$$U \coloneqq \begin{pmatrix} \gamma_0 & 0 & \ldots & 0 & 1 \\ 0 & \gamma_1 & \ldots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \gamma_n & 1 \end{pmatrix} \qquad \text{and} \qquad V \coloneqq \begin{pmatrix} \frac{\rho}{\gamma_0} & 0 & \ldots & 0 & 0 \\ 0 & \frac{\rho}{\gamma_1} & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \frac{\rho}{\gamma_n} & 0 \end{pmatrix}.$$

The proof is in four steps.

1. *ψ is a Nash equilibrium if and only if it is regular, continuously differentiable and solves the system of differential equations*

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi = M\psi - V^\top y, \tag{3.16}$$

where $y = U\psi(T)$.

Necessity: Let ψ be a Nash equilibrium. It is clear from (3.12) and (3.14) that ψ is continuously differentiable. Let $t = T$ in (3.12) to obtain $y = U\psi(T)$. From (3.13) and (3.14), deduce (3.16).

Sufficiency: Let $\psi = (\alpha_0, \alpha_1, \ldots, \alpha_n, S)$ be regular and suppose it solves (3.16) with $y = U\psi(T)$. Let $\alpha \coloneqq (\alpha_0, \alpha_1, \ldots, \alpha_n)$. Fix $i = 0, 1, \ldots, n$. Define

$$f_i(t) \coloneqq \int_t^T e^{\rho(t-s)}\alpha_i(s)\,\mathrm{d}s, \qquad t \in [0, T].$$

Clearly, $f_i' = \rho f_i - \alpha_i$. Let $g_i \coloneqq \eta_i - \gamma_i\alpha_i - S$. Conclude $g_i' = \rho g_i - \alpha_i$ with (3.16). Hence $f_i' - g_i' = \rho(f_i - g_i)$. Applying the boundary condition $y = U\psi(T)$ shows that $f_i(T) - g_i(T) = 0$. It follows that $f_i = g_i$ and

$$(F\alpha)_i = \gamma_i\alpha_i + S + f_i = \gamma_i\alpha_i + S + g_i = \eta_i.$$

This is true for all $i = 0, 1, \ldots, n$. Hence ψ is a Nash equilibrium by Lemma 3.4.

2. *M is invertible.*

Define the $(n+2)$-dimensional column vectors

$$v_1 \coloneqq \Big(\rho + \frac{1}{\gamma_0}, \rho + \frac{1}{\gamma_1}, \ldots, \rho + \frac{1}{\gamma_n}, -\frac{1}{2}\Big), \qquad v_2 \coloneqq \Big(-\frac{1}{\gamma_0}, -\frac{1}{\gamma_1}, \ldots, -\frac{1}{\gamma_n}, 1\Big),$$
$$v_3 \coloneqq (1, 1, \ldots, 1, -2\rho), \qquad u \coloneqq (1, 1, \ldots, 1).$$

Then $M = (\operatorname{diag}(v_1) + v_2 u^\top)\operatorname{diag}(v_3)$ and by the matrix determinant lemma,

$$\det M = -\rho\Big(1 + \sum_{i=0}^{n} \frac{1}{\rho\gamma_i + 1}\Big)\prod_{i=0}^{n}\Big(\rho + \frac{1}{\gamma_i}\Big) \neq 0.$$

3. *ψ is a Nash equilibrium if and only if it is regular and there is a $z \in \mathbb{R}^{n+2}$ such that*

$$\psi(t) = (e^{Mt} + N_1 e^{MT})z, \qquad t \in [0,T]. \tag{3.17}$$

The general solution of (3.16) is $e^{Mt}z + M^{-1}V^\top y$, $t \in [0,T]$, for $z \in \mathbb{R}^{n+2}$. By Step 1, the condition $y = U(e^{MT}z + M^{-1}V^\top y)$ must be satisfied. The matrix $M - V^\top U$ is nonsingular, which can be verified by checking that $(M - V^\top U)^{-1} = (N_1 - vv^\top)/\rho$. By the Woodbury matrix identity,

$$(I_{n+1} - UM^{-1}V^\top)^{-1} = I_{n+1} + U(M - V^\top U)^{-1}V^\top.$$

Hence

$$y = (I_{n+1} + U(M - V^\top U)^{-1}V^\top)Ue^{MT}z.$$

It holds that $V^\top U(I_{n+2} + N_1) = MN_1$, or equivalently, $(M - V^\top U)^{-1}V^\top U = N_1$. With this, obtain

$$M^{-1}V^\top y = M^{-1}V^\top U(I_{n+2} + N_1)e^{MT}z = N_1 e^{MT}z.$$

4. Once ψ is defined by (3.17), integrating over $[0,T]$ shows that ψ is regular if and only if $N_2 z = \tilde{x}^0$. It remains to show that N_2 is invertible. Although this follows from Theorem 3.6, a separate proof is given here. Consider the case where all investors must trade zero net amounts, i.e., $x^0 = (0,0,\dots,0)$. It is easy to check that in this case, $\psi^0 := \mathbf{0}$ is a Nash equilibrium. By Lemma 3.2 (ii), this is the only Nash equilibrium. According to Step 3, there exists a $z^0 \in \mathbb{R}^{n+2}$ such that $\psi^0(t) = (e^{Mt} + N_1 e^{MT})z^0$ for $t \in [0,T]$. This shows that $(0,0,\dots,0) = \frac{\mathrm{d}}{\mathrm{d}t}\psi^0(t) = Me^{Mt}z^0$ for all $t \in [0,T]$. The matrix Me^{Mt} is invertible, hence $z^0 = (0,0,\dots,0)$. It follows that the equation $N_2 z = (0,0,\dots,0)$ has only the trivial solution $z^0 = (0,0,\dots,0)$, showing that N_2 is invertible. $\square$

The Nash equilibrium (3.15) can be approximated numerically. The next section does so and studies how the presence of opportunistic investors affects optimal strategies.

3.4 Order anticipation strategies

In practice, optimal execution of large orders is accomplished with the help of execution algorithms. These algorithms, including the popular VWAP (volume weighted average price, see for instance Cartea and Jaimungal, 2015), are typically based on the observation that price impact depends on the relative size of an order: Price impact is smaller when markets are busy. An execution algorithm might exploit this fact by trading every thirty seconds over the course of one trading day, placing large positions when market volume is high and small positions when market volume is low; while ensuring that the liquidation constraint is satisfied by the end of the day. The major weakness of such an algorithm is its predictability. Opportunistic investors with access to high-resolution data on financial markets (e.g., the entire limit order book) can detect the algorithm and reverse-engineer it to predict future trades. Once this is accomplished, they can pursue an *order anticipation strategy:* Trade in the same direction as the algorithm, but a little earlier; then wait until the execution algorithm has traded and clear inventory directly afterwards. With this simple strategy—also known as front-running—the execution algorithm's price impact becomes a predictable source of profit.

Order anticipation strategies require sophisticated detection algorithms (Hirschey, 2016) and a quick alternation of buy and sell orders. Hence they are typically associated with high frequency traders, for instance by the Securities and Exchange Commission (2010). Notice however that order anticipation strategies do not require the breathtaking speed necessary for "true" high frequency strategies such as stale order sniping or non-designated market making (MacKenzie, 2011).

That order anticipation strategies have been described as aggressive (Benos and Sagade, 2012), predatory (Brunnermeier and Pedersen, 2005) and "algo-sniffing" (MacKenzie, 2011) suggests that the Securities and Exchange Commission (2010) is not alone in suspecting that they "may present serious problems in today's market structure" (p. 3609). Indeed, Tong (2015) reports that "one standard deviation increase in the intensity of [high frequency trading] activities increases institutional execution shortfall costs by a third" (p. 4). Brunnermeier and Pedersen (2005) even suggest a direct connection between order anticipation strategies and financial breakdowns: Front-running amplifies the price drop caused by a large sell order, an effect known as *price overshooting.* This might trigger further sell orders (e.g., from pending stop-loss orders), which are again subject to front-running, causing further price overshooting and, ultimately, a complete market crash.

Even with high-quality data, empirical studies cannot perfectly identify order anticipation strategies in the market. This is why it can be helpful to study them in a theoretical model. This section analyzes the influence that order anticipation strategies have on a liquidating investor's costs of execution and on the asset price evolution in the model derived in the previous section. Special consideration is given to Brunnermeier and Pedersen's (2005) claim that order anticipation strategies cause price overshooting.

Let $n \geq 1$. Assume that investor 0 executes a net sell order $x_0^0 < 0$, while all other investors $i = 1, 2, \ldots, n$ trade zero net amounts $x_i^0 = 0$. The case $x_0^0 > 0$ is perfectly symmetric. Investors $i = 1, 2, \ldots, n$ will only trade if they can generate profits (that is, negative costs) from the price impact generated by investor 0. In this sense, they are *opportunistic investors.* Investor 0 will be referred to as the *liquidating investor.*

Assume that all opportunistic investors have identical levels of transaction costs, i.e., $\gamma_1 = \gamma_2 = \cdots = \gamma_n$. It follows from Lemma 3.2 that in equilibrium, all opportunistic investors pursue the same strategy α_1^*. Hence $\overline{\alpha} \coloneqq \sum_{i=1}^n \alpha_i^*$ equals $n\alpha_1^*$.

For the subsequent analysis, it is most illustrative to study the remaining net amounts

$$X_0^*(t) = x_0^0 - \int_0^t \alpha_0^*(s)\,\mathrm{d}s \qquad \text{and} \qquad \overline{X}(t) \coloneqq -n\int_0^t \alpha_1^*(s)\,\mathrm{d}s, \qquad t \in [0, T].$$

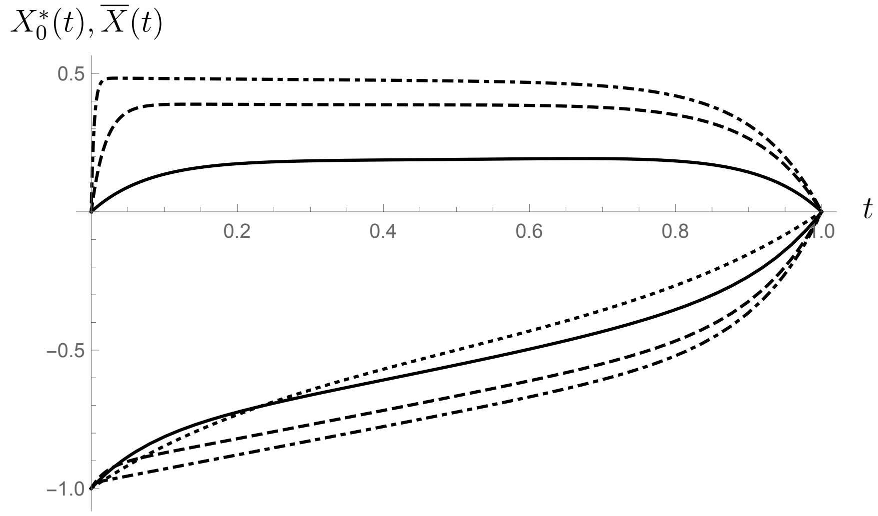

Figure 3.5: Remaining net amounts $X_0^*(t)$ and $\overline{X}(t)$ for $n = 0$ (dotted line), $n = 1$ (solid line), $n = 5$ (dashed line) and $n = 25$ (dot-dashed line). Parameters: $T = 1$, $\rho = 0.95$, $\gamma_0 = \gamma_1 = 0.1$, $x_0^0 = -1$ and $x_1^0 = 0$.

Figure 3.5 is representative of equilibrium strategies in general. Opportunistic investors engage in *front-running*: They build up short positions in the beginning and buy back for the rest of the trading period. The liquidating investor sells throughout the trading period, such that opportunistic investors generate a profit from selling high and buying low. A larger number of opportunistic investors implies more front-running shortly after $t = 0$. The liquidating investor reacts accordingly to avoid selling in a falling market and shifts more trading activity to the end.
This pattern holds for many parameter combinations, but it is not universal: If γ_1 is small and ρ large, opportunistic investors may start selling again shortly after $T/2$, such that $\overline{\alpha}$ changes sign three times.

Costs of execution

Front-running amplifies the liquidating investor's price impact and increases his total costs significantly (see Figure 3.6).
If opportunistic investors have low transaction costs, one or two of them suffice to fully realize the profit potential of order anticipation. Further increasing n results in competition among opportunistic investors, reducing their total profit and also reducing the liquidating investor's total costs. This is different if transaction costs are high. Transaction costs restrict the degree to which investors can benefit from opportunistic trading. High transaction costs leave unrealized profit potential for further opportunistic investors. Consequently, the liquidating investor's total costs increase in n. Notice that while total profits of opportunistic investors generally increase in n, these profits are divided (equally) among an increasing number of investors; each opportunistic investor earns less if n increases.
It is interesting that the liquidating investor's costs rise significantly if γ_1 decreases

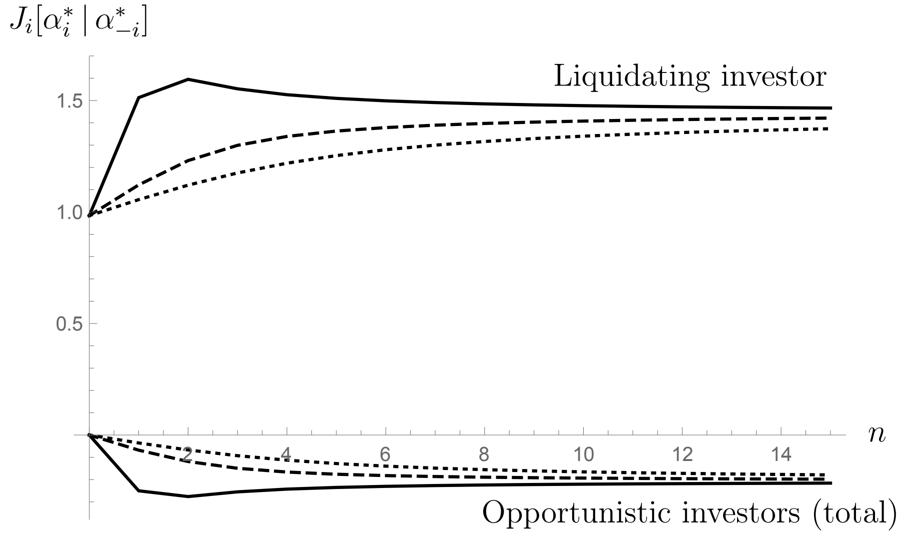

Figure 3.6: Total costs of the liquidating investor and sum of total (negative) costs of the opportunistic investors in dependence of n; for $\gamma_1 = 0.1$ (solid line), $\gamma_1 = 0.5$ (dashed line) and $\gamma_1 = 1$ (dotted line). Parameters: $T = 1$, $\rho = 0.1$, $\gamma_0 = 1$, $x_0^0 = -1$ and $x_1^0 = 0$.

(at least for small n), while simulations show that his optimal strategy hardly changes. This suggests that the liquidating investor can do little to avoid exploitation from order anticipation strategies.

Asset price and price overshooting

Figure 3.7 shows that in the absence of opportunistic investors, the asset price S^* decreases steadily over time: it exhibits a persistent drift. This changes drastically once opportunistic investors enter the picture, especially if there are many of them. Opportunistic investors build up short positions very quickly, causing a sudden price drop right after $t = 0$. The asset price remains almost constant afterwards. This seems to support claims about opportunistic investors improving price discovery: The price drop caused by the order x_0^0 occurs earlier and more quickly (see Sections 6.2 and 6.3 in Benos and Sagade, 2012, for a discussion). But inferences about price discovery are outside the scope of this model because all liquidation constraints are known to all investors.

Consider now the maximum deviation of the asset price,

$$\Sigma := \sup_{t \in [0,T]} |S(t) - S(0)|.$$

Brunnermeier and Pedersen (2005) claim that opportunistic investors cause price overshooting, i.e., Σ is larger for $n \geq 1$ than for $n = 0$. They argue that this may lead to a domino effect: The price drop caused by the liquidating investor and amplified by the opportunistic investors triggers additional sell orders (for instance from

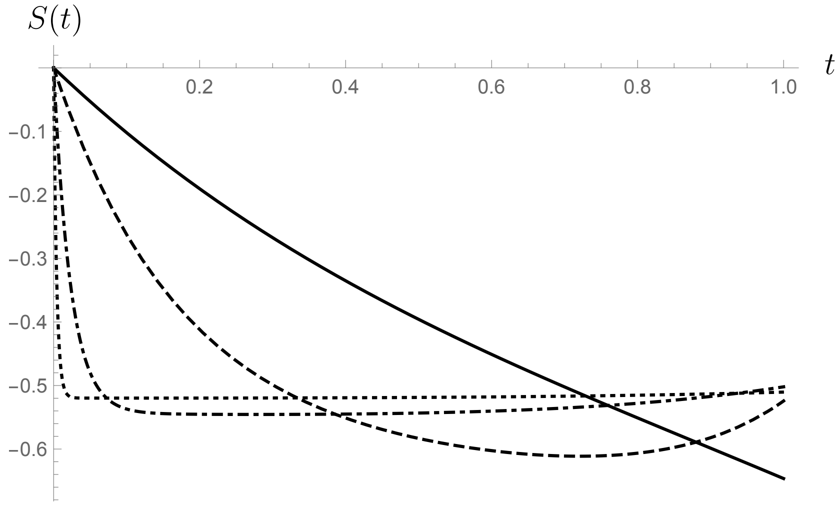

Figure 3.7: Asset price $S(t)$ for $n = 0$ (solid line), $n = 1$ (dashed line), $n = 5$ (dot-dashed line) and $n = 25$ (dotted line). Parameters: $T = 1$, $\rho = 0.95$, $\gamma_0 = 1$, $\gamma_1 = 0.1$, $x_0^0 = -1$ and $x_1^0 = 0$.

pending stop-loss orders). This causes an even more extreme price drop, triggering further sell orders, etc.

The model of price impact in which Brunnermeier and Pedersen observe price overshooting only features temporary and permanent price impact. Both impose few constraints on the opportunistic traders, and generate little feedback from the opportunistic traders to the liquidating trader. Consequently, opportunistic traders trade aggressively and scale their strategies to an exogenously given maximum size. Brunnermeier and Pedersen arrive at a grim picture in which "predators" (opportunistic traders) exploit "distressed traders" (liquidating investors) and may even cause a "panic" (the domino effect described above). Price overshooting is also observed by Oehmke (2014), again in a model with temporary and permanent price impact only.

Figure 3.8 shows that in the present model, price overshooting is the exception, not the rule: In general, Σ decreases if n increases. This is most evident for markets with a "short memory", i.e., for large values of ρ. A possible explanation is that the price overshooting observed by Brunnermeier and Pedersen is a consequence of permanent (or long-lived transient) price impact, rather than an inherent consequence of opportunistic trading.

Another possible explanation is that quadratic transaction costs prevent price overshooting. Quadratic costs imply that a (statistical) arbitrage strategy cannot be scaled indefinitely without becoming unprofitable. As n increases, competition among opportunistic investors increases, and transaction costs increasingly work against them. Figure 3.8 shows, however, that there is no obvious relationship between Σ and the level of transaction costs γ_1.

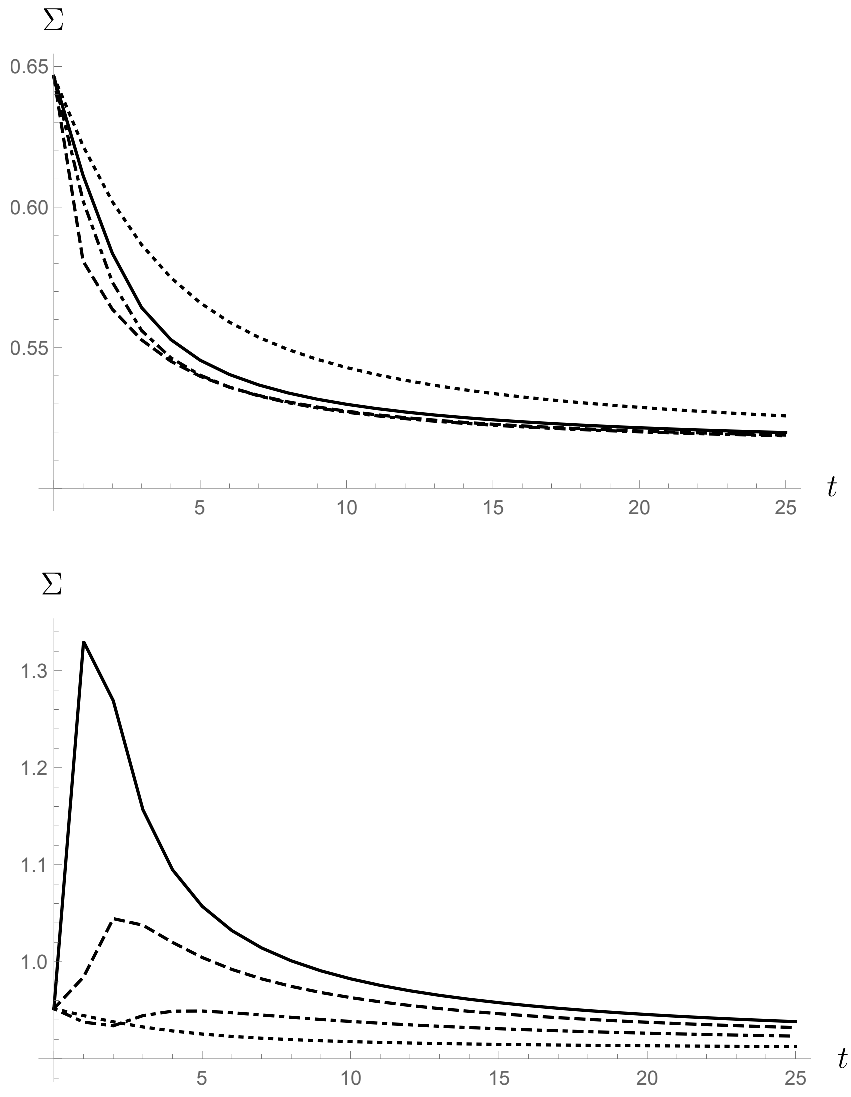

Figure 3.8: Maximum deviation Σ of the asset price for $\rho = 0.95$ (top) and $\rho = 0.1$ (bottom) and for $\gamma_1 = 0.1$ (solid line), $\gamma_1 = 0.25$ (dashed line), $\gamma_1 = 0.5$ (dot-dashed line) and $\gamma_1 = 1$ (dotted line). Parameters: $T = 1$, $\gamma_0 = 1$, $x_0^0 = -1$ and $x_1^0 = 0$.

3.5 Different time frames

Opportunistic investors do not necessarily have the same time horizon as the liquidating investor. Admati and Pfleiderer (1991) point out that the liquidating investor may *preannounce* his liquidation constraint: "By informing potential traders who can take the other side of the preannounced orders and by allowing the market to prepare to absorb these orders, preannouncement facilitates the match between the demand and supply of liquidity in the market" (p. 444). This practice is known as sunshine trading. It can be implemented by demanding that the liquidating investor only trade after some time $T_0 > 0$, as in Brunnermeier and Pedersen (2005).

In their analysis of sunshine trading, Schöneborn and Schied (2009) argue that opportunistic investors may also have additional time to unwind their position after the liquidating investor has fully executed his order. This can be implemented in the model by demanding that the liquidating investor only trade until some time $T_1 < T$. Consequently, divide $[0, T]$ into three periods: The *acquisition period* $[0, T_0]$, the *main period* $[T_0, T_1]$ and the *unwinding period* $[T_1, T]$. Suppose for now that T_0 and T_1 are fixed. The liquidating investor is only allowed to trade during the main period. Opportunistic investors begin and end with a flat inventory $X_1(0) = X_1(T) = 0$. They use the acquisition period to build up a position $X_1(T_0)$, then trade alongside the liquidating investor during the main period. At the end of the main period, they hold a position $X_1(T_1)$, which they liquidate in the unwinding period. Notice that in equilibrium, all opportunistic investors still behave identically.

Given $X_1(T_0)$, the acquisition period is described by the model in Section 3.1, where investors $i = 1, 2, \dots, n$ acquire identical amounts $x_1^0 = x_2^0 = \cdots = x_n^0 = X_1(T_0)$ over the time horizon $[0, T_0]$. Theorem 3.14 yields the equilibrium strategies.

During the main and the liquidation period, the situation is more complicated. The model by Schöneborn and Schied features linear temporary and permanent price impact. This has the advantage that price impact generated in earlier periods has no influence on equilibrium strategies in subsequent periods. With transient price impact, trades from earlier periods cause a (deterministic) price drift in subsequent periods. During the main period, the asset price becomes

$$S(t) = e^{-\rho(t-T_0)} S(T_0) + \int_{T_0}^{t} e^{-\rho(t-s)} \sum_{i=0}^{n} \alpha_i(s)\, \mathrm{d}s, \qquad t \in [T_0, T_1].$$

In the same way, price impact from the main period generates a price drift during the liquidation period. Theorem 3.14 must be generalized by replacing S with

$$\tilde{S}(t) := e^{-\rho(t-\tau_0)} s + S(t), \qquad t \in [\tau_0, \tau_1],$$

for $(\tau_0, \tau_1) \in \{(T_0, T_1), (T_1, T)\}$ and $s \in \mathbb{R}$. Repeating the arguments from Section 3.1, one sees that the Nash equilibrium ψ^* still satisfies a system of differential equations of the form $\frac{\mathrm{d}}{\mathrm{d}t}\psi^* = M\psi^* + m$, but now $m = m(t)$ is not constant.

Once this system is solved, one may calculate the optimal strategies for the liq-

uidating investor during the main period, and for the opportunistic investors during the main and liquidation periods, in dependence of $X_1(T_0)$ and $X_1(T_1)$. This yields the total (negative) costs for opportunistic investors over the entire time horizon $[0, T]$ in dependence of $X_1(T_0)$ and $X_1(T_1)$. It remains to minimize these costs over $(X_1(T_0), X_1(T_1)) \in \mathbb{R}^2$.
But go one step further. In the current setting, a liquidating investor engaging in sunshine trading may not only announce his liquidation constraint x_0^0, but also his time horizon $[T_0, T_1]$. Schöneborn and Schied show that a shorter trading horizon $T_1 < T$ can be beneficial to the liquidating investor in certain market conditions. Although there may be an exogenous upper bound on T_1, it is reasonable to assume that the liquidating investor can voluntarily commit to a shorter trading horizon. The liquidating investor also has some control over T_0 because he can choose the time of announcement $t = 0$. Hence T_0 and T_1 should not be viewed as exogenous. One should rather perform a final optimization over $(T_0, T_1) \in \mathbb{R}^2$, this time minimizing the liquidating investor's total costs during the main period.
This extension promises interesting results, with opportunistic investors possibly engaging in liquidity provision instead of front-running, as in Schöneborn and Schied (2009).

Chapter 4

Completely monotone decay kernels

Consider again the problem of optimal execution with a single risk-neutral investor. The investor must trade a fixed net amount x^0 over the time horizon $[0, T]$. Price impact is transient and modeled via a *decay kernel* G of positive type. Distinguish two cases:

(i) There are no transaction costs. This corresponds to the setting in Section 2.4 for $\tilde{\gamma} = 0$. In this case, an *(admissible) strategy* is a right-continuous function $X\colon [0-, T] \to \mathbb{R}$ of bounded total variation which satisfies the liquidation constraint $X(0-) = x^0$ and $X(T) = 0$. As in Section 2.4, $X(t)$ describes the remaining net order at time t. Lemma 2.3 in Gatheral et al. (2012) shows that the *costs of execution* of a given admissible strategy X are

$$J^0[X] := \frac{1}{2} \int_{[0,T]} \int_{[0,T]} G(|t-s|)\, \mathrm{d}X(s)\, \mathrm{d}X(t)$$

plus a constant that is the same for all admissible strategies.

(ii) There are positive transaction costs of size $\gamma > 0$ on the instantaneous rate of trading. This corresponds to the setting in Section 3.2. In this case, an *(admissible) strategy* is a function $\alpha \in L^2[0, T]$ satisfying the liquidation constraint $\int_0^T \alpha(t)\, \mathrm{d}t = x^0$. As shown in the previous chapter, the *costs of execution* of a given admissible strategy α are

$$J^\gamma[\alpha] := \frac{1}{2} \int_0^T \big(\gamma \alpha(t)^2 + \int_0^T G(|t-s|) \alpha(t) \alpha(s)\, \mathrm{d}s \big)\, \mathrm{d}t$$

plus a constant that is the same for all admissible strategies.

The case $\gamma = 0$ must be treated separately because optimal strategies in this case are usually not absolutely continuous (Gatheral et al., 2012, Theorem 2.23). Hence it is necessary to specify the remaining net amount directly. Positive transaction costs $\gamma > 0$ imply that jumps in the remaining net amount cannot be optimal. If α is an admissible strategy for the case $\gamma > 0$, then $X(t) := x^0 - \int_0^t \alpha(s)\, \mathrm{d}s$ is an admissible strategy for the case $\gamma = 0$.

If G is square-integrable on $[0, T]$ and $G(0)$ is finite, a unique minimizer of J^γ in the class of admissible strategies exists for every $\gamma > 0$ (compare Lemma 3.2 and Theorem 3.6). But these assumptions on G are too rigid: Bouchaud et al. (2004) and Almgren et al. (2005) report that empirical observations of transient price impact are described well by *power-law decay kernels* $G(t) = t^{-\rho}$ for $0 < \rho < 1$.
Notice also that square-integrability is not sufficient to ensure existence of a minimizer of J^0. A counterexample is the Gaussian decay kernel $G(t) = e^{-t^2}$ (Gatheral et al., 2012, Example 2.16).
In this section, I study the minimization of J^γ for $\gamma \geq 0$ and decay kernels G that are continuous and integrable, but may have a weak singularity $\lim_{t\to 0} G(t) = \infty$. The main results will additionally assume that G is convex and nonincreasing. Under these assumptions, existence and uniqueness of a minimizer X^* of J^0 was shown by Gatheral et al. (2012), who also prove that X^* is nonincreasing. In Section 4.1, I extend this result to the case $\gamma > 0$.
In Section 4.2, I additionally assume that G is *completely monotone* on $(0, T)$, i.e., it is smooth on $(0, T)$ and $(-1)^n G^{(n)} \geq 0$ for all $n \in \mathbb{N}$. Completely monotone kernels lead to the study of *symmetrically totally monotone* functions. These are functions f that are analytic and have a power series development in $T/2$ of the form

$$f(t) = \sum_{k=0}^{\infty} z_{2k}(t - T/2)^{2k}$$

for nonnegative coefficients $(z_{2k})_{k\in\mathbb{N}}$. I show that if G is completely monotone, then for every $\gamma > 0$, the minimizer α^* of J^γ is symmetrically totally monotone, and the minimizer X^* of J^0 is differentiable on $(0, T)$ and $-\frac{\mathrm{d}}{\mathrm{d}t}X^*(t)$ is symmetrically totally monotone.
This chapter is a revised version of Schied and Strehle (2017). I am grateful to Alexander Schied for the fruitful collaboration, and two anonymous referees for helpful remarks.

4.1 Existence and nonnegativity

Assume that the decay kernel $G\colon (0, \infty) \to [0, \infty)$ is nonconstant, continuous and satisfies $\int_0^\tau G(t)\,\mathrm{d}t < \infty$ for all $\tau > 0$. Assume further that G is of *positive type*, i.e.,

$$\int_0^\tau \int_0^\tau G(|t-s|)\alpha(t)\alpha(s)\,\mathrm{d}s\,\mathrm{d}t \geq 0$$

for every $\alpha \in L^2[0, \tau]$ for which the double integral is well-defined and every $\tau > 0$. Define $G(0) := \lim_{t\to 0} G(t)$. Notice that $G(0) \in (0, \infty]$.
Assume $x^0 = 1$ without loss of generality. For $\gamma = 0$, let $\mathcal{A}^0$ denote the set of admissible strategies X for which $\int_0^T \int_0^T G(|t-s|)\,\mathrm{d}|X|(s)\,\mathrm{d}|X|(t)$ is finite. Here, $|X|$ denotes the total variation process of X. Notice that the liquidation constraint ensures $X(0-) = 1$ and $X(T) = 0$ for every $X \in \mathcal{A}^0$.

For $\gamma > 0$, let $\mathcal{A}^1$ denote the set of admissible strategies α for which the double integral $\int_0^T \int_0^T G(|t-s|)\alpha(t)\alpha(s)\,\mathrm{d}s\,\mathrm{d}t$ is well-defined and finite.
Gatheral et al. (2012) state sufficient conditions for the existence and uniqueness of a minimizer of J^0.

Theorem 4.1. (Theorem 2.24 in Gatheral et al., 2012). *If G is convex and nonincreasing, then J^0 admits a unique minimizer X^* in $\mathcal{A}^0$. In this case, X^* is nonincreasing.*

Notice that $-X^*$ is a probability distribution function on $[0,T]$. The goal of this section is to prove a parallel result for the case $\gamma > 0$.
Many arguments that follow will be simplified by the alternative representations of J^γ derived in the following lemma. For $X \in \mathcal{A}^0$, let $\hat{X}$ denote the Fourier-Stieltjes transform of X, i.e., $\hat{X}(t) = \int e^{ist}\,\mathrm{d}X(s)$. For $\alpha \in \mathcal{A}^1$, let $\hat{\alpha}$ denote the Fourier transform of α, i.e., $\hat{\alpha}(t) = \int e^{ist}\alpha(s)\,\mathrm{d}s$ (following the convention in Gatheral et al., 2012).

Lemma 4.2. *If G is convex and nonincreasing, then there is a positive Radon measure μ on $(0,\infty)$ such that*

$$J^0[X] = \frac{1}{2}\int |\hat{X}(t)|^2\mu(\mathrm{d}t) \qquad \textit{for every } X \in \mathcal{A}^0, \tag{4.1}$$

and at the same time, for every $\gamma > 0$,

$$J^\gamma[\alpha] = \frac{\gamma}{2}\int |\hat{\alpha}(t)|^2\,\mathrm{d}t + \frac{1}{2}\int |\hat{\alpha}(t)|^2\mu(\mathrm{d}t) \qquad \textit{for every } \alpha \in \mathcal{A}^1. \tag{4.2}$$

Proof. Define $G(\infty-) := \lim_{t\to\infty} G(t) < \infty$. By Lemma 4.1 in Gatheral et al. (2012), there is a positive Radon measure ν on $(0,\infty)$ such that $\int_{(0,\infty)} t \wedge t^2\,\nu(\mathrm{d}t) < \infty$ and

$$G(t) = G(\infty-) + \int_{(0,\infty)} (s-t)^+\nu(\mathrm{d}s), \qquad t \in (0,\infty).$$

Define $f\colon (0,\infty) \to \mathbb{R}$ via

$$f(t) := \frac{1}{\pi}\int_{(0,\infty)} \frac{1-\cos(st)}{t^2}\nu(\mathrm{d}s) \tag{4.3}$$

and a measure μ on $(0,\infty)$ via $\mu(\mathrm{d}t) := G(\infty-)\delta_0(\mathrm{d}t) + f(t)\,\mathrm{d}t$. Then μ is a positive Radon measure, and $G(|\cdot|)$ is the Fourier transform of μ, i.e. $G(|t|) = \int e^{ist}\mu(\mathrm{d}s)$ for every $t \in \mathbb{R}$ (Gatheral et al., 2012, Lemma 4.2).
Let $T > 0$. Proposition 4.5 in Gatheral et al. (2012) shows (4.1). This, in combination with the Plancherel theorem, shows (4.2). □

Notice that the representation of $J^\gamma[\alpha]$ in (4.2) extends to all $\alpha \in L^2[0,T]$ for which $\int_0^T \int_0^T G(|t-s|)\alpha(t)\alpha(s)\,\mathrm{d}s\,\mathrm{d}t$ is well-defined and finite. Denote the set of all

these functions by $L^2_G[0,T]$. Conclude with Lemma 4.2 and the Minkowski inequality that $L^2_G[0,T]$ is a vector space.
The following result shows that the characterization of minimizers of J^γ for $\gamma > 0$ via a Fredholm integral equation (as in Lemma 3.4) remains valid under the given assumptions, even if G is weakly singular and fails to be square-integrable.

Lemma 4.3. *Suppose $\gamma > 0$. The function $\alpha^* \in \mathcal{A}^1$ minimizes J^γ if and only if there is an $\eta \in \mathbb{R}$ such that*

$$\gamma\alpha^*(t) + \int_0^T G(|t-s|)\alpha^*(s)\,\mathrm{d}s = \eta, \qquad t \in [0,T]. \tag{4.4}$$

In this case, η is strictly positive and equals $2J^\gamma[\alpha^*]$.

Proof. Let $\gamma > 0$ and define $L^2_G[0,T]$ as above. For $\alpha, \beta \in L^2_G[0,T]$, define the symmetric bilinear form

$$J^\gamma[\alpha,\beta] := \frac{1}{2}\big(J^\gamma[\alpha+\beta] - J^\gamma[\alpha] - J^\gamma[\beta]\big).$$

Suppose $\alpha^* \in \mathcal{A}^1$ minimizes $J^\gamma[\,\cdot\,]$. Choose $y \in \mathbb{R}$ and a nonzero $\beta \in L^2_G[0,T]$ satisfying $\int_0^T \beta(t)\,\mathrm{d}t = 0$. Then $\alpha^* + y\beta \in \mathcal{A}^1$. A straightforward calculation shows that

$$J^\gamma[\alpha^* + y\beta] = J^\gamma[\alpha^*] + 2yJ^\gamma[\alpha^*,\beta] + y^2 J^\gamma[\beta].$$

Since α^* is optimal, the right-hand side must be minimized at $y = 0$. Clearly,

$$J^\gamma[\beta] \geq \frac{\gamma}{2}\int_0^T \beta(t)^2\,\mathrm{d}t > 0.$$

Hence $J^\gamma[\alpha^*,\beta] = 0$. Conclude that $\gamma\alpha^*(t)\int_0^T G(|t-s|)\alpha^*(s)\,\mathrm{d}s$ is orthogonal to every $\beta \in L^2_G[0,T]$ satisfying $\int_0^T \beta(t)\,\mathrm{d}t = 0$. This implies (4.4).
Now suppose on the other hand that α^* satisfies (4.4). Every $\alpha \in \mathcal{A}^1$ can be written as $\alpha = \alpha^* + y\beta$ for some $y \in \mathbb{R}$ and some $\beta \in L^2_G[0,T]$ satisfying $\int_0^T \beta(t)\,\mathrm{d}t = 0$. By Fubini's theorem,

$$\begin{aligned} J^\gamma[\alpha^* + \beta] &= J^\gamma[\alpha^*] + J^\gamma[\beta^*] + 2\int_0^T \beta(t)\big(\gamma\alpha^*(t) + \int_0^T G(|t-s|)\alpha^*(s)\,\mathrm{d}s\big)\,\mathrm{d}t \\ &= J^\gamma[\alpha^*] + J^\gamma[\beta^*] + 2\eta\int_0^T \beta(t)\,\mathrm{d}t. \end{aligned}$$

Conclude that $J^\gamma[\alpha^*,\beta] = 0$ and thus $J^\gamma[\alpha] = J^\gamma[\alpha^*] + y^2 J^\gamma[\beta] \geq J^\gamma[\alpha^*]$.
Finally, it is clear that $J[\alpha^*] = 2\eta$ if α^* satisfies (4.4). □

The following theorem is the main result of this section. It extends Theorem 4.1 to the case $\gamma > 0$. The proof is based on a similar result on the monotonicity of trading strategies in discrete time that is due to Alfonsi et al. (2012).

Theorem 4.4. *If G is convex and nonincreasing, then J^γ admits a unique minimizer α^* in $\mathcal{A}^1$ for every $\gamma > 0$. In this case, α^* is nonnegative.*

Proof. Let $\gamma > 0$. Lemma 3.2 (i) remains valid under the given assumptions on G, so J^γ has at most one minimizer in $\mathcal{A}^1$.
Consider first the case $G(0) := \lim_{t\to 0} G(t) < \infty$. Then $G(|\cdot|)$ is bounded and continuous on $\mathbb{R}$.
For $n \in \mathbb{N}$, define $t_k := k2^{-n}T$ for $k = 0, 1, \ldots, 2^n$. Let $\mathcal{A}^1_n$ denote the set of functions $\alpha \in L^2[0,T]$ which are constant on the intervals $[t_k, t_{k+1})$ and satisfy the liquidation constraint $\int_0^T \alpha(t)\,\mathrm{d}t = 1$. Any $\alpha \in \mathcal{A}^1_n$ is of the form

$$\alpha = \sum_{k=0}^{2^n-1} a_k \mathbb{1}_{[t_k,t_{k+1})} \tag{4.5}$$

for constants $a_0, a_1, \ldots, a_{2^n-1}$ that sum to $2^n/T$. Since each $\alpha \in \mathcal{A}^1_n$ is bounded, conclude that $\mathcal{A}^1_n \subseteq \mathcal{A}^1$. Now prove the following auxiliary result.

Lemma 4.5. *Let $n \in \mathbb{N}$ and define $G_n\colon [0,\infty) \to [0,\infty)$ via*

$$\begin{aligned} G_n(0) &:= \gamma 2^{-(n+1)}T + 2^{-2n+1}T^2 \int_0^1 G(2^{-n}Ts)\,(1-s)\,\mathrm{d}s, \\ G_n(t) &:= 2^{-2n}T^2 \int_{-1}^1 G(t + 2^{-n}Ts)\,(1-|s|)\,\mathrm{d}s, \qquad t \in [2^{-n}T, \infty), \end{aligned}$$

and $G_n(t)$ linearly interpolated between $G_n(0)$ and $G_n(2^{-n}T)$ for $t \in (0, 2^{-n}T)$. Let $\alpha \in \mathcal{A}^1_n$ be of the form (4.5). Then

$$J^\gamma[\alpha] = \sum_{i,j=0}^{2^n-1} G_n(|t_i - t_j|)a_i a_j.$$

Proof. Calculate

$$\frac{\gamma}{2}\int_0^T \alpha(t)^2\,\mathrm{d}t = \frac{\gamma}{2}\sum_{k=0}^{2^n-1} a_k^2(t_{k+1}-t_k) = \gamma 2^{-(n+1)}\,T\sum_{k=0}^{2^n-1} a_k^2$$

and

$$\frac{1}{2}\int_0^T\int_0^T G(|t-s|)\alpha(t)\alpha(s)\,\mathrm{d}s\,\mathrm{d}t = \frac{1}{2}\sum_{i,j=0}^{2^n-1} a_i a_j \int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}} G(|t-s|)\,\mathrm{d}s\,\mathrm{d}t.$$

Notice that by Fubini's theorem, for every integrable function f,

$$\begin{aligned} \int_0^1\int_0^1 f(x-y)\,\mathrm{d}y\,\mathrm{d}x &= \int_0^1\int_{x-1}^x f(y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^1\int_{x-1}^0 f(y)\,\mathrm{d}y\,\mathrm{d}x + \int_0^1\int_0^x f(y)\,\mathrm{d}y\,\mathrm{d}x \end{aligned}$$

$$
\begin{aligned}
&= \int_{-1}^{0} f(y) \int_{0}^{1+y} \mathrm{d}x \, \mathrm{d}y + \int_{0}^{1} f(y) \int_{y}^{1} \mathrm{d}x \, \mathrm{d}y \\
&= \int_{-1}^{1} f(y)(1-|y|) \, \mathrm{d}y.
\end{aligned}
$$

Hence for $i < j$, it holds that

$$
\begin{aligned}
\frac{1}{2} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} G(|t-s|) \, \mathrm{d}s \, \mathrm{d}t &= \frac{1}{2} \int_{0}^{2^{-n}T} \int_{0}^{2^{-n}T} G(t_j - t_i + s - t) \, \mathrm{d}s \, \mathrm{d}t \\
&= 2^{-2n+1}T^2 \int_0^1 \int_0^1 G\big(t_j - t_i + 2^{-n}T(s-t)\big) \, \mathrm{d}s \, \mathrm{d}t \\
&= G_n(t_j - t_i).
\end{aligned}
$$

The same argument shows $\int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} G(|t-s|) \, \mathrm{d}s \, \mathrm{d}t = G_n(t_i - t_j)$ for $i > j$ and

$$
\frac{1}{2} \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_{i+1}} G(|t-s|) \, \mathrm{d}s \, \mathrm{d}t = 2^{-2n}T^2 \int_0^1 G(2^{-n}Ts)\,(1-s) \, \mathrm{d}s.
$$

This proves the claim. Notice that all values $G_n(t)$ for $t \notin \{t_0, t_1, \ldots, t_{2^n-1}\}$ can in fact be chosen arbitrarily. □

Let G_n be as in Lemma 4.5.
On $[0, 2^{-n}T]$, the function G_n is linear. On $[2^{-n}T, \infty)$, it is a mixture of the convex, nonincreasing and nonnegative functions $t \mapsto G(t+2^{-n}Ts)$ for $s \in [-1,1]$: hence G_n has the same properties there. Conclude that G_n is convex if and only if the left-hand derivative $G'_{n,-}(2^{-n}T)$ of G_n in $2^{-n}T$ is smaller than or equal to the right-hand derivative $G'_{n,+}(2^{-n}T)$.
Let G'_+ denote the right-hand derivative of G. Recall that $G(0)$ is finite by assumption. The value $G'_+(0) \in (-\infty, 0)$ provides a lower bound on G'_+, hence

$$
G'_{n,+}(2^{-n}T) \geq 2^{-2n}T^2 \int_{-1}^{1} G'_+(0)(1-|s|) \, \mathrm{d}s = 2^{-2n}T^2 G'_+(0).
$$

Define $G(\infty-) := \lim_{t\to\infty} G(t)$. Notice that for every $t \in (0,\infty)$, it holds that $0 \leq G(\infty-) \leq G(t) \leq G(0) < \infty$. Conclude from the linearity of G_n on $[0, 2^{-n}T]$ that

$$
\begin{aligned}
G'_{n,-}(2^{-n}T) = \frac{1}{2^{-n}T}\Big(& 2^{-2n}T^2 \int_{-1}^{1} G\big(2^{-n}T(1+s)\big)(1-|s|) \, \mathrm{d}s \\
& - \gamma 2^{-(n+1)}T - 2^{-2n+1}T^2 \int_0^1 G(2^{-n}Ts)(1-s) \, \mathrm{d}s \Big) \\
\leq 2^{-n}T\Big(& G(0) \int_{-1}^{1} (1-|s|) \, \mathrm{d}s - 2G(\infty-) \int_0^1 (1-s) \, \mathrm{d}s \Big) - \gamma/2 \\
= 2^{-n}T\big(& G(0) - G(\infty-)\big) - \gamma/2.
\end{aligned}
$$

For large n, the factor $-\gamma/2$ becomes dominating. Hence there is an $n_0 \in \mathbb{N}$ such that $G'_{n,-}(2^{-n}T) \leq G'_{n,+}(2^{-n}T)$ for all $n \geq n_0$.

For $n \geq n_0$, consider the problem of minimizing $J^\gamma[\alpha]$ over $\alpha \in \mathcal{A}^1_n$. By Lemma 4.5, this is equivalent to the minimization of the quadratic form

$$\sum_{i,j=0}^{2^n-1} G_n(|t_i - t_j|)a_i a_j \tag{4.6}$$

over constants $a_0, a_1, \ldots, a_{2^n-1}$ that sum to $2^n/T$. Conclude from (4.1) and Lemma 4.5 that (4.6) is always nonnegative. Hence the matrix with entries $G_n(|t_i - t_j|)$ is positive definite and the minimization of the quadratic form admits a unique solution. By Theorem 1 in Alfonsi et al. (2012), all components $a_0, a_1, \ldots, a_{2^n-1}$ of this solution are nonnegative. Thus, the problem of minimizing $J^\gamma[\alpha]$ over $\alpha \in \mathcal{A}^1_n$ admits a unique solution α^*_n, which is nonnegative.

It holds that $\mathcal{A}^1_n \subseteq \mathcal{A}^1_{n+1}$, and therefore $J^\gamma[\alpha^*_{n_0}] \geq J^\gamma[\alpha^*_n]$ for every $n \geq n_0$. Furthermore, since G is of positive type,

$$J^\gamma[\alpha^*_n] = \frac{1}{2}\int_0^T \big(\gamma\alpha^*_n(t)^2 + \int_0^T G(|t-s|)\alpha^*_n(t)\alpha^*_n(s)\,\mathrm{d}s\big)\,\mathrm{d}t \geq \frac{\gamma}{2}\|\alpha^*_n\|^2,$$

where $\|\cdot\|$ denotes the standard L^2-norm on $[0,T]$. Conclude that the L^2-norms of $(\alpha^*_n)_{n\geq n_0}$ are uniformly bounded. By passing to a subsequence if necessary, assume that the sequence $(\alpha^*_n)_{n\geq n_0}$ converges weakly in $L^2[0,T]$ to some nonnegative function α^*. It remains to show that α^* minimizes J^γ over $\mathcal{A}^1$.

Define the filtered probability space $([0,T], \mathcal{B}([0,T]), (\mathcal{F}_n)_{n\in\mathbb{N}}, \lambda)$, where $\mathcal{B}([0,T])$ is the Borel σ-algebra on $[0,T]$, each σ-field $\mathcal{F}_n$ is generated by the intervals $[t_k, t_{k+1})$ for $k = 0, 1, \ldots, 2^n - 1$, and λ is the normalized Lebesgue measure on $[0,T]$. Notice that $\mathcal{F}_\infty \coloneqq \sigma\big(\bigcup_{n\in\mathbb{N}}\mathcal{F}_n\big)$ equals $\mathcal{B}([0,T])$.

Choose an arbitrary $\alpha \in \mathcal{A}^1$. For $n \in \mathbb{N}$, let α_n denote the conditional expectation of α with respect to $\mathcal{F}_n$ under λ. Then α_n belongs to $\mathcal{A}^1_n$, hence $J^\gamma[\alpha^*_n] \leq J^\gamma[\alpha_n]$. Since the sequence $(\alpha_n)_{n\in\mathbb{N}}$ is bounded in $L^2[0,T]$, martingale convergence shows that it converges in $L^2[0,T]$ to $\mathbb{E}_\lambda[\alpha \,|\, \mathcal{F}_\infty] = \alpha$. Recall that G is bounded and continuous. Hence $\lim_{n\to\infty} J^\gamma[\alpha_n] = J^\gamma[\alpha]$. Since J^γ is lower semicontinuous, conclude that α^* is indeed the desired minimizer:

$$J^\gamma[\alpha^*] \leq \liminf_{n\to\infty} J^\gamma[\alpha^*_n] \leq \liminf_{n\to\infty} J^\gamma[\alpha_n] = J^\gamma[\alpha].$$

This concludes the proof for the case $G(0) < \infty$.

Now suppose $G(0) = \infty$. Let ν be the measure defined in the proof of Lemma 4.2. For $n \in \mathbb{N}$, define measures ν_n via $\nu_n(\mathrm{d}t) \coloneqq \mathbb{1}_{(1/n,\infty)}(t)\nu(\mathrm{d}t)$. As in the proof of Proposition 4.5 in Gatheral et al. (2012), consider approximations G_n of G defined via

$$G_n(t) \coloneqq G(\infty-) + \int_{(0,\infty)} (s-t)^+ \nu_n(\mathrm{d}s). \tag{4.7}$$

Each G_n is continuous, nonincreasing, nonnegative and satisfies $G_n(0) < \infty$. The same arguments as in Lemma 4.2 show that every $G_n(|\cdot|)$ corresponds to a function f_n as in (4.3) and a functional J_n^γ as in (4.2), where ν is replaced with ν_n.

Let $(\alpha_n)_{n\in\mathbb{N}}$ be a minimizing sequence for J^γ in $\mathcal{A}^1$. Define the function f as in (4.3). Since $f \geq f_n$, conclude from (4.2) that $J^\gamma[\alpha_n] \geq J_n^\gamma[\alpha_n]$. For each $n \in \mathbb{N}$, it was already shown that a unique minimizer α_n^* of J_n^γ exists and is nonnegative. Conclude that $J^\gamma[\alpha_n] \geq J_n^\gamma[\alpha_n] \geq J_n^\gamma[\alpha_n^*]$. Since $(\alpha_n)_{n\in\mathbb{N}}$ is a minimizing sequence, the preceding inequalities imply that the sequence

$$\Big(\int_0^T \alpha_n^*(t)^2 \,\mathrm{d}t\Big)_{n\in\mathbb{N}}$$

is uniformly bounded. By passing to a subsequence if necessary, assume without loss of generality that the sequence $(\alpha_n^*)_{n\in\mathbb{N}}$ converges weakly in $L^2[0,T]$ to some nonnegative function α^*.

The set $[0,T]$ is compact, hence the Fourier transforms $\hat{\alpha}_n^*$ converge pointwise to $\hat{\alpha}^*$. Since f_n increases pointwise to f, conclude with Fatou's lemma that

$$\begin{aligned}
\inf_{\alpha\in\mathcal{A}^1} J^\gamma[\alpha] &= \lim_{n\to\infty} J^\gamma[\alpha_n] \\
&\geq \liminf_{n\to\infty} J_n^\gamma[\alpha_n^*] \\
&= \liminf_{n\to\infty} \Big(\frac{\gamma}{2}\int |\hat{\alpha}_n^*(t)|^2 \,\mathrm{d}t + G_n(\infty-)|\hat{\alpha}_n^*(0)|^2 + \frac{1}{2}\int |\hat{\alpha}_n^*(t)|^2 f_n(t)\,\mathrm{d}t\Big) \\
&\geq J^\gamma[\alpha^*],
\end{aligned}$$

as desired. □

4.2 Symmetric total monotonicity

Numerical simulations show that for many, but not all, convex and nonincreasing decay kernels, the minimizer α^* of J^γ for $\gamma > 0$ is convex. The results in this section originate from a search for sufficient conditions that ensure the convexity of α^*. It will turn out that total monotonicity is such a condition, but in fact implies much more: In particular, α^* is itself completely monotone on $[0, T/2]$.

A decay kernel $G\colon (0,\infty) \to [0,\infty)$ is said to be *completely monotone* if it is smooth on $(0,\infty)$ and satisfies $(-1)^n G^{(n)} \geq 0$ for all $n \in \mathbb{N}$. If G is completely monotone, then it is convex and nonincreasing and thus of positive type. Furthermore, it can be represented as the Laplace transform of a positive Radon measure on $[0,\infty)$ (Bernstein, 1929). Notice that this is not necessarily true if G is only completely monotone on $(0,\tau)$ for some $\tau < \infty$.

A function $f\colon [0,T] \to \mathbb{R}$ shall be called *symmetrically totally monotone* if it is analytic on $(0,T)$ and there are nonnegative coefficients $(z_{2k})_{k\in\mathbb{N}}$ such that its power series development in $T/2$ is of the form

$$f(t) = \sum_{k=0}^{\infty} z_{2k}(t - T/2)^{2k}, \qquad t \in (0, T).$$

The following theorem is the main result of this chapter.

Theorem 4.6. *Suppose G is completely monotone. Then the following statements are true:*

(i) For every $\gamma > 0$, the unique minimizer of J^γ is symmetrically totally monotone.

(ii) For $\gamma = 0$, let X^* be the unique minimizer of J^0 in $\mathcal{A}^0$. Then $-X^*$ admits a symmetrically totally monotone derivative on $(0, T)$.

The proof of Theorem 4.6 requires some general results about symmetrically totally monotone functions.
For $h > 0$, introduce the notation $\Delta_h f(t) := f(t+h) - f(t)$. Furthermore, say that a function $f\colon [0, T] \to \mathbb{R}$ is *symmetric around* $T/2$ if $f(t) = f(T - t)$ for all $t \in (0, T)$, and that it is *absolutely monotone* if $f^{(n)} \geq 0$ for all $n \in \mathbb{N}$.

Lemma 4.7. *Let $f\colon (0, T) \to \mathbb{R}$ be analytic. Then the following conditions are equivalent:*

(i) f is symmetrically totally monotone.

(ii) f is symmetric around $T/2$, completely monotone on $(0, T/2)$ and absolutely monotone on $(T/2, T)$.

(iii) f is symmetric around $T/2$ and, for every $t \in (T/2, T)$, $n \in \{1, 2, \dots\}$ and $h > 0$ with $t + nh < T$,

$$\Delta_h^n f(t) = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} f(t + kh) \geq 0.$$

Proof. (i) $\Rightarrow$ (iii) is easily proved with induction over n.
(iii) $\Rightarrow$ (ii) follows from results due to Bernstein (1914, p. 451): A function satisfying (iii) (without necessarily being analytic a priori) is absolutely monotone on $(T/2, T)$ and admits an analytic continuation g to all of $(0, T)$. By analyticity, g must coincide with f. Conclude from symmetry around $T/2$ that f is completely monotone on $(0, T/2)$.
(ii) $\Rightarrow$ (i) is again straightforward: Since f is completely monotone on $(0, T/2)$ and absolutely monotone on $(T/2, T)$, conclude that $f^{(2n)}(T/2) \geq 0$ and $f^{(2n+1)}(T/2) = 0$ for all $n \in \mathbb{N}$. Develop f into a power series in $T/2$ to prove the claim. $\square$

The condition of analyticity cannot be dropped. The function $t \mapsto \arcsin(|1-t|)$ on $[0,2]$ provides a counterexample: It is straightforward to show that it satisfies condition (ii); but its first derivative jumps in $t = 1$, so it is not analytic, and therefore not symmetrically totally monotone.

Lemma 4.8. *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of symmetrically totally monotone functions on $(0,T)$, and $\mathbb{T}$ a dense subset of $(0,T)$ such that $\lim_{n\to\infty} f_n(t)$ exists and is finite for every $t \in \mathbb{T}$. Then*

$$f\colon (0,T) \to \mathbb{R}, \qquad f(t) = \lim_{n\to\infty} f_n(t),$$

is well-defined and symmetrically totally monotone. Furthermore, $(f_n)_{n\in\mathbb{N}}$ converges uniformly to f on compact subsets of $(0,T)$, and the coefficients in the power series development $f_n(t) = \sum_{k=0}^\infty z_k^n (t-T/2)^k$ converge to those in the power series development $f(t) = \sum_{k=0}^\infty z_k (t-T/2)^k$.

Proof. Each function f_n is convex. Theorem 10.8 in Rockafellar (1970) shows that f is well-defined, that $(f_n)_{n\in\mathbb{N}}$ converges uniformly on compact subsets, and that f is convex and thus continuous on $(0,T)$.

Clearly, $\lim_{n\to\infty} \Delta_h^k f_n(t) = \Delta_h^k f(t)$ for every $t \in (T/2,T), n \in \{1,2,\dots\}$ and $h > 0$ with $t + nh < T$. As stated in the proof of Lemma 4.7, it follows from Bernstein (1914) that f is analytic on $(T/2,T)$ and admits an analytic continuation g to all of $(0,T)$.

Since each f_n is symmetrically totally monotone, there are sequences $(z_k^n)_{k\in\mathbb{N}}$ satisfying $z_{2k}^n \geq 0$ and $z_{2k+1}^n = 0$ for every $k \in \mathbb{N}$ such that $f_n(t) = \sum_{k=0}^\infty z_k^n (t-T/2)^k$. Furthermore, there is a sequence $(z_k)_{k\in\mathbb{N}}$ such that $g(t) = \sum_{k=0}^\infty z_k (t-T/2)^k$.

Notice that

$$z_0 = g(T/2) = \lim_{n\to\infty} f_n(T/2) = \lim_{n\to\infty} z_0^n$$

by continuity. Next, consider the functions $f_{n,1}(t) = \sum_{k=0}^\infty z_{k+1}^n |t-T/2|^k$. Conclude that each $f_{n,1}$ is convex and that $f_{n,1}(T/2) = z_1^n$ and

$$f_{n,1}(t) = \operatorname{sgn}(t-T/2) \sum_{k=0}^\infty z_{k+1}^n (t-T/2)^k = \frac{f_n(t) - z_0^n}{|t-T/2|}$$

for $t \in (0,T/2) \cup (T/2,T)$. Hence the sequence $(f_{n,1})_{n\in\mathbb{N}}$ converges pointwise on $(0,T/2) \cup (T/2,T)$ to the function $g_1(t) := (g(t) - z_0)/|t-T/2|$. Using once again Theorem 10.8 from Rockafellar (1970), conclude that g_1 has a continuous and convex extension to all of $(0,T)$ and that the sequence $(f_{n,1})_{n\in\mathbb{N}}$ converges to g_1 locally uniformly. It follows that

$$z_1 = g_1(T/2) = \lim_{n\to\infty} f_{n,1}(T/2) = \lim_{n\to\infty} z_1^n.$$

By considering the convex functions $f_{n,2}(t) = \sum_{k=0}^\infty z_{k+2}^n (t-T/2)^k$, conclude with the same arguments that $z_2 = \lim_{n\to\infty} z_2^n$. Iterate this argument to show that $z_k = \lim_{n\to\infty} z_k^n$ and thus $z_{2k} \geq 0$ and $z_{2k+1} = 0$ for every $k \in \mathbb{N}$.

This implies that g is symmetrically totally monotone. Recall that $f = g$ on $(T/2, T)$ by construction. By symmetry of the f_n, the function f is symmetric around $T/2$, and so is g. Thus $f = g$ on $(0, T/2)$. By continuity, $f(T/2) = g(T/2)$. □

Lemma 4.9. *The class of all symmetrically totally monotone functions in $L^2[0,T]$ is weakly closed in $L^2[0,T]$.*

Proof. Let $\mathcal{M}$ denote the class of all finite measures on $[0,T]$ whose restrictions to $(0,T)$ admit a symmetrically totally monotone Lebesgue density. Show first that $\mathcal{M}$ is weakly closed with respect to convergence of measures.
Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}$ that converges weakly to a finite measure μ on $[0,T]$. Denote by $F_n(t) := \mu_n([0,t])$ and $F(t) := \mu([0,T])$ the corresponding distribution functions. Then $F_n(t) \to F(t)$ for all continuity points of F and hence on a dense subset of $(0,T)$. Since each F_n is the integral of an absolutely monotone function on $(T/2, T)$, it is absolutely monotone itself. In particular, it is convex on $[T/2, T]$. By symmetry, $F_n(t) = F_n(T) + \mu_n(\{T-t\}) - F_n(T-t)$ for $t \in [0, T/2]$. Since $\mu_n(\{T\}) \geq 0$ and $\mu_n(\{T-t\}) = 0$ for all $t \in (0, T/2]$, conclude that F_n is concave on $[0, T/2]$.
As in the proof of Lemma 4.8, conclude that the sequence $(F_n)_{n\in\mathbb{N}}$ converges to F for all $t \in (0,T)$ with a symmetrically totally monotone derivative on $(0,T)$.
Now let $(f_n)_{n\in\mathbb{N}}$ be a sequence of symmetrically totally monotone functions that converge in $L^2[0,T]$ to some function f. Then f is nonnegative, and the finite measures $f_n(t)\,\mathrm{d}t$ converge weakly to a finite measure $f(t)\,\mathrm{d}t$. Since $\mathcal{M}$ is weakly closed with respect to convergence of measures, f is symmetrically totally monotone. Conclude that the class of all symmetrically totally monotone functions is closed in $L^2[0,T]$ and thus weakly closed. □

The proof of Theorem 4.6 is in two steps: First, the statement is shown for $\gamma > 0$ and *generalized exponential decay kernels*, i.e., decay kernels $G\colon (0,\infty) \to [0,\infty)$ of the form

$$G(t) = \sum_{k=1}^{n} a_k e^{-\sqrt{b_k}t}, \qquad t \in (0,\infty), \tag{4.8}$$

where $n \in \mathbb{N}$ and $a_1, a_2, \ldots, a_n > 0$ and $b_n > b_{n-1} > \cdots > b_1 > 0$. Then the general case is deduced from this.

Proof of Theorem 4.6 for $\gamma > 0$ and generalized exponential kernels.
Let $\gamma > 0$ and assume that G is of the form (4.8). Since G is completely monotone, a unique minimizer $\alpha^* \in \mathcal{A}^1$ of J^γ exists by Theorem 4.4. Furthermore, G is square-integrable and $G(0) := \lim_{t\to 0} G(t)$ is finite. Recall the following facts from Section 3.2: By Lemma 3.9, the function α^* is symmetric around $T/2$, and by Lemma 3.4, there is an $\eta \in \mathbb{R}$ such that α^* satisfies the following Fredholm integral equation of the second kind:

$$\gamma\alpha^*(t) + \int_0^T G(|t-s|)\alpha^*(s)\,\mathrm{d}s = \eta, \qquad t \in [0,T]. \tag{4.9}$$

The latter also follows from Lemma 4.3.
It is enough to consider the case $\eta = \gamma$. All matrices considered in this proof are n-dimensional square matrices, and all vectors n-dimensional column vectors. Denote the diagonal matrix with $x_1, x_2, \ldots, x_n$ on its main diagonal as $\operatorname{diag}(x_i)_{i=1,2,\ldots,n}$ and say that a matrix is a *positive diagonal matrix* if it is diagonal and all diagonal entries are positive.
Define $A := \operatorname{diag}(a_i)_{i=1,2,\ldots,n}$ and $B := \operatorname{diag}(b_i)_{i=1,2,\ldots,n}$. Define $\psi = (\psi_1, \psi_2, \ldots, \psi_n)$ via

$$\psi_k(t) := a_k \int_0^T e^{-\sqrt{b_k}|t-s|} \alpha^*(s)\,\mathrm{d}s, \qquad t \in [0,T], \quad k = 1, 2, \ldots, n.$$

Let $\lambda := 1/\gamma$ and $\mathbf{1} := (1, 1, \ldots, 1) \in \mathbb{R}^n$. Conclude from (4.9) that

$$\alpha^* = 1 - \lambda \sum_k \psi_k = 1 - \lambda\, \mathbf{1}^\top \psi. \tag{4.10}$$

The proof proceeds as follows:

1. Show that ψ solves a system of n ordinary differential equations $\psi'' = M\psi - 2AB^{1/2}\mathbf{1}$ with boundary conditions $\psi(0) = \psi(T)$ and $\psi'(0) = B^{1/2}\psi(0)$. Here, M is a nonsingular matrix.

2. Show that M has n distinct, real eigenvalues $c_n > c_{n-1} > \cdots > c_1 > 0$. Define $C := \operatorname{diag}(c_i)_{i=1,2,\ldots,n}$. Obtain an eigendecomposition $M = QCQ^{-1}$, where Q is a nonsingular matrix.

3. Conclude with Step 1. that

$$\alpha^*(t) = d\left(1 + 2\lambda\, \mathbf{1}^\top \left(e^{M^{1/2}t} + e^{M^{1/2}(T-t)}\right) N^{-1}\mathbf{1}\right), \qquad t \in [0,T], \tag{4.11}$$

   where d is a positive constant and N is a nonsingular matrix.

4. Use the eigendecomposition of M to rewrite (4.11) as

$$\alpha^*(t) = d\left(1 + \mathbf{1}^\top E(t) \tilde{N}^{-1}\mathbf{1}\right), \qquad t \in [0,T]. \tag{4.12}$$

   Here, $\tilde{N}$ is a nonsingular matrix and $\big(E(t)\big)_{t\in[0,T]}$ is a class of positive diagonal matrices where each diagonal entry is a symmetrically totally monotone function of t.

5. Decompose $\tilde{N}^{-1} = \tilde{N}_1(\tilde{N}_2 + \tilde{N}_3)^{-1}\tilde{N}_4$ such that $\tilde{N}_1$ and $\tilde{N}_3$ are positive diagonal matrices, $\tilde{N}_2$ is positive definite, and all off-diagonal entries of $(\tilde{N}_2 + \tilde{N}_3)^{-1}$ are nonpositive.

6. Show that all entries of $\tilde{N}_2^{-1}\tilde{N}_4\,\mathbf{1}$ are nonnegative. Show that this implies that all entries of $\tilde{N}^{-1}\mathbf{1}$ are nonnegative.

7. Conclude with (4.12) and Step 6. that α^* is symmetrically totally monotone.

1. Recall that $A = \operatorname{diag}(a_i)_{i=1,2,\dots,n}$ and $B = \operatorname{diag}(b_i)_{i=1,2,\dots,n}$ are positive diagonal matrices, and that $b_n > b_{n-1} > \cdots > b_1$. Notice that $\mathbf{1}\mathbf{1}^\top$ is the matrix containing only ones. Define

$$M := B + 2\lambda\, AB^{1/2}\mathbf{1}\mathbf{1}^\top = \begin{pmatrix} b_1 + 2\lambda a_1\sqrt{b_1} & 2\lambda a_1\sqrt{b_1} & \cdots & 2\lambda a_1\sqrt{b_1} \\ 2\lambda a_2\sqrt{b_2} & b_2 + 2\lambda a_2\sqrt{b_2} & \cdots & 2\lambda a_2\sqrt{b_2} \\ \vdots & \vdots & \ddots & \vdots \\ 2\lambda a_n\sqrt{b_n} & 2\lambda a_n\sqrt{b_n} & \cdots & b_n + 2\lambda a_n\sqrt{b_n} \end{pmatrix}.$$

1.1 *ψ solves the system of n ordinary differential equations $\psi'' = M\psi - 2AB^{1/2}\mathbf{1}$.*
Let $t \in [0,T]$ and $k = 1,2,\dots,n$. Differentiating and plugging in from (4.10) shows

$$\begin{aligned} \psi_k''(t) &= a_k\sqrt{b_k}\,\frac{\mathrm{d}}{\mathrm{d}t}\Big[-\int_0^t e^{-\sqrt{b_k}(t-s)}\alpha^*(s)\,\mathrm{d}s + \int_t^T e^{-\sqrt{b_k}(s-t)}\alpha^*(s)\,\mathrm{d}s\Big] \\ &= a_k\sqrt{b_k}\Big(\sqrt{b_k}\int_0^T e^{-\sqrt{b_k}|t-s|}\alpha^*(s)\,\mathrm{d}s - 2\alpha^*(t)\Big) \\ &= b_k\psi_k(t) - 2a_k\sqrt{b_k}\alpha^*(t) \\ &= b_k\psi_k(t) - 2a_k\sqrt{b_k}\big(1 - \lambda\sum_l \psi_l(t)\big). \end{aligned}$$

Conclude $\psi'' = (B + 2\lambda\, AB^{1/2}\mathbf{1}\mathbf{1}^\top)\psi - 2AB^{1/2}\mathbf{1} = M\psi - 2AB^{1/2}\mathbf{1}$.

1.2 *$\psi(0) = \psi(T)$.*
By Lemma 3.9, α^* is symmetric around $T/2$, i.e., $\alpha^*(t) = \alpha^*(T-t)$ for $t \in [0,T]$. Let $t \in [0,T]$ and $k = 1,2,\dots,n$. Integration by substitution shows

$$\begin{aligned} \psi_k(t) &= a_k\int_0^T e^{-\sqrt{b_k}|t-s|}\alpha^*(s)\,\mathrm{d}s \\ &= a_k\int_0^T e^{-\sqrt{b_k}|(T-t)-(T-s)|}\alpha^*(T-s)\,\mathrm{d}s \\ &= a_k\int_0^T e^{-\sqrt{b_k}|(T-t)-s|}\alpha^*(s)\,\mathrm{d}s \\ &= \psi_k(T-t). \end{aligned}$$

In particular, $\psi_k(0) = \psi_k(T)$.

1.3 *$\psi'(0) = B^{1/2}\psi(0)$.*
Let $k = 1,2,\dots,n$. Then

$$\begin{aligned} \psi_k'(0) &= a_k\sqrt{b_k}\Big[-\int_0^t e^{-\sqrt{b_k}(t-s)}\alpha^*(s)\,\mathrm{d}s + \int_t^T e^{-\sqrt{b_k}(s-t)}\alpha^*(s)\,\mathrm{d}s\Big]_{t=0} \\ &= a_k\sqrt{b_k}\int_0^T e^{-\sqrt{b_k}s}\alpha^*(s)\,\mathrm{d}s \\ &= \sqrt{b_k}\,\psi_k(0). \end{aligned}$$

2.1 *M has n distinct, real eigenvalues $c_1, c_2, \ldots, c_n$. They satisfy the interlacing inequalities $c_n > b_n > c_{n-1} > b_{n-1} > \cdots > c_1 > b_1 > 0$.*
Let $v := 2\lambda\,(a_1\sqrt{b_1}, a_2\sqrt{b_2}, \ldots, a_n\sqrt{b_n})$ and $x \in \mathbb{R}_+ \setminus \{b_1, b_2, \ldots, b_n\}$. By the matrix determinant lemma,

$$\begin{aligned}
\det(xI - M) &= \det(xI - B - v\mathbf{1}^\top) \\
&= \big(1 - v^\top (xI - B)^{-1}\mathbf{1}\big)\det(xI - B) \\
&= \Big(1 - 2\lambda \sum_k \frac{a_k\sqrt{b_k}}{x - b_k}\Big)\prod_k (x - b_k).
\end{aligned}$$

The following argument is due to Terrell (2017). Define $f \colon \mathbb{R}_+ \setminus \{b_1, b_2, \ldots, b_n\} \to \mathbb{R}$ via

$$f(x) := 1 - 2\lambda \sum_k \frac{a_k\sqrt{b_k}}{x - b_k}.$$

Let $k = 1, 2, \ldots, n-1$. Then f is continuous on (b_k, b_{k+1}), with

$$\lim_{x \searrow b_k} f(x) = -\infty \qquad \text{and} \qquad \lim_{x \nearrow b_{k+1}} f(x) = +\infty.$$

Conclude that f has a root $c_k \in (b_k, b_{k+1})$. Furthermore,

$$\lim_{x \searrow b_n} f(x) = -\infty \qquad \text{and} \qquad \lim_{x \nearrow +\infty} f(x) = 1,$$

showing that f has another root $c_n \in (b_n, +\infty)$. Since $\det(c_k I - M) = 0$ for $k = 1, 2, \ldots, n$, each c_k is an eigenvalue of M.

2.2 *If c is an eigenvalue of M, then*

$$\Big(\frac{a_1\sqrt{b_1}}{c - b_1}, \frac{a_2\sqrt{b_2}}{c - b_2}, \ldots, \frac{a_n\sqrt{b_n}}{c - b_n}\Big)$$

is a corresponding eigenvector.
Let $c \in \mathbb{R}_+ \setminus \{b_1, b_2, \ldots, b_n\}$ be an eigenvalue of M, and $v = (v_1, v_2, \ldots, v_n)$ a corresponding eigenvector. The definition $Mv = cv$ translates into the following system of equations:

$$b_k v_k + 2\lambda a_k \sqrt{b_k} \sum_l v_l = c v_k, \qquad k = 1, 2, \ldots, n.$$

It must be true that $\sum_l v_l \neq 0$. Otherwise $b_k v_k = c v_k$ for all $k = 1, 2, \ldots, n$. Since $c \notin \{b_1, b_2, \ldots, b_n\}$ (see Step 2.1), this implies $v = 0$, which contradicts the definition of an eigenvector. Hence one may set $\sum_l v_l = 1/(2\lambda)$ without loss of generality. Obtain

$$v = \Big(\frac{a_1\sqrt{b_1}}{c - b_1}, \frac{a_2\sqrt{b_2}}{c - b_2}, \ldots, \frac{a_n\sqrt{b_n}}{c - b_n}\Big).$$

Now let $c_n > c_{n-1} > \cdots > c_1 > 0$ be the eigenvalues of M.

Define $C := \operatorname{diag}(c_i)_{i=1,2,\dots,n}$,

$$\tilde{Q} := \Big(\frac{1}{c_j - b_i}\Big)_{i,j=1,2,\dots,n} \qquad \text{and} \qquad Q := AB^{1/2}\tilde{Q}.$$

2.3 $M = QCQ^{-1}$.
By Step 2.2, the columns of Q are eigenvectors corresponding to the eigenvalues $c_1, c_2, \dots, c_n$. Eigenvectors corresponding to different eigenvalues are linearly independent, hence Q is nonsingular. Obtain the eigendecomposition $M = QCQ^{-1}$.

2.4 $\mathbf{1}^\top Q = 1/(2\lambda)\,\mathbf{1}^\top$.
This follows from Step 2.2, where each eigenvector in Q was assumed to sum to $1/(2\lambda)$.

3. Define

$$d := \Big(1 + 2\lambda \sum_k \frac{a_k}{\sqrt{b_k}}\Big)^{-1} > 0.$$

Let $M^{1/2} := Q\operatorname{diag}(\sqrt{c_i})_{i=1,2,\dots,n}\,Q^{-1}$ and denote by

$$e^{M^{1/2}T} = Q\operatorname{diag}(e^{\sqrt{c_i}T})_{i=1,2,\dots,n}\,Q^{-1}$$

the matrix exponential of $M^{1/2}T$. Define

$$N := A^{-1}\big(M^{1/2}\big(e^{M^{1/2}T} - I\big) + B^{1/2}\big(e^{M^{1/2}T} + I\big)\big),$$

where I denotes the identity matrix.
The general solution of $f'' = Mf - 2AB^{1/2}\mathbf{1}$ is

$$f(t) = e^{M^{1/2}t}z_0 + e^{M^{1/2}(T-t)}z_1 + 2dAB^{-1/2}\mathbf{1}, \qquad t \in [0,T],$$

for $z_0, z_1 \in \mathbb{R}^n$. To see this, let $t \in [0,T]$ and $z_0, z_1 \in \mathbb{R}^n$. Writing $d = 1/(1 + 2\lambda\,\mathbf{1}^\top AB^{-1/2}\mathbf{1})$ shows

$$\begin{aligned} d\,MAB^{-1/2}\mathbf{1} &= d\,\big(AB^{1/2}\mathbf{1} + 2\lambda\,AB^{1/2}\mathbf{1}\,\mathbf{1}^\top AB^{-1/2}\mathbf{1}\big) \\ &= d\,\Big(1 + \frac{1}{d} - 1\Big)AB^{1/2}\mathbf{1} \\ &= AB^{1/2}\mathbf{1}. \end{aligned}$$

Therefore,

$$\begin{aligned} f''(t) &= M\big(e^{M^{1/2}t}z_0 + e^{M^{1/2}(T-t)}z_1\big) \\ &= Mf(t) - 2d\,MAB^{-1/2}\mathbf{1} \\ &= Mf(t) - 2AB^{1/2}\mathbf{1}. \end{aligned}$$

It remains to choose z_0 and z_1 in such a way that the boundary conditions from Steps 1.2 and 1.3 are satisfied. First, $f(0) - f(T) = (e^{M^{1/2}T} - I)(z_1 - z_0)$.

By Step 2.3, the matrix

$$e^{M^{1/2}T} - I = Q\,\mathrm{diag}\left(e^{\sqrt{c_i}T}\right)_{i=1,2,\dots,n} Q^{-1} - I = Q\,\mathrm{diag}\left(e^{\sqrt{c_i}T} - 1\right)_{i=1,2,\dots,n} Q^{-1}$$

is nonsingular. Hence $f(0) = f(T)$ if and only if $z_0 = z_1$. Set $z_0 = z_1$. Second,

$$\begin{aligned} f'(0) - B^{1/2} f(0) &= \left(M^{1/2}\left(I - e^{M^{1/2}T}\right) - B^{1/2}\left(I + e^{M^{1/2}T}\right)\right) z_0 - 2dA\mathbf{1} \\ &= -A(Nz_0 + 2d\,\mathbf{1}). \end{aligned}$$

It is shown in Step 5.5 that N is nonsingular. Hence, $f'(0) = B^{1/2} f(0)$ if and only if $z_0 = -2d\,N^{-1}\mathbf{1}$. Conclude

$$\begin{aligned} \psi(t) &= e^{M^{1/2}t} z_0 + e^{M^{1/2}(T-t)} z_1 + 2d\,AB^{-1/2}\mathbf{1} \\ &= \left(e^{M^{1/2}t} + e^{M^{1/2}(T-t)}\right) z_0 + 2d\,AB^{-1/2}\mathbf{1} \\ &= 2d\left(AB^{-1/2} - \left(e^{M^{1/2}t} + e^{M^{1/2}(T-t)}\right) N^{-1}\right)\mathbf{1} \end{aligned}$$

for all $t \in [0,T]$. Notice that $1 - 2d\lambda\,\mathbf{1}^\top AB^{-1/2}\mathbf{1} = 1 - d(1/d - 1) = d$, so

$$\begin{aligned} \alpha^*(t) &= 1 - \lambda\,\mathbf{1}^\top \psi(t) \\ &= 1 - 2d\lambda\,\mathbf{1}^\top AB^{-1/2}\mathbf{1} + 2d\lambda\,\mathbf{1}^\top\left(e^{M^{1/2}t} + e^{M^{1/2}(T-t)}\right) N^{-1}\mathbf{1} \\ &= d\left(1 + 2\lambda\,\mathbf{1}^\top\left(e^{M^{1/2}t} + e^{M^{1/2}(T-t)}\right) N^{-1}\mathbf{1}\right) \end{aligned}$$

for all $t \in [0,T]$.

4. Define

$$E(t) := \mathrm{diag}\left(\frac{e^{\sqrt{c_i}t} + e^{\sqrt{c_i}(T-t)}}{e^{\sqrt{c_i}T} - 1}\right)_{i=1,2,\dots,n,} \qquad t \in [0,T].$$

$E(t)$ is a positive diagonal matrix for all $t \in [0,T]$. In particular, it is nonsingular. Define further

$$\tilde{N} := A^{-1}\left(QC^{1/2} + B^{1/2}QE(T)\right).$$

With Step 2.3, obtain

$$\begin{aligned} N &= A^{-1}\left(QC^{1/2}\,\mathrm{diag}\left(e^{\sqrt{c_i}T} - 1\right)_{i=1,2,\dots,n} Q^{-1} + B^{1/2}Q\,\mathrm{diag}\left(e^{\sqrt{c_i}T} + 1\right)_{i=1,2,\dots,n} Q^{-1}\right) \\ &= A^{-1}\left(QC^{1/2} + B^{1/2}QE(T)\right)\mathrm{diag}\left(e^{\sqrt{c_i}T} - 1\right)_{i=1,2,\dots,n} Q^{-1} \\ &= \tilde{N}\,\mathrm{diag}\left(e^{\sqrt{c_i}T} - 1\right)_{i=1,2,\dots,n} Q^{-1}. \end{aligned}$$

Hence N is nonsingular if and only if $\tilde{N}$ is nonsingular. This, in combination with Steps 2.3, 2.4 and 3., shows, for all $t \in [0,T]$,

$$\begin{aligned} \alpha^*(t) &= d\left(1 + 2\lambda\,\mathbf{1}^\top Q\,\mathrm{diag}\left(e^{\sqrt{c_i}t} + e^{\sqrt{c_i}(T-t)}\right)_{i=1,2,\dots,n} Q^{-1}N^{-1}\mathbf{1}\right) \\ &= d\left(1 + \mathbf{1}^\top E(t)\tilde{N}^{-1}\mathbf{1}\right). \end{aligned}$$

5. Define the real-valued functions

$$\beta(x) := \prod_l (x - b_l), \qquad \gamma(x) := \prod_l (x - c_l).$$

Let

$$D_1 := \operatorname{diag}\Big(\frac{\beta(c_k)}{\gamma'(c_k)}\Big)_{k=1,2,\ldots,n,} \qquad \text{and} \qquad D_2 := \operatorname{diag}\Big(-\frac{\gamma(b_k)}{\beta'(b_k)}\Big)_{k=1,2,\ldots,n.}$$

It will be shown in Step 5.2 that D_1 and D_2 are positive diagonal matrices. In particular, they are nonsingular.

5.1 *$\tilde{Q}$ is nonsingular. It holds that $\tilde{Q}^{-1} = D_1\tilde{Q}^\top D_2$ and $\tilde{Q}^{-1}\mathbf{1} = D_1\mathbf{1}$.*
The matrix $-\tilde{Q}$ is known as a Cauchy matrix. The results are due to Schechter (1959).

5.2 *D_1 and D_2 are positive diagonal matrices.*
Let $k = 1, 2, \ldots, n$. Then

$$\frac{\beta(c_k)}{\gamma'(c_k)} = \frac{\prod_l (c_k - b_l)}{\sum_m \prod_{l\neq m}(c_k - c_l)} = \frac{\prod_l (c_k - b_l)}{\prod_{l\neq k}(c_k - c_l)} = (c_k - b_k)\prod_{l\neq k}\frac{c_k - b_l}{c_k - c_l}.$$

Recall from Step 2.1 that $c_k > b_k$, and, for each $l = 1, 2, \ldots, n$, that $c_k > b_l$ if and only if $c_k > c_l$. Similarly,

$$-\frac{\gamma(b_k)}{\beta'(c_k)} = (c_k - b_k)\prod_{l\neq k}\frac{b_k - c_l}{b_k - b_l} > 0.$$

Now define

$$\tilde{N}_1 := C^{-1/2}, \quad \tilde{N}_2 := \tilde{Q}^\top D_2 B^{-1/2}\tilde{Q}, \quad \tilde{N}_3 := D_1^{-1}E(T)C^{-1/2}, \quad \tilde{N}_4 := \tilde{Q}^\top D_2 B^{-1}.$$

All four matrices are nonsingular (see Steps 5.1 and 5.2 in particular). It will be shown in Step 5.5 that $\tilde{N}_2 + \tilde{N}_3$ is nonsingular.

5.3 $\tilde{N}^{-1} = \tilde{N}_1(\tilde{N}_2 + \tilde{N}_3)^{-1}\tilde{N}_4$.
By definition, $\tilde{Q} = A^{-1}B^{-1/2}Q$. Using Step 5.1:

$$\begin{aligned}
\tilde{N}_4^{-1}(\tilde{N}_2 + \tilde{N}_3)\tilde{N}_1^{-1} &= BD_2^{-1}\tilde{Q}^{-T}\big(\tilde{Q}^\top D_2 B^{-1/2}\tilde{Q} + D_1^{-1}E(T)C^{-1/2}\big)C^{1/2} \\
&= B^{1/2}\tilde{Q}C^{1/2} + B(D_1\tilde{Q}^\top D_2)^{-1}E(T) \\
&= A^{-1}QC^{1/2} + B\tilde{Q}E(T) \\
&= A^{-1}\big(QC^{1/2} + B^{1/2}QE(T)\big) \\
&= \tilde{N}.
\end{aligned}$$

5.4 *$\tilde{N}_1$ and $\tilde{N}_3$ are positive diagonal matrices, and $\tilde{N}_2$ is positive definite.*
B, C and $E(T)$ are positive diagonal matrices. By Step 5.2, the same is true for D_1 and D_2. Hence $D_2 B^{-1/2}$ is positive definite. Since $\tilde{Q}$ is nonsingular (see Step 5.1), $\tilde{N}_2 = \tilde{Q}^\top D_2 B^{-1/2} \tilde{Q}$ is also positive definite.

5.5 *$\tilde{N}_2 + \tilde{N}_3, \tilde{N}$ and N are nonsingular.*
It follows from Step 5.4 that $\tilde{N}_2 + \tilde{N}_3$ is positive definite. Every positive definite matrix is nonsingular (Horn and Johnson, 2013, Corollary 7.1.7). Since $\tilde{N}_1$ and $\tilde{N}_3$ are nonsingular, $\tilde{N}$ is nonsingular. It was shown in Step 4. that N is nonsingular if (and only if) $\tilde{N}$ is nonsingular.

A square matrix is called a *Z-matrix* if all its off-diagonal entries are nonpositive. Given that some matrix U is a Z-matrix, the following two conditions are equivalent:

(M1) There exists a positive diagonal matrix V such that $UV + VU^\top$ is positive definite.

(M2) U is nonsingular and all entries of U^{-1} are nonnegative.

In this case, U is called an *M-matrix.* In particular, (M1) implies that every positive definite Z-matrix is an M-matrix. See Theorem 2.3 in Berman and Plemmons (1994) for proofs and further equivalent characterizations of M-matrices.

5.6 *$\tilde{N}_2^{-1}$ is a Z-matrix.*
With Step 5.1, obtain

$$\tilde{N}_2^{-1} = \tilde{Q}^{-1} B^{1/2} D_2^{-1} \tilde{Q}^{-T} = D_1 \tilde{Q}^\top D_2 B^{1/2} \tilde{Q} D_1.$$

D_1 is a positive diagonal matrix, so it suffices to show that all off-diagonal entries of $\tilde{Q}^\top D_2 B^{1/2} \tilde{Q}$ are nonpositive. Fix $i, j \in \{1, 2, \ldots, n\}$ such that $i \neq j$. Define

$$\begin{aligned} z &:= \left(\tilde{Q}^\top D_2 B^{1/2} \tilde{Q}\right)_{ij} \\ &= -\sum_k \frac{\sqrt{b_k}\,\gamma(b_k)}{(b_k - c_i)(b_k - c_j)\beta'(b_k)} \\ &= -\sum_k \frac{\sqrt{b_k}\prod_{l\neq i,j}(b_k - c_l)}{\prod_{l\neq k}(b_k - b_l)}. \end{aligned}$$

The following argument is due to Petrov (2017). Define $f\colon \mathbb{R}_+ \to \mathbb{R}$,

$$f(x) = -\sqrt{x}\prod_{l\neq i,j}(x - c_l).$$

There are positive constants $z_0, z_1, \ldots, z_{n-2}$ such that

$$f(x) = -\sum_{k=0}^{n-2}(-1)^{n-2-k} z_k\, x^{k+1/2} = \sum_{k=0}^{n-2}(-1)^{n-1-k} z_k\, x^{k+1/2}.$$

Differentiating $n-1$ times yields

$$f^{(n-1)}(x) = \sum_{k=0}^{n-2} \Big((-1)^{n-1-k} z_k \, x^{k-n+3/2} \prod_{l=0}^{n-2} (k+1/2-l) \Big).$$

Let $k = 0, 1, \ldots, n-2$. The factor $k+1/2-l$ is positive if $l = 0, 1, \ldots, k$ and negative if $l = k+1, k+2, \ldots, n-2$. Hence

$$\begin{aligned}
(-1)^{n-1-k} \prod_{l=0}^{n-2} (k+1/2-l) &= (-1)^{n-1-k} (-1)^{n-2-(k+1)+1} \prod_{l=0}^{n-2} |k+1/2-l| \\
&= -\prod_{l=0}^{n-2} |k+1/2-l| \\
&< 0.
\end{aligned}$$

Conclude that $f^{(n-1)}(x) < 0$ for all $x > 0$.

The Lagrange polynomial interpolation $p\colon [0,\infty) \to \mathbb{R}$ of f in $b_1, b_2, \ldots, b_n$ is

$$\begin{aligned}
p(x) &= \sum_k f(b_k) \prod_{l \neq k} \frac{x-b_l}{b_k-b_l} \\
&= \Big(-\sum_k \frac{\sqrt{b_k} \prod_{l \neq i,j} (b_k - c_l)}{\prod_{l \neq k} (b_k - b_l)} \Big) x^{n-1} + q(x) \\
&= z x^{n-1} + q(x)
\end{aligned}$$

for some polynomial q of degree at most $n-2$. The interpolation is exact in $x = b_1, b_2, \ldots, b_n$. By Rolle's theorem, there is an $\tilde{x} > 0$ such that $f^{(n-1)}(\tilde{x}) = p^{(n-1)}(\tilde{x})$ (Milne-Thomson, 2000, Chapter 1). Hence

$$0 > f^{(n-1)}(\tilde{x}) = p^{(n-1)}(\tilde{x}) = (n-1)!\, z,$$

showing that $z = \big(\tilde{Q}^\top D_2 B^{1/2} \tilde{Q}\big)_{ij}$ is nonpositive if $i \neq j$.

5.7 *$\tilde{N}_2^{-1}$ is a nonsingular M-matrix.*
According to Step 5.6, $\tilde{N}_2^{-1}$ is a Z-matrix. Since $\tilde{N}_2$ is positive definite by Step 5.4, its inverse $\tilde{N}_2^{-1}$ is positive definite as well (Horn and Johnson, 2013, Theorem 7.2.1). Hence $\tilde{N}_2^{-1}$ is a nonsingular M-matrix by (M1).

5.8 *All entries of $(\tilde{N}_2^{-1} + \tilde{N}_3^{-1})^{-1}$ are nonnegative.*
As a positive diagonal matrix, $\tilde{N}_3^{-1}$ is positive definite and a nonsingular M-matrix. The sum of positive definite Z-matrices is again a positive definite Z-matrix. Conclude that $\tilde{N}_2^{-1} + \tilde{N}_3^{-1}$ is a positive definite Z-matrix (see Step 5.7); and therefore an M-matrix. By (M2), all entries of $(\tilde{N}_2^{-1} + \tilde{N}_3^{-1})^{-1}$ are nonnegative.

5.9 *All off-diagonal entries of* $(\tilde{N}_2 + \tilde{N}_3)^{-1}$ *are nonpositive.*
By the Woodbury matrix identity,

$$(\tilde{N}_2 + \tilde{N}_3)^{-1} = \tilde{N}_3^{-1} - \tilde{N}_3^{-1}(\tilde{N}_2^{-1} + \tilde{N}_3^{-1})^{-1}\tilde{N}_3^{-1}.$$

Recall that $\tilde{N}_3^{-1}$ is a positive diagonal matrix. It follows from Step 5.8 that all off-diagonal entries of $\tilde{N}_3^{-1}(\tilde{N}_2^{-1} + \tilde{N}_3^{-1})^{-1}\tilde{N}_3^{-1}$ are nonnegative.

6.1 *All entries of* $\tilde{N}_2^{-1}\tilde{N}_4\,\mathbf{1}$ *are nonnegative.*
Using Step 5.1, obtain

$$\begin{aligned}\tilde{N}_2^{-1}\tilde{N}_4\,\mathbf{1} &= \tilde{Q}^{-1}B^{1/2}D_2^{-1}\tilde{Q}^{-T}\tilde{Q}^\top D_2B^{-1}\mathbf{1}\\ &= D_1\tilde{Q}^\top D_2B^{-1/2}\,\mathbf{1}\\ &= D_1\tilde{Q}^\top D_2B^{-1/2}\tilde{Q}\,\tilde{Q}^{-1}\mathbf{1}\\ &= D_1\tilde{N}_2D_1\mathbf{1}.\end{aligned}$$

By Step 5.7, $\tilde{N}_2^{-1}$ is a nonsingular M-matrix. Hence all entries of $\tilde{N}_2$ are nonnegative by (M2). The same is true for D_1 by Step 5.2.

6.2 *All entries of* $(\tilde{N}_2 + \tilde{N}_3)^{-1}\tilde{N}_2$ *are nonnegative.*
Define $U := (\tilde{N}_2 + \tilde{N}_3)^{-1}\tilde{N}_2$. Writing $U = I - (\tilde{N}_2 + \tilde{N}_3)^{-1}\tilde{N}_3$ shows that all off-diagonal entries of U are nonnegative (see Steps 5.4 and 5.9).

Now use the following result about positive definite matrices: If two matrices M_1 and M_2 are positive definite, then $M_1 - M_2$ is positive definite if and only if $M_2^{-1} - M_1^{-1}$ is positive definite (Horn and Johnson, 2013, Corollary 7.7.4). The matrices $(\tilde{N}_2 + \tilde{N}_3), \tilde{N}_3$ and $(\tilde{N}_2 + \tilde{N}_3) - \tilde{N}_3 = \tilde{N}_2$ are positive definite (see Step 5.4). Hence

$$\tilde{N}_3^{-1} - (\tilde{N}_2 + \tilde{N}_3)^{-1} = U\tilde{N}_3^{-1}$$

is positive definite. All entries on the main diagonal of a positive definite matrix are nonnegative. Therefore, all entries on U's main diagonal are nonnegative.

6.3 *All entries of* $\tilde{N}^{-1}\mathbf{1}$ *are nonnegative.*
All entries of $\tilde{N}_1$, $(\tilde{N}_2+\tilde{N}_3)^{-1}\tilde{N}_2$ and $\tilde{N}_2^{-1}\tilde{N}_4\,\mathbf{1}$ are nonnegative (see Steps 6.1 and 6.2). Hence all entries of the product

$$\tilde{N}_1(\tilde{N}_2 + \tilde{N}_3)^{-1}\tilde{N}_2\tilde{N}_2^{-1}\tilde{N}_4\,\mathbf{1} = \tilde{N}_1(\tilde{N}_2 + \tilde{N}_3)^{-1}\tilde{N}_4\,\mathbf{1} = \tilde{N}^{-1}\mathbf{1}$$

are nonnegative.

7. Recall that

$$\alpha^*(t) = d\left(1 + \mathbf{1}^\top E(t)\tilde{N}^{-1}\mathbf{1}\right), \qquad t \in [0, T].$$

Since

$$E(t) = \operatorname{diag}\left(\frac{e^{\sqrt{c_i}T/2}\left(e^{\sqrt{c_i}(t-T/2)} + e^{\sqrt{c_i}(T/2-t)}\right)}{e^{\sqrt{c_i}T} - 1}\right), \qquad t \in [0, T],$$

conclude that there are $y_1, y_2, \dots, y_n \in \mathbb{R}$ for which

$$\begin{aligned}
\alpha^*(t) &= d\Big(1 + \sum_{i=1}^{n} y_i \big(e^{\sqrt{c_i}(t-T/2)} + e^{\sqrt{c_i}(T/2-t)}\big)\Big) \\
&= d\Big(1 + \sum_{i=1}^{n} y_i \sum_{k=0}^{\infty} \frac{\big(1+(-1)^k\big)\big(\sqrt{c_i}(t-T/2)\big)^k}{k!}\Big) \\
&= d\Big(1 + \sum_{k=0}^{\infty} \Big(\sum_{i=1}^{n} \frac{2 y_i c_i^k}{(2k)!}\Big)(t-T/2)^{2k}\Big).
\end{aligned}$$

Step 6.3 shows that $y_1, y_2, \dots, y_n \geq 0$. Recall that $d > 0$. Hence α^* is symmetrically totally monotone. □

Proof of Theorem 4.6, general case.
(i) Let $\gamma > 0$. Suppose first that $G(0) \coloneqq \lim_{t\to 0} G(t) < \infty$. Assume without loss of generality that $G(0) = 1$. By Bernstein's theorem, there is a Borel probability measure μ on $[0, \infty)$ such that G is the Laplace transform of μ, i.e., $G(t) = \int_{[0,\infty)} e^{-st} \mu(\mathrm{d}s)$ for $t \in [0, \infty)$.
The set of finite convex combinations of Dirac measures is dense in the set of all Borel probability measures on $[0, \infty)$ with respect to weak convergence. Hence there exists a sequence $(\mu_n)_{n\in\mathbb{N}}$ of such measures that converges weakly to μ. The corresponding Laplace transforms G_n of μ_n are all generalized exponential kernels of the form (4.8). Weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ to μ implies that $\lim_{t\to\infty} G_n(t) = G(t)$ for all $t \in [0, \infty)$.
For $n \in \mathbb{N}$ and $\alpha \in \mathcal{A}^1$, define

$$J_n^\gamma[\alpha] \coloneqq \frac{1}{2} \int_0^T \big(\gamma \alpha(t)^2 + \int_0^T G_n(|t-s|)\alpha(t)\alpha(s)\,\mathrm{d}s\big)\,\mathrm{d}t.$$

Notice that the restrictions to $[0, T]$ of all G_n and G are elements of $L^2[0, T]$. Let $\|\cdot\|$ denote the standard L^2-norm on $[0, T]$. Apply the Cauchy-Schwarz inequality twice to see that

$$\begin{aligned}
\big|J^\gamma[\alpha] - J_n^\gamma[\alpha]\big| &\leq \|\alpha\| \int_0^T |\alpha(t)| \Big(\int_0^T \big(G(|t-s|) - G_n(|t-s|)\big)^2\,\mathrm{d}s\Big)^{1/2}\,\mathrm{d}t \\
&\leq \|\alpha\|^2 \Big(\int_0^T \int_0^T \big(G(|t-s|) - G_n(|t-s|)\big)^2\,\mathrm{d}s\,\mathrm{d}t\Big)^{1/2} \\
&\leq \sqrt{2T}\,\|\alpha\|^2\,\|G - G_n\|.
\end{aligned}$$

By dominated convergence, $\|G - G_n\| \to 0$. Hence $J_n^\gamma[\alpha]$ converges to $J^\gamma[\alpha]$ uniformly in functions α from any bounded subset of $L^2[0, T]$.
For $n \in \mathbb{N}$, denote by α_n^* the minimizer of J_n^γ in $\mathcal{A}^1$. As shown in the first step of the proof, each α_n^* is symmetrically totally monotone. Since the constant function $\alpha(t) = 1/T$ belongs to $\mathcal{A}^1$, conclude that there is a constant C independent of n such that $\|\alpha_n^*\| \leq C$. By passing to a subsequence if necessary, assume without loss of

generality that the sequence $(\alpha_n^*)_{n\in\mathbb{N}}$ converges weakly in $L^2[0,T]$ to a function α^*. By Lemma 4.9, the function α^* is symmetrically totally monotone.
Choose an arbitrary $\alpha \in \mathcal{A}^1$. Then $J_n^\gamma[\alpha] \geq J_n^\gamma[\alpha_n^*]$ and by uniform convergence of the functionals J_n^γ and the lower semicontinuity of J^γ,

$$J^\gamma[\alpha] = \lim_{n\to\infty} J_n^\gamma[\alpha] \geq \liminf_{n\to\infty} J_n^\gamma[\alpha_n^*] = \liminf_{n\to\infty} J^\gamma[\alpha_n^*] \geq J^\gamma[\alpha^*].$$

This concludes the proof for $\gamma > 0$ and $G(0) < \infty$.
If $\gamma > 0$ and $G(0) = \infty$, approximate G via finite kernels G_n as in (4.7). As in the final part of the proof of Theorem 4.4, conclude that the symmetrically totally monotone minimizers for G_n converge weakly in $L^2[0,T]$ to the minimizer for G. Thus, the latter is also symmetrically totally monotone by Lemma 4.9.
(ii) Let $\gamma = 0$. Denote by X^* the minimizer of J^0 in $\mathcal{A}^0$. Approximate the probability measure $-\,\mathrm{d}X^*(t)$ in the weak topology by probability measures $\alpha_n(t)\,\mathrm{d}t$, where each α_n is a bounded nonnegative function on $[0,T]$ satisfying $\int_0^T \alpha_n(t)\,\mathrm{d}t = 1$. Then choose a sequence $(\gamma_n)_{n\in\mathbb{N}}$ of positive real numbers that decreases to zero and satisfies

$$\lim_{n\to\infty} \gamma_n \int_0^T \alpha_n(t)^2\,\mathrm{d}t = 0.$$

Conclude from (4.1) that $\lim_{n\to\infty} J^{\gamma_n}[\alpha_n] = J^0[X^*]$.
Let α_n^* denote the minimizer of J^{γ_n} in $\mathcal{A}^1$. By passing to a subsequence if necessary, assume without loss of generality that the probability measures $\mu_n(\mathrm{d}t) = \alpha_n^*(t)\,\mathrm{d}t$ on $[0,T]$ converge weakly to a probability measure μ on $[0,T]$. Conclude from the proof of Lemma 4.9 that the restriction of μ to $(0,T)$ is absolutely continuous with respect to the Lebesgue measure and admits a symmetrically totally monotone density. Define $X_n := 1 - \mu_n([0,t])$ and $X := 1 - \mu([0,t])$. Then all X_n and X are elements of $\mathcal{A}^0$. Obtain

$$J^0[X] = \lim_{n\to\infty} J^0[X_n] \leq \liminf_{n\to\infty} J^{\gamma_n}[\alpha_n^*] \leq \liminf_{n\to\infty} J^{\gamma_n}[\alpha_n] = J^0[X^*].$$

By uniqueness of the minimizer, $X = X^*$. □

Chapter 5

Outlook

This thesis studies optimal execution strategies in the presence of transient price impact and transaction costs. Open questions remain, both economic and mathematical. I would like to briefly discuss two promising directions for future research.

1. The Nash equilibria obtained in Chapters 2 and 3 are partial equilibria: While all strategic investors are aware of each other, price impact is completely mechanic, and random fluctuations in the asset price are implicitly attributed to non-strategic "noise traders". Almost all mathematical models of optimal execution are partial equilibrium models (Kyle, 1985, comes close to being an exception). In this respect, they provide a middle ground between models where the asset price is purely exogenous (such as the Black-Scholes model) and *full equilibrium* models where all market participants, including market makers and noise traders, behave strategically.

In my view, it would be worthwhile to study optimal execution in the context of a full equilibrium. This would merge research on optimal execution with research on optimal market making (see, e.g., Guéant, 2017, and the references therein) and efficient markets (see the discussion and references in the introduction of Bouchaud et al., 2004). A full equilibrium model would have to address two important issues: First, in what sense is the market equilibrium free of arbitrage and price manipulation? Second, different market participants—such as liquidating investors, arbitrage traders and market makers—are primarily characterized by differences in information. How can this be incorporated into the model, and what is the appropriate equilibrium concept?

One of the economic problems that could be studied in a full equilibrium model is the influence that non-designated market making has on market quality. Bershova and Rakhlin (2013) echo concerns that liquidity provided by non-designated market makers (e.g., by high frequency traders) could be "fictitious; although such liquidity is plentiful during 'normal' market conditions, it disappears at the first sign of trouble" (p. 3). To test this claim in a theoretical setting, it is necessary to model non-designated market makers who interact strategically with liquidity takers with different information sets.

2. One of the observations in Chapter 3 was that order anticipation strategies, which are typically associated with high frequency traders, do not necessarily cause *price overshooting* and in fact often reduce the price drop caused by a large sell order. Brunnermeier and Pedersen (2005) make a strong point about the danger of price overshooting. Linking opportunistic trading with increased (or reduced) price overshooting would serve as a strong argument for (or against) the harmfulness of high frequency trading. Theoretical research should determine for what types of price impact and transaction costs price overshooting occurs; and empirical research is necessary to study price overshooting in real financial markets.

Bibliography

Admati, A. R. and Pfleiderer, P. (1991). Sunshine trading and financial market equilibrium. *Review of Financial Studies*, 4(3):443–481.

Alfonsi, A., Fruth, A., and Schied, A. (2010). Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157.

Alfonsi, A. and Schied, A. (2013). Capacitary measures for completely monotone kernels via singular control. *SIAM Journal on Control and Optimization*, 51(2):1758–1780.

Alfonsi, A., Schied, A., and Slynko, A. (2012). Order book resilience, price manipulation, and the positive portfolio problem. *SIAM Journal on Financial Mathematics*, 3(1):511–533.

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

Almgren, R., Thum, C., Hauptmann, E., and Li, H. (2005). Equity market impact. *Risk*, 18(7):57–62.

Almgren, R. F. (2003). Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1):1–18.

Arfken, G. B. and Weber, H. J. (2005). *Mathematical Methods for Physicists*. Elsevier, 6th edition.

Atkinson, K. and Han, W. (2001). *Theoretical Numerical Analysis: A Functional Analysis Framework*. Springer.

Atkinson, K. E. (1997). *The Numerical Solution of Integral Equations of the Second Kind*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.

Avron, J. E. (2003). Non-local Lagrangian. `http://phsites.technion.ac.il/wp-content/uploads/sites/3/2013/07/non-local-lagrangian.pdf`, Version: 12.01.2003.

Benos, E. and Sagade, S. (2012). High-frequency trading behaviour and its impact on market quality: Evidence from the UK equity market. Working paper 469, Bank of England.

Berman, A. and Plemmons, R. J. (1994). *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial and Applied Mathematics.

Bernstein, S. (1914). Sur la définition et les propriétés des fonctions analytiques d'une variable réelle. *Mathematische Annalen*, 75(4):449–468.

Bernstein, S. (1929). Sur les fonctions absolument monotones. *Acta Mathematica*, 52(1):1–66.

Bershova, N. and Rakhlin, D. (2013). High-frequency trading and long-term investors: A view from the buy-side. *Journal of Investment Strategies*, 2(2):25–69.

Bertsimas, D. and Lo, A. W. (1998). Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50.

Bochner, S. (1932). *Vorlesungen über Fouriersche Integrale*. Akademische Verlagsgesellschaft.

Bouchaud, J.-P. (2010). Price impact. In *Encyclopedia of Quantitative Finance*. R. Cont (ed.), Wiley Online Library.

Bouchaud, J.-P., Farmer, J. D., and Lillo, F. (2009). How markets slowly digest changes in supply and demand. In *Handbook of Financial Markets: Dynamics and Evolution*. T. Hens and K. Schenk-Hoppe (eds.), Elsevier.

Bouchaud, J.-P., Gefen, Y., Potters, M., and Wyart, M. (2004). Fluctuations and response in financial markets: The subtle nature of 'random' price changes. *Quantitative Finance*, 4(2):176–190.

Brunnermeier, M. K. and Pedersen, L. H. (2005). Predatory trading. *The Journal of Finance*, 60(4):1825–1863.

Cardaliaguet, P. and Lehalle, C.-A. (2017). Mean field game of controls and an application to trade crowding. Working paper (arXiv:1610.09904v2).

Carlin, B. I., Lobo, M. S., and Viswanathan, S. (2007). Episodic liquidity crises: Cooperative and predatory trading. *The Journal of Finance*, 62(5):2235–2274.

Carmona, R. A. and Yang, J. (2011). Predatory trading: A game on volatility and liquidity. Quantitative Finance, forthcoming.

Cartea, Á. and Jaimungal, S. (2015). A closed-form execution strategy to target VWAP. Working paper (SSRN 2542314).

Choi, J. H., Larsen, K., and Seppi, D. J. (2015). Information and trading targets in a dynamic market equilibrium. Working paper (arXiv:1502.02083v3).

Curato, G., Gatheral, J., and Lillo, F. (2017). Optimal execution with non-linear transient market impact. *Quantitative Finance*, 17(1):41–54.

El-Mikkawy, M. E. A. (2004). On the inverse of a general tridiagonal matrix. *Applied Mathematics and Computation*, 150(3):669–679.

Gatheral, J. (2010). No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7):749–759.

Gatheral, J., Schied, A., and Slynko, A. (2012). Transient linear price impact and Fredholm integral equations. *Mathematical Finance*, 22(3):445–474.

Gelfand, I. M. and Fomin, S. V. (1963). *Calculus of Variations.* Prentice-Hall.

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

Guéant, O. (2017). Optimal market making. Working paper (arXiv:1605.01862v5).

Hirschey, N. (2016). Do high-frequency traders anticipate buying and selling pressure? Working paper (SSRN 2238516).

Ho, T. and Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1):47–73.

Horn, R. A. and Johnson, C. R. (2013). *Matrix Analysis.* Cambridge University Press, 2nd edition.

Huang, X., Jaimungal, S., and Nourian, M. (2017). Mean-field game strategies for optimal execution. Working paper (SSRN 2578733).

Huberman, G. and Stanzl, W. (2004). Price manipulation and quasi-arbitrage. *Econometrica*, 72(4):1247–1275.

Huberman, G. and Stanzl, W. (2005). Optimal liquidity trading. *Review of Finance*, 9(2):165–200.

Kac, M., Murdock, W. L., and Szegö, G. (1953). On the eigen-values of certain Hermitian forms. *Journal of Rational Mechanics and Analysis*, 2:767–800.

Kahane, C. (1965). Analyticity of solutions of mildly singular integral equations. *Communications on Pure and Applied Mathematics*, 18(4):593–626.

Kaper, H. G. and Kellogg, R. B. (1977). Asymptotic behavior of the solution of the integral transport equation in slab geometry. *SIAM Journal on Applied Mathematics*, 32(1):191–200.

Kelley, W. G. and Peterson, A. C. (1991). *Difference Equations. An Introduction with Applications.* Academic Press.

Kissell, R., Glantz, M., and Malamut, R. (2004). A practical framework for estimating transaction costs and developing optimal trading strategies to achieve best execution. *Finance Research Letters*, 1(1):35–46.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335.

Lachapelle, A., Lasry, J.-M., Lehalle, C.-A., and Lions, P.-L. (2016). Efficiency of the price formation process in presence of high frequency participants: A mean field game analysis. *Mathematics and Financial Economics*, 10(3):223–262.

Lorenz, C. and Schied, A. (2013). Drift dependence of optimal trade execution strategies under transient price impact. *Finance and Stochastics*, 17(4):743–770.

Love, E. R. (1949). The electrostatic field of two equal circular co-axial conducting disks. *The Quarterly Journal of Mechanics and Applied Mathematics*, 2(4):428–451.

MacKenzie, D. (2011). How to make money in microseconds. *London Review of Books*, 33(10):16–18.

Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London, Series A*, 209:415–446.

Milne-Thomson, L. M. (2000). *The Calculus of Finite Differences*. American Mathematical Society.

Moallemi, C. C., Park, B., and Van Roy, B. (2012). Strategic execution in the presence of an uninformed arbitrageur. *Journal of Financial Markets*, 15(4):361–391.

Nyström, E. J. (1930). Über die praktische Auflösung von Integralgleichungen mit Anwendungen auf Randwertaufgaben. *Acta Mathematica*, 54(1):185–204.

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

Oehmke, M. (2014). Liquidating illiquid collateral. *Journal of Economic Theory*, 149(1):183–210.

Petrov, F. (2017). Inequality connected to Lagrange's interpolation formula. MathOverflow contribution (`http://mathoverflow.net/q/263277`, Version: 28.02.2017).

Pitkäranta, J. (1980). Estimates for the derivatives of solutions to weakly singular Fredholm integral equations. *SIAM Journal on Mathematical Analysis*, 11(6):952–968.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton Mathematical Series, No. 28. Princeton University Press.

Schechter, S. (1959). On the inversion of certain matrices. *Mathematical Tables and Other Aids to Computation*, 13(66):73–77.

Schied, A. and Strehle, E. (2017). On the minimizers of energy forms with completely monotone kernel. Working paper (arXiv:1706.04844v4).

Schied, A., Strehle, E., and Zhang, T. (2017). High-frequency limit of Nash equilibria in a market impact game with transient price impact. To appear in *SIAM Journal on Financial Mathematics*.

Schied, A. and Zhang, T. (2015). A state-constrained differential game arising in optimal portfolio liquidation. *Mathematical Finance*, 27(3):779–802.

Schied, A. and Zhang, T. (2017). A market impact game under transient price impact. Working paper (arXiv:1305.4013v7).

Schöneborn, T. (2008). *Trade Execution in Illiquid Markets: Optimal Stochastic Control and Multi-Agent Equilibria*. Technische Universität Berlin, doctoral thesis.

Schöneborn, T. and Schied, A. (2009). Liquidation in the face of adversity: Stealth vs. sunshine trading. In *EFA 2008 Athens Meetings*.

Securities and Exchange Commission (2010). Concept release on equity market structure (17 CFT Part 242). *Federal Register*, 75(13):3594–3614.

Strehle, E. (2017). Are order anticipation strategies harmful? A theoretical approach. Working paper (arXiv:1609.00599v3).

Terrell, B. (2017). How to show that this polynomial has n positive roots? Math StackExchange contribution (`http://math.stackexchange.com/q/2106024`, Version: 20.01.2017).

Tong, L. (2015). A blessing or a curse? The impact of high frequency trading on institutional investors. In *The Impact of High Frequency Trading on Institutional Investors (October 5, 2015). European Finance Association Annual Meetings*.

Usmani, R. A. (1994a). Inversion of a tridiagonal Jacobi matrix. *Linear Algebra and its Applications*, 212:413–414.

Usmani, R. A. (1994b). Inversion of Jacobi's tridiagonal matrix. *Computers & Mathematics with Applications*, 27(8):59–66.

Vainikko, G. (2006). *Multidimensional Weakly Singular Integral Equations*. Springer.

Wazwaz, A.-M. (2011). *Linear and Nonlinear Integral Equations: Methods and Applications*. Springer.

Yeung, D. W. K. and Petrosjan, L. A. (2006). *Cooperative Stochastic Differential Games*. Springer.

Young, W. H. (1913). On the Fourier series of bounded functions. *Proceedings of the London Mathematical Society*, 2(1):41–70.

Zhang, T. (2014). *Nash Equilibria in Market Impact Models: Differential Game, Transient Price Impact and Transaction Costs*. University of Mannheim, doctoral thesis.