Optimal Trading with Multiplicative Transient Price Impact for Non-Stochastic or Stochastic Liquidity

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Dipl.-Math. Peter Frentrup

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:

1. Prof. Dr. Dirk Becherer
2. Prof. Dr. Giorgio Ferrari
3. Prof. Dr. Jan Kallsen

Eingereicht am: 14. 5. 2019

Tag der Verteidigung: 10. 9. 2019

Abstract

In this thesis, we study a class of multiplicative price impact models for trading a single risky asset. When a large investor trades in a risky asset, her actions have adverse effect on the price at which trading happens in the market, depressing the price when she sells and increasing it when she buys. We model price impact to be multiplicative so that prices are guaranteed to stay non-negative, a feature that additive impact models, which are often used in the optimal liquidation literature, lack. Our risk-neutral large investor seeks to maximize expected gains from trading.

We first introduce a basic variant of our model, wherein the transient impact is a deterministic functional of the trading strategy. We draw the connection to limit order books and give the optimal strategy to liquidate or acquire an asset position in an a priori infinite time horizon. Building on these results for an unconstrained time horizon, we subsequently introduce a clearing condition at a fixed horizon. We solve the corresponding optimization problem in a two step manner. Calculus of variations allows to identify the free boundary surface that separates buy and sell regions and moreover show its local optimality, which is a crucial ingredient for the verification giving (global) optimality. This result allows us to conduct a qualitative comparison to the according additive impact variant.

In the second part of the thesis, we introduce uncertainty about the actual price impact by adding stochasticity to the auxiliary impact process. This feature causes optimal strategies to dynamically adapt to random changes in liquidity. Again focusing on the infinite time horizon case, we identify the optimal liquidation strategy as the reflection local time which keeps the market impact process below some non-constant free boundary level. We describe this free boundary curve explicitly in terms of an integral equation. Similar to the previous chapter the proof technique involves a combination of classical calculus of variations and direct methods, showing first a local optimality result and then augmenting it to a global one.

In order to again impose a constraint on the time horizon, allowing for intermediate buy actions without further transaction costs, we need to enlarge the set of admissible controls to include semimartingales. We address the issue with an in-depth analysis of the stability of the proceeds functional with respect to the trading strategy in a broad class of market models including multiplicative and additive price impact, with deterministic or stochastic liquidity. Skorokhod's M_1 topology is key to extend the class of admissible controls from finite variation to general càdlàg strategies.

Subsequently, we return to the problem of optimal liquidation and introduce proportional transaction costs into our stochastic liquidity model. We solve the related one-dimensional free boundary problem of optimal trading without constraints on time or asset position and highlight possible solution methods for the corresponding liquidation problem where trading stops as soon as all assets are sold.

With the last chapter we depart from optimal control problems. Inspired by the reflection local time nature of the optimal liquidation strategy for non-deterministic impact, we develop an approximation scheme for diffusions with reflection at an elastic boundary which is a function of the reflection local time. This leads to a probabilistic functional limit result and naturally gives an explicit expression for the Laplace transform of the inverse local time.

Zusammenfassung

Diese Arbeit untersucht eine Reihe multiplikativer Preiseinflussmodelle für das Handeln in einer riskanten Anlage. Handelt ein Großinvestor mit riskanten Finanzanlagen, so beeinflusst er den entsprechenden Preis in einer für ihn ungünstigen Weise; Preise fallen, wenn er verkauft, und steigen, sobald er kauft. Wir modellieren diesen Preiseinfluss multiplikativ, um die Möglichkeit negativer Preise auszuschließen, ein Merkmal, dass additiven Preiseinflussmodellen, wie sie häufig in der finanzmathematischen Literatur zur optimalen Portfolioliquidierung vorkommen, fehlt. Unser risikoneutraler Investor versucht seine zu erwartenden Handelserlöse zu maximieren.

Wir beginnen mit einer einfachen Variante unseres Modells, in der der vorübergehende Preiseinfluss ein deterministisches Funktional der Handelsstrategie darstellt. Wir stellen den Zusammenhang mit Limit-Orderbüchern her und besprechen die optimale Strategie zur Liquidierung bzw. zum Aufbau einer Anlageposition bei a priori unbeschränkem Anlagehorizont. Aufbauend auf diesen Resultaten für den zeitunbeschränkten Fall führen wir im Anschluss eine feste Zeitschranke zur Liquidierung ein. Das daraus entstehende Optimierungsproblem lösen wir in zwei Schritten. Mittels Variationsrechnung lässt sich die freie Grenzefläche, welche Kauf- und Verkaufsregionen trennt, als lokales Optimum identifizieren. Diese lokale Optimalität ist entscheidend für die Verifikation globaler Optimalität. Dieses Resultat erlaubt einen qualitativen Vergleich mit der entsprechenden additiven Preiseinflussvariante.

Im zweiten Teil der Arbeit führen wir Unsicherheit bezüglich des tatsächlichen Preiseinflusses ein, indem wir den zwischengeschalteten Markteinflussprozess um eine stochastische Komponente erweitern. Dies bedingt, dass optimale Strategien dynamisch an zufällige Liquiditätsschwankungen adaptieren. Wir behandeln zunächst wieder den zeitunbeschränkten Fall und bestimmen die optimale Liquidierungsstrategie als die reflektierende Lokalzeit, die den Markteinflussprozess unterhalb eines nicht-konstanten freien Grenzlevels hält. Diese Grenzkurve beschreiben wir explizit über eine Integralgleichung. Wie im vorherigen Kapitel umfasst der Beweis wieder eine Kombination aus Variationsrechnung, um zunächst lokale Optimalität zu zeigen, und direkten Methoden, die diese dann zu globaler Optimalität erweitern.

Um erneut eine Beschränkung des Anlagehorizonts und zwischenzeitliches Kaufen ohne zusätzliche Transaktionskosten zu ermöglichen, ist es nötig, die Klasse der zulässigen Strategien um Semimartingale zu erweitern. Dazu betreiben wir eine detailierte Analyse zur Stabilität des Erlösfunktionals bezüglich der Handelsstrategie in einer umfangreichen Klasse von Preiseinflussmodellen, welche sowohl additiven, als auch multiplikativen Preiseinfluss umfasst, mit deterministischer oder stochastischer Liquidität. Skorochods M_1-Toplogie erweist sich als Schlüsselelement, um die Klasse der möglichen Strategien von endlichen Variations- auf allgemeine rechtsstegige Strategien mit linksseitigen Limiten zu erweitern.

Zurück beim optimalen Liquidierungsproblem führen wir anschließend proportionale Transaktionskosten in unser stochastisches Preiseinflussmodell ein. Wir lösen das entsprechende eindimensionale freie Grenzproblem des optimalen Handels ohne Zeit- oder Anlagepositionsbeschränkungen und beleuchten mögliche Lösungsansätze für das Liquidierungsproblem, welches mit dem Verkauf der letzten Anleihe endet.

Mit dem letzten Kapitel entfernen wir uns von optimalen Kontrollproblemen. Inspiriert durch die Struktur der optimalen Strategie bei stochastischer Liquidität entwickeln wir ein Approximationsschema für Diffusionen mit Reflexion an einer elastischen Grenze, die als Funktion der reflektierenden Lokalzeit darstellbar ist. Dies führt zu einem funktionalem Grenzwertresultat und liefert auf natürliche Weise einen explizite Ausdruck für die Laplace-Transformation der inversen Lokalzeit.

Acknowledgment

First and foremost, I want to express my deep gratitude to Dirk Becherer for his enduring support and fruitful discussions. He introduced me to the field of stochastic analysis, supervised my diploma thesis, and guided me through my Ph.D. studies. I especially appreciate his encouragement to present our research at many conferences, workshops, and seminars.

I would like to thank Giorgio Ferrari and Jan Kallsen for agreeing to co-examine this thesis. The discussions with Giorgio and his suggestions proved invaluable to bring forward parts of this thesis.

I owe special thanks to Todor Bilarev. We published four papers together. It is his precise comments and explanations that made these results possible. Let me thank all friends and colleges at Humboldt University who accompanied my research and teaching.

To Helena.

Contents

1 Introduction

A large trader seeks to liquidate her position in a risky financial asset. Being large means that her trading actions have adverse impact on the asset's price, with the effect that selling a large amount of assets in short time will depress the price and, conversely, buying a large amount will cause increasing prices. As a consequence, the large trader needs to split large trades into smaller chunks to be executed over time, balancing the losses caused by such price movements against her preference to finish the trade early. One explanation for such adverse price effect might be that other market participants observe her actions, learn or anticipate that she would continue trading in the same direction for the near future and try to exploit this knowledge. A different explanation that does not require modeling competitive agents builds on the structure of the market in form of a limit order book. A limit order book (LOB) lists the amount of assets available for sale or buy at each particular price. These quantities correspond to (not yet executed) limit orders. An incoming limit order is placed in the order book unless it can be executed immediately. In contrast, an incoming market order is executed directly by "eating" into the order book and thereby moving the price, cf. Remark 2.1.1. From this limit order book perspective, our large investor issues only market orders, but no limit orders. Not having to wait for later execution of placed (limit) orders means that there is a one-to-one correspondence between her trading strategy and her asset position at each time point, so we can identify these two notions.

Our large investor is risk-neutral, she merely aims at maximizing expected proceeds from trading. We will solve her optimal liquidation problem (and the related problem of optimally acquiring an asset) in infinite and finite time horizons, when she either expects fundamental prices to have no trend, to rise, or to decline in expectation. Since seminal work by [BL98, AC00], optimal execution problems have been a subject of extensive research. We mention [OW13, AFS10, KP10, ASS12, FKTW12, LS13, Kat14, GH17] and refer to [PSS11, GS13] for further references and application background.

Unlike electricity or other goods that involve storage cost, financial assets such as options or stocks do not have negative prices. As already noted by [Sam65], forgetting this property may lead to anomalous and counter-intuitive results. Nonetheless, in the field of mathematical finance most literature on optimal execution with market impact considers additive impact models, wherein the price at which trading occurs is the sum of an unperturbed fundamental price and a function of an auxiliary market impact level. Such modeling causes negative prices with (small) positive probability, which means that these additive impact models are to be used for short time horizons only, where the probability of reaching negative prices may be negligible, see e.g. [GSS12, footnote 3]. Conceptual difficulties would thus arise for applications with longer time horizons, as they can occur e.g. for large institutional trades [CL95, KMS17], or for hedging problems with longer maturities. Furthermore, additive impact models are better suited to an arithmetic Brownian motion specification for the fundamental price, as in the Bachelier model, and may cause modeling artifacts when combined with an e.g. a geometric Brownian motion, as in the Black-Scholes model, for the fundamental price process, cf. Remark 3.2.7.

To remedy the possible occurrence of negative prices, alternative price impact models were proposed early on in mathematical finance and economics literature, cf. e.g. [Jar94, Fre98, BL98, FJ02, FGL$^+$04, HM05, FKTW12, Kat14]. As noted in [GS13, Sect. 3.2], computing optimal strategies in such models may be more difficult than in additive price impact models. This is certainly a reason for the wide adoption of the latter in the literature. One goal of this thesis is to demonstrate how to attain explicit analytical solutions in a multiplicative impact model and thereby help to fill the gap between (analytically tractable) additive and more plausible multiplicative modeling of price impact. To be more precise, we consider multiplicative impact similar to [Løk12, GZ15], i.e. additive impact on the returns. Herein, the actual price at which trading occurs is the product of a fundamental price and a function of the market impact level which, in turn, incorporates past and present trades. In order to investigate the role of resiliency in the market, we take this persistent effect on price to be transient instead of permanent. It turns out that transience is essential for the optimal liquidation results, whereas [GZ15] have shown that purely permanent impact would lead to trivial results, cf. Remark 2.3.4. Mathematically, our setup of transience implies that the state variable $(Y_t)_{t\geq 0}$, that represents impact, cannot be decoupled from the control strategy $(\Theta_t)_{t\geq 0}$. It is this strong dependence that makes the (non-convex) optimization problems we study challenging. This feature is similar to (and in fact tightly connected to, as we will see in Chapter 4) the difference between a reflected Ornstein-Uhlenbeck process and a reflected Brownian motion. While Skorokhod's lemma applies for the latter, giving that it is the difference of an unreflected diffusion and the reflection process (its running maximum), such direct approach is not available in case of a state-dependent drift as it appears for the Ornstein-Uhlenbeck case. We examine such reflected diffusions in Chapter 7.

In Chapter 2, we elaborate the basic ingredients of our multiplicative market impact model and present the main results on optimal liquidation in infinite time horizon with deterministic impact dynamics. Since we consider continuous time trading, our optimization objective has the form of a singular stochastic control *monotone follower* problem of *finite fuel* type (the term finite fuel dates back to [BC67], see e.g. [Kar85], or [BE08] and the references therein for a more recent treatment). We express the optimal trading strategy in feedback form through an explicit formula for the free boundary separating sell- from no-sell-regions. One outcome of our analysis is that – assuming initial impact is small, so that prices have (almost) no upward trend – intermediate buying is sub-optimal for the large investor.

Whereas an infinite time horizon allows for easier analysis since the state space is two-dimensional, this setup requires discounting and a sign constraint on admissible strategies. By introducing an exogenous restriction on the time to liquidation in Chapter 3, we are able to allow for short sales and to incorporate beliefs of the large trader about short-time evolution of prices (upward or downward trend) at the cost of increasing the state space to three dimensions. The optimal trading strategy in this non-concave maximization problem (cf. Remark 2.3.1) is characterized by a non-constant free boundary surface between sale- and no-sale regions that we construct explicitly.

Up to Chapter 3, we follow the majority of optimal execution literature with price impact in that we take the inter-temporal impact to be a deterministic function of the single large trader's strategy, thereby effectively keeping all aspects of market liquidity static. When all relevant market characteristics are deterministic functions of the large investor's activity, the optimal strategy will naturally be deterministic, unless one introduces model artifacts by e.g. mixing additive and multiplicative dynamics (such as additive impact with geometric Brownian motion for the fundamental price; or multiplicative impact with additive bid-ask spread). Working in an essentially deterministic setup greatly simplifies the analysis, but also limits the richness of possible results. On longer time-horizons, one would instead expect some liquidity aspects to vary stochastically. In Chapter 4, we incorporate such effects by introducing own stochasticity in the large trader's controlled market impact process. In order to retain a tractable problem, we simplify other aspects like the specific form of resilience and again concentrate on the infinite horizon problem without intermediate buying. The optimal liquidation strategy in this setup turns out to be the (non-deterministic) reflection local time which keeps the impact process below some varying critical level that we describe explicitly as a function of the current position (the non-constant free boundary curve). The original model from Chapter 2 can be understood as an approximation for the our impact model with stochastic liquidity, cf. Section 4.7.

This result motivates a more in-depth analysis of diffusions that are reflected at a non-constant boundary which in turn depends on the reflection local time (in a sense of "how often the boundary was hit before") in Chapter 7, where we develop this intuitive perception of a boundary that retreats with every hit into an approximation scheme. See e.g. [DI93] for general reflected diffusions. As a byproduct, this analysis gives a probabilistic proof for the explicit formula of the Laplace transform of the inverse local time (cf. Theorem 7.2.2), for which we also have an analytic proof (see Theorem 4.3.2) in a slightly less general setup of strictly increasing reflection boundary functions.

The initial problem in Chapter 2 is classic in the sense that a usual "guess and verify" approach can be carried out, wherein one first identifies a candidate value function, e.g. by assuming a smooth fit condition, and then performs a direct verification of the variational inequalities. In contrast, direct verification seems out of reach in Chapters 3 and 4; at least I can not see any convexity structure, that would help. Instead, there we divide the verification procedure into two manageable parts. Using sufficiency conditions available from calculus of variations methods, we first obtain a local optimality result in a class of strategies that can be described by smooth boundary curves (Theorems 3.3.8 and 4.4.6, respectively). This local result causes the variational inequalities to hold in a neighborhood of the candidate boundary (cf. Lemma 3.3.10 for Chapter 3 and the second part of the proof of Lemma 4.5.7 for Chapter 4). Finally, the proof of the variational inequalities in the whole state space reduces to their validity in a neighborhood of the candidate boundary.

One aspect, that we initially take as given and explain only heuristically through a limit order book interpretation, is the concrete objective functional which represents the cumulative proceeds of a trading strategy. Of particular importance are the trading gains from large block orders, because these are associated with two natural prices – immediately before and after the trade. Defining these gains in an ad-hoc manner may lead to surprising consequences, such as the large investor being able to completely circumvent her price impact – an undesirable outcome for a model that tries to explain market impact effects – cf. Example 5.2.2. Chapter 5 provides a thorough analysis of the subject for a broad class of impact models. Since we deal with mechanic price impact and disregard strategic considerations, fast trading in small blocks, or even continuously, should lead to similar proceeds as trading the same amount in one large block. Skorokhod's M_1 topology is the correct choice on the space of càdlàg paths to encode this desired property, because it has the feature that continuous functions form a dense subset, unlike in the J_1 and uniform topologies. Equipped with the M_1 topology, we find the unique continuous extension of the proceeds functional from continuous finite variation strategies to general (predictable) càdlàg strategies. As an example, this extension allows to again restrict the time horizon through an expectation constraint in the stochastic resilience setup of Chapter 4 and consider non-monotone strategies, see Section 5.3. There, a convexity argument in the spirit of [PSS11] gives that optimal strategies are of infinite variation and such that the impact process remains at a fixed level until a terminal block trade.

When considering non-monotone strategies, the question of how to accommodate for bid-ask spread becomes important. For a deterministic resilience specification and natural parameter choices, the difference between bid and ask prices is irrelevant, because the large investor would still just implement a monotone strategy, cf. Remark 2.3.6. In comparison, with stochastic resilience, where the large investor cannot fully control her impact on price, the optimizer among non-monotone strategies cannot be expected to be monotone. With zero spread, it would even be of infinite variation, see Section 5.3. In Chapter 6, we provide an outlook of how to incorporate proportional transaction costs into our stochastic resilience model of Chapter 4. We perform a preliminary analysis of the optimal liquidation problem in this extended market model with infinite time horizon in Section 6.1. The optimal strategy should be characterized by two free boundary curves in $\mathbb{R}^2$ separating the wait region from buy and sell regions, which we can describe quite explicitly through a system of ordinary differential equations with boundary condition at infinity. We solve explicitly the *infinite fuel* variant of optimal trading indefinitely without liquidation constraint in Section 6.2. Informally, this (one-dimensional) infinite fuel problem can be understood as the limiting case of the (two-dimensional) *finite fuel* liquidation problem. Such connections between finite and infinite fuel problems are known in the singular control literature, see e.g. [KOWZ00]. While the infinite fuel variant is amenable to a smooth pasting and direct verification combination, verification (and even existence of the candidate optimizer described by an ODE, cf. Remark 6.1.6) for the finite fuel case remains an open problem. For comparison, the calculus of variations ansatz in Chapter 4 to get local optimality crucially depends on monotonicity of the strategies, cf. equation (4.16).

We proceed with a more concrete outline of each chapter.

A deterministic price impact model for optimal liquidation (Chapter 2)

This chapter introduces the overall objective and notation for the thesis. It considers the optimal execution problem for a large trader in an illiquid financial market, who aims to sell (or buy, cf. Remark 2.2.11) a given amount of a risky asset while her price impact is a deterministic functional of her trading strategy. We present explicit solutions for the optimal control and the related free boundary. Serving more as an extended introduction, we defer proofs to the article [BBF18b] on which this chapter is based.

Since orders of the large trader have an adverse impact on the prices at which they are executed, she needs to balance the incurred liquidity costs against her preference to complete a trade early. Posing the problem in continuous time leads to a singular stochastic control problem of finite fuel type. We note that our control objective, see (2.8)–(2.9), involves control cost terms like in [Tak97, DZ98, DM04], depending explicitly on the state process $(\overline{S}, Y)$ with a summation of integrals for each jump in the control strategy Θ. We refer to these articles for more background on singular stochastic control. The articles [Tak97, DM04] show general results on existence for optimal singular controls; explicit descriptions of those can be obtained only for special problems, see e.g. [KS86, Kob93, DZ98], but these examples differ from the one considered here in several aspects. In particular, their setups are such that integrator Θ and auxiliary process Y for the integrand can be decoupled in the objective functional.

The multiplicative limit order book model we investigate here is closely related to the additive limit order book models of [PSS11, AFS10, OW13, LS13], a key difference being that the price impact of orders is multiplicative instead of additive. In absence of large trader activity, the risky asset price follows some unaffected non-negative price evolution $\overline{S} = (\overline{S}_t)$, for instance geometric Brownian motion. The trading strategy (Θ_t) of the large trader has a multiplicative impact on the actual asset price which is evolving as $S_t = \overline{S}_t f(Y_t)$, $t \geq 0$, for a process Y that describes the level of market impact. This process is defined by a mean-reverting differential equation $\mathrm{d}Y_t = -h(Y_t)\,\mathrm{d}t + \mathrm{d}\Theta_t$, which is driven by the amount Θ_t of risky assets held, and can be interpreted as a volume effect process like in [PSS11, AFS10], see Remark 2.1.1. Subject to suitable properties for the functions f, h (see Assumption 2.2.2), asset sales (buys) are depressing (increasing) the level of market impact Y_t and thereby the actual price S_t in a transient way, with some finite rate of resilience. For f being positive, multiplicative price impact ensures that risky asset prices S_t are positive, like in the continuous-time variant [GS13, Sect. 3.2] of the model in [BL98]. We admit for general non-linear impact functions f, corresponding to general density shapes of a multiplicative limit order book whose shapes are specified with respect to relative price perturbations $S/\overline{S}$, and depth of the order book could be infinite or finite, cf. Remark 2.1.1. The rate of resilience $h(Y_t)/Y_t$ may be non-constant and (unaffected) transient recovery of Y_t could be non-exponential, while the problem still remains Markovian in $(\overline{S}, Y)$ through Y, like in [PSS11] but differently to [AFS10, LS13]. Following [PSS11, GZ15], we admit for general (monotone) bounded variation strategies in continuous time, while [AFS10, KP10] consider trading at discrete times.

Most of the related literature, like [AFS10, PSS11], on transient additive price impact assumes that the unaffected (discounted) price dynamics exhibit no drift, and such a martingale property allows for different arguments in the analysis. Without drift, a convexity argument as in [PSS11] can be applied readily also for multiplicative impact to identify the optimal control in the finite horizon problem with a free boundary that is constant in one coordinate, see Remark 2.2.10. [Løk12] has shown how a multiplicative limit order book (cf. Remark 2.1.1) could be transformed into an additive one with further intricate dependencies, to which the result by [PSS11] may be applied. For additive impact, [LS13] investigate the problem with general drift for finite horizon, while in this chapter we derive explicit solutions for multiplicative impact, infinite horizon and negative drift (we will extend our setup and preform a qualitative comparison with the [LS13] findings later in Chapter 3). The interesting articles [KP10, FKTW12, GZ15] also solve optimal trade execution problems in a model with multiplicative instead of additive price impact, but models and results differ in key aspects. The article [GZ15] considers permanent price impact, non-zero bid-ask spread (proportional transaction costs) and a particular exponential parametrization for price impact from block trades, whereas we study transient price impact, general impact functions f, and zero spread (in Section 2.3). Numerical solutions of the Hamilton-Jacobi-Bellman equation derived by heuristic arguments are investigated in [FKTW12] for a different optimal execution problem on finite horizon in a Black-Scholes model with permanent multiplicative impact. The authors of [KP10] obtain viscosity solutions and their nonlinear transient price impact is a functional of the present order size and the time lag from (only) the last trade, while we consider impact which depends via Y on the times and sizes of all past orders, as in [PSS11].

We obtain explicit solutions for two variants of the optimal liquidation problem to maximize expected discounted liquidation proceeds over an infinite time horizon, in a model with multiplicative price impact and drift that is introduced in Section 2.1. In the first variant (I), whose solution is presented in Section 2.2, the large trader may only sell but is not permitted to buy, whereas for the second variant (II) in Section 2.3 intermediate buying is admitted, even though the trader ultimately wants to liquidate her position. Variant I may be of interest, if a bank selling a large position on behalf of a client is required by regulation to execute only sell orders. The second variant might fit for an investor trading for herself and is mathematically needed to explore, whether a multiplicative limit order book model admits profitable round trips or transaction triggered price manipulations, as studied by [ASS12] for additive impact, see Remark 2.3.5. Notably, the free boundaries coincide for both variants, and the time to complete liquidation is finite, varies continuously with the discounting parameter (i.e. the investor's impatience) and tends to zero for increasing impatience in suitable parametrizations.

Optimal execution with price trends – a three-dimensional free boundary problem (Chapter 3)

Our large investor who faces the problem of maximizing expected proceeds from liquidating a risky asset position has some beliefs about price trends, meaning that the (unaffected) fundamental price process $\overline{S}$ is not necessarily a martingale, but may have an increasing or decreasing drift component. She needs to clear her position in a given finite time horizon (even when she expects prices to increase), while her trading activities cause adverse transient impact on the asset price.

We consider a finite time horizon $[0, T]$ in this chapter, so that the state space is three-dimensional – involving time to liquidation, current impact level, and current asset position. Consequently, we assume slightly more regularity on the impact function f and resilience function h than in Chapter 2, cf. Assumption 3.1.1.

Like [AFS10, PSS11], most of the related literature on transient price impact models assumes the fundamental prices process $\overline{S}$ to be a martingale. This ansatz can be of great help for the solution and verification by allowing a richer set of mathematical tools to be applied, like convexity arguments, as explained in Remark 2.2.10 for our multiplicative impact model. When we interpret the large investor's measure $\mathbb{P}$, with respect to which she maximizes her expected trading gains, as a proxy for her prospects about market dynamics, it is reasonable to take the (unaffected) fundamental price $\overline{S}$ to not be a martingale under $\mathbb{P}$, but to expose some drift or price trend. The findings of [LS13] on optimal execution in finite time horizon with additive impact and general drift specifications allows us to directly compare additive and multiplicative market impact models in Remark 3.2.7 and hint on their respective advantages and disadvantages.

Imposing a drift term on the fundamental price subverts the convexity approach à la [PSS11] of Remark 2.2.10. We construct the optimal trading strategy (Theorem 3.2.1) in two steps. First, classical calculus of variations methods, for which we refer to [GF00], yield a candidate for the (non-constant) free boundary surface that separates the three-dimensional state space of impact, asset position and remaining time into buy and sell regions. On the one hand, classical calculus of variations deals with absolutely continuous controls only. On the other hand, it provides us with sufficient conditions to prove local optimality of the candidate solution in this restricted set of controls in Theorem 3.3.8 via the second variation. Now, local optimality among absolutely continuous strategies implies that the corresponding value function necessarily satisfies the Hamilton-Jacobi-Bellman equation in a neighborhood of the (candidate) boundary surface. This observation is essential for the second step of expanding the region in which the variational inequality holds to the whole state space and thus proving global optimality among the larger set of bounded variation controls.

When prices are generally increasing, it may be optimal to start buying (first en bloc, then in rates) and only begin to sell later on, cf. Remark 3.2.4 and Figure 3.1a. When there is no drift in the (unaffected) fundamental price process, optimal trading happens at constant rate (apart from initial and terminal block trades). In contrast, with decreasing prices, it is generally optimal to start selling and to possibly go short.

As a variant of the optimization problem, we also consider the case when the large investor is not allowed to place buy orders, but can only wait or sell. We solve this variant for the case of generally decreasing prices. A naïve ansatz would be to follow the optimal strategy from the infinite horizon problem from Theorem 2.2.4 as long as possible and then finish early with a terminal block trade. Such a strategy would be optimal only in special cases like [Kar85] where the controlled diffusion is a controlled Brownian motion. Here, such a strategy is suboptimal, cf. Remark 3.4.3, but suitably combined with the optimizer from Theorem 3.2.1 where buying is allowed, it forms a building block of the optimal sell-only strategy in finite horizon, see Theorem 3.4.1.

Optimal liquidation under stochastic liquidity (Chapter 4)

In the majority of literature on price impact models the inter-temporal impact is typically a deterministic function of the strategy of the (single) large trader. In reality, we would rather expect some aspects of market liquidity (where [Kyl85] has distinguished resilience, depth and tightness) to vary stochastically over time, and a sophisticated trader to adapt her optimal strategy accordingly. Even for the extensively studied problem of optimal liquidation, there are relatively few recent articles on models in continuous time where the optimal liquidation strategy is adaptive to random changes in liquidity, cf. [Alm12, LS13, FSU19, GHS16, GH17].

We consider a model where temporary market imbalances involve *own stochasticity*. Price impact is transient, i.e. it could be persistent but eventually vanishes over time. Moreover, it is non-linear, corresponds to a general shape for the density of the limit order book as in Chapter 2, and is multiplicative to ensure positive risky asset prices. More precisely, our price process $S = (S_t)_{t\geq 0} = (f(Y_t)\overline{S}_t)_{t\geq 0}$ observed in the market deviates by a positive factor $f(Y_t)$ from the fundamental price $\overline{S}_t$ that would prevail in the absence of large traders. Our stochastic impact process Y is of a controlled Ornstein-Uhlenbeck (OU) type, namely it is driven by a Brownian motion and the large trader's holdings in the risky asset (see eq. (4.3) below). The mean-reversion of Y models the transience of impact. The additional noise in Y gives a stochastic limit order book density compared to the static one in Remark 2.1.1, or it can be seen as the accumulated effect from other non-strategic large traders, see Remark 4.1.4.

For our multiplicative model with transient impact, we take the fundamental price $\overline{S}$ to be an exponential Brownian motion and permit for non-zero correlation with the stochastic volume effect process Y. In this setup, we study the optimal liquidation problem for infinite time horizon as a singular stochastic control problem of finite fuel type and construct its explicit solution. Our main result in this chapter, Theorem 4.2.1, gives the optimal strategy as the local time process of a diffusion reflected obliquely at a curved free boundary in $\mathbb{R}^2$, the state space being the impact level and the holdings in the risky asset. The stochasticity of our optimal strategy arises from its adaptivity to the transient component of the price dynamics and is of local time type. In contrast to the models with additive price impact, where the martingale part of the fundamental price is irrelevant for a risk-neutral trader, here the volatility of $\overline{S}$ is relevant, cf. Remark 4.2.3.

We solve the singular control problem by explicitly constructing the value function as a classical solution of the HJB variational inequality. Our verification arguments differ from a more common approach (outlined in Remark 4.5.4) since we were not able to verify the optimality more directly for general impact functions f, due to the technical complications arising from the implicit nature of the eigenfunctions of the infinitesimal generator for the OU process. For particular choices of f, a more direct verification would be available, see Section 4.5.3. In contrast, we first restrict the set of optimization strategies to those described by diffusions reflected at monotone boundaries, and optimize over the set of possible boundaries. To be able to apply methods from calculus of variations, we derive an explicit formula (eq. (4.17)) for the Laplace transform of the inverse local times of diffusions reflected at elastic boundaries, i.e. boundaries which vary with the local time that the reflected process has spent at the boundary, and employ a change of coordinates. By solving the calculus of variations problem, we construct the candidate optimal free boundary and, moreover, show (one-sided) local optimality in the sense of Theorem 4.4.6. The latter is crucial for our verification of optimality. In Section 4.6, we present an optimal stopping problem which the directional derivative of the singular control value function satisfies. This connections is interesting, because it may hint on an alternative verification utilizing the rich literature on optimal stopping (we refer to the book [PS06]). Finally, with Section 4.7, we investigate the deterministic liquidity limit and show that for vanishing stochasticity of market liquidity, the free boundary converges to its counterpart from Chapter 2.

Skorokhod M_1/J_1 stability for gains from large investors' strategies (Chapter 5)

One important aspect in the theory of stochastic differential equations is how stably the solution process behaves, as a functional of its integrand and integrator processes, see e.g. [KP96] and [Pro05, Chapter V.4]. A typical question is how to extend such a functional sensibly to a larger class of input processes. Continuity in suitable topologies is a key property for addressing such problems, cf. e.g. in [Mar81] for his canonical extension of Stratonovich SDEs.

For instance in singular control problems, the non-linear objective functional may initially be only defined for finite variation or even absolutely continuous control strategies. Existence of an optimizer might require a continuous extension of the functional to a more general class of controls, e.g. semimartingale controls for the problem of hedging. Herein the question of which topology to embrace arises, and this depends on the problem at hand, see e.g. [Kar13] for an example of utility maximization in a frictionless financial market where the Emery topology turns out to be useful for the existence of an optimal wealth process. For our application we need suitable topologies on the Skorokhod space of càdlàg functions. The two most common choices here are the uniform topology and Skorokhod J_1 topology; they share the property that a jump in a limiting process can only be approximated by jumps of comparable size at the same time or, respectively, at nearby times. But this is overly restrictive for the optimal trading applications we have in mind, where a large jump may be approximated sensibly by many small jumps in fast succession or by continuous processes such as Wong-Zakai-type approximations. The M_1 topology by Skorokhod [Sko56] captures such approximations of *unmatched jumps*. This choice serves as the starting point to identify the relevant non-linear objective functional for càdlàg controls as a continuous extension from (absolutely) continuous controls. See [Whi02] for a profound survey on the M_1 topology.

Applications of the M_1 topology include queuing theory, functional statistics, mathematical neuroscience and scaling limits for random walks. We refer the reader to [Led16], which also contains an extensive list of literature. We tackle the old subject of stability of SDEs with jumps, when considered with respect to the M_1 topology, in the context of trading a single risky asset in an illiquid financial market, where a large investor's trading causes transient impact on the asset price. Our setting for this chapter is rather general. It can accommodate for instance for models where price impact is basically additive, see Example 5.1.1; yet, some extra provisions are required here to ensure M_1 continuity, which can actually fail to hold in common additive models that lack a monotonicity property and positivity of prices, cf. Remark 5.2.9. In line with the rest of this thesis, our framework here also permits for multiplicative impact. The details of this aspect are worked out in a previous chapter in Remark 3.2.7.

The large trader's feedback effect on prices causes the proceeds (negative expenses) to be a non-linear functional of her control strategy for dynamic trading in risky assets. Having specified the evolution for an affected price process at which trading of infinitesimal quantities would occur, one still has, even for a simple block trade, to define the variations in the bank account by which the trades in risky assets are financed, i.e. the so-called self-financing condition. Choosing a seemingly sensible, but ad-hoc, definition could lead to surprising and undesirable consequences, in that the large investor can evade her liquidity costs entirely by using continuous finite variation strategies to approximate her target control strategy, cf. Example 5.2.2. Optimal trade execution proceeds may be only approximately attainable in such models. Indeed, the analysis in [BB04, ÇJP04] shows that approximations by continuous strategies of finite variation play a particular role. This is, of course, a familiar theme in stochastic analysis, at least since [WZ65]. A notion of approximately realizable gains is relevant for the mathematical analysis of price impact models, cf. also Remark 5.2.3. For example, in the models in [BB04, ÇJP04] the aforementioned strategies have zero liquidity costs, permitting the large trader to avoid those costs entirely by simply approximating more general strategies. This appears not desirable from an application point of view. To settle this issue, a stability analysis for proceeds for a class of price impact models should address in particular the M_1 topology, in which continuous finite variation strategies are dense in the space of càdlàg strategies (in contrast to the uniform or J_1 topologies), see Remark 5.2.5.

We contribute a systematic study on stability of the proceeds functional. Starting with an unambiguous definition (5.4) for continuous finite-variation strategies, we identify the approximately realizable gains for a large set of controls. In particular for càdlàg finite-variation controls, we obtain the form of objective functional that is usually employed in stochastic control problems, see e.g. [Zhu92]. A mathematical challenge for stability of the stochastic integral functional is that both the integrand and the integrator depend on the control strategy. Our main Theorem 5.2.7 in this chapter shows continuity, in probability, of this non-linear controlled functional in the uniform, J_1 and M_1 topologies on the space of (predictable) semimartingale or càdlàg strategies which are bounded in probability. Another direct implication of M_1 continuity is that proceeds of general (optimal) strategies can be approximated by those of simple strategies with only small jumps. Whereas the former property is typical for common stochastic integrals, it is far from obvious for our non-linear controlled SDE functional (5.15).

The topic of stability for the proceeds process, that keeps track of gains from dynamically trading risky assets in illiquid markets, where the dynamics of wealth and proceeds for a large trader are non-linear in her strategies because of her market impact, is showing up at several places in the literature. But the mathematical topic appears to have been touched mostly in-passing so far. The focus of few notable investigations has been on the application context and on different topologies, see e.g. [RS13, Prop. 6.2] for uniform convergence in probability (ucp). In [LS13, Lem. 2.5] a cost functional is extended from simple strategies to semimartingales via convergence in ucp. [Roc11, Def. 2.1] and [ÇJP04, Sect. A.2] use particular choices of approximating sequences to extend their definition of self-financing trading strategies from simple processes to semimartingales by limits in ucp. Trading gains of semimartingale strategies are defined in [BLZ16, Prop. 1.1–1.2] as L^2-limits of gains from simple trading strategies via rebalancing at discrete times and large order split. In contrast, we contribute a study of M_1-, J_1- and ucp-stability for general approximations of càdlàg strategies in a class of price impact models with transient impact (5.3), driven by quasi-left continuous martingales (5.1).

As one particular application example that calls for a larger class of admissible strategies (than only finite variation trading), we solve in Section 5.3 the liquidation problem with stochastic liquidity à la Chapter 4, but where the time horizon is bounded by an expectation constraint for stopping times. This relies on M_1 convergence to define the trading proceeds. It provides an example of a liquidation problem where the optimum of singular controls is not attained in a class of finite variation strategies, but a suitable extension to semimartingale strategies is needed.

Proportional bid-ask spreads in optimal trading – a double obstacle problem (Chapter 6)

A large investor who optimizes her portfolio or tries to liquidate her position in a risky asset can freely choose whether to buy or to sell at each time. This is different from e.g. a bank that liquidates some position on a client's behalf, where trading in the opposite direction (intermediate buying) is typically forbidden by regulations or legislation. We adopt the view of the large investor acting on herself. When her price impact from trading involves uncertainty, a simple price model without bid-ask spread would suggest to follow an infinite variation strategy where to buy and sell persistently in order to pinpoint her price impact at a desired level, as we already saw in Section 5.3. Implementing such a strategy becomes costly though, since even for relatively liquid assets a one-tick-spread is typical, cf. e.g. [CdL13].

One way to tackle this problem is to not complicate the impact model further but leave it as it is – a zero spread idealization – and accordingly interpret an infinite variation strategy as an ideal to approximate by e.g. many small jumps. In our price impact model, such approximation would give similar trading gains, cf. Chapter 5. However, this qualitative result falls short of giving an quantitative estimate of the induced approximation error. In Chapter 6, we take a different route and instead adjust our market impact model slightly to incorporate bid-ask-spread by means of a fixed transaction cost factor.

Related to Chapter 4 we pose the optimal liquidation problem on an infinite time horizon with finite variation strategies in Section 6.1. Due to the bid-ask spread, cost terms for buying and selling differ. See e.g. [KW01, DAF14, FP14] for such kind of *reversible investment* problems of finite fuel type. These articles differ from our setup, in that the costs for (infinitesimal) buying and selling depend solely on time there, while in our case also depend on the (controlled) diffusion.

We explicitly solve the unconstrained (optimal investment) problem, where a risk-neutral large investor trades in infinite time horizon to maximize her expected proceeds, without a priori constraint on her asset position. The state space herein is one-dimensional and the optimal strategy turns out to be given by two reflection local times that keep the market impact level in some finite interval, see Theorem 6.2.8.

For the more involved problem of liquidating a given asset position, based on a smooth pasting approach, we conjecture the optimal strategy again to be the difference of two reflection local times which keep the impact process in a now non-constant (asset position dependent) interval until full liquidation. Verification seems difficult though, regarding the hurdles in verifying the one-sided analog in Chapter 4, which is the infinite transaction cost limit. An alternative approach would be to connect our non-monotone singular control problem to a Dynkin game (double obstacle problem), as is was successfully carried out in [DY09].

Approximating diffusion reflections at elastic boundaries (Chapter 7)

The classical Skorokhod problem is that of reflecting a path at a boundary. It is a standard tool to construct solutions to SDEs with reflecting boundary conditions. The fundamental example is Brownian motion with values in $[0, \infty)$ being reflected at a constant boundary at zero, solved by Skorokhod [Sko61]. Starting with Tanaka [Tan79], well-known generalizations concern diffusions in multiple dimensions with normal or oblique reflection at the boundary of some given (time-invariant) domain in the Euclidean space of certain smoothness or other kinds of regularity, cf. e.g. [LS84, DI93]. Other generalizations admit for an a-priori given but time-dependent boundary, see for instance [NO10].

Our contribution is a functional limit result for reflection at a boundary which is a function of the reflection local-time L, for general one-dimensional diffusions X. Because of the mutual interaction between boundary and diffusion, see Figure 7.1a, we call the boundary *elastic*. Such elastic boundaries appear typically in solutions to singular control problems of finite fuel type, where the optimal control is the reflection local time that keeps a diffusion process within a no-action region, cf. Karatzas and Shreve [KS86]. In order to explicitly construct the control (pathwise via Skorokhod's Lemma), finite fuel studies typically assume that the dynamics of the diffusion can be expressed without reference to the control (see e.g. [Kob93, EKK91]). This is different to our setup, where the non-linear mutual interdependence between diffusion and control (local time) subverts direct construction by Skorokhod's lemma, already for Ornstein-Uhlenbeck processes [WG03, Remark 1]. We relate to a concrete application in the context of optimal liquidation for a financial asset position in Remark 7.2.4.

A natural idea for approximation is to replace 'infinitesimal' reflections with small ε-jumps ΔL^{ε}, thereby inducing jumps of the elastic reflection boundary, see Figure 7.2. This allows to express excursion lengths of the approximating diffusion X^{ε} in terms of independent hitting times for continuous diffusions, what naturally leads to an explicit expression (7.12) for the Laplace transform of the inverse local time of X. In our singular control context, L^{ε} is asymptotically optimal at first order if L is optimal, see Remark 7.2.4. Our main result is Theorem 7.2.2. In Section 7.3, we prove ucp-convergence of $(X^{\varepsilon}, L^{\varepsilon})$ to (X, L) by showing tightness of the approximation sequence $(X^{\varepsilon}, L^{\varepsilon})_{\varepsilon}$ and using Kurtz–Protter's notion of uniformly controlled variations (UCV), introduced in [KP91].

Part I Deterministic multiplicative and transient price impact

2 A deterministic price impact model for optimal liquidation

Throughout the whole thesis, we consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. The filtration $(\mathcal{F}_t)_{t\geq 0}$ is assumed to satisfy the usual conditions of right-continuity and completeness, all semimartingales have càdlàg paths, and (in-)equalities of random variables are meant to hold almost everywhere. We refer to [JS03] for terminology and notations from stochastic analysis. We take $\mathcal{F}_0$ to be trivial and let also $\mathcal{F}_{0-}$ denote the trivial σ-field. Wherever we talk about given (semi-)martingales, we assume our probability space to be large enough to contain these and the martingale property to hold w.r.t. the filtration $(\mathcal{F}_t)_{t\geq 0}$.

This chapter introduces the general ideas of our transient market impact model and presents the main results of optimal liquidation in infinite time horizon with deterministic impact dynamics, Theorems 2.2.4 and 2.3.2, from the article [BBF18b], where the proofs can be found. Unlike now classical and more recent contributions to the optimal execution literature such as [AC00, OW13, AS10, PSS11, ASS12, LS13, GH17] that model price impact of a large trader to be additive, we follow earlier treatments in the economics and mathematical finance literature like [Jar94, Fre98, BL98, FJ02, FGL+04, HM05] by imposing a multiplicative structure for price impact. In subsequent Chapters 4 to 6 and Section 5.3 we will extend our market impact model to incorporate stochasticity and also consider the more involved optimal liquidation and execution problems in finite time horizon in Chapter 3.

2.1 Transient and multiplicative price impact

We consider a market with a risky asset in addition to the riskless numéraire asset, whose (discounted) price is constant at 1. Without trading activity of a large trader, the unaffected (fundamental) price process $\overline{S}$ of the risky asset would be of the form

$$\overline{S}_t = e^{\mu t} M_t, \qquad \overline{S}_0 \in (0, \infty), \tag{2.1}$$

with $\mu \in \mathbb{R}$ and with M being a non-negative martingale that is square integrable on any compact time interval, i.e. $\sup_{t\leq T} \mathbb{E}[M_t^2] < \infty$ for all $T \in [0, \infty)$, and quasi-left continuous (cf. [JS03]), i.e. $\Delta M_\tau := M_\tau - M_{\tau-} = 0$ for any finite predictable stopping time τ. Let us assume that the unaffected market is free of arbitrage for small investors in the sense that $\overline{S}$ is a local $\mathbb{Q}$-martingale under some probability measure $\mathbb{Q}$ that is locally equivalent to $\mathbb{P}$, i.e. $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F}_T$ for any $T \in [0, \infty)$. This implies no free lunch with vanishing risk [DS98] on any finite horizon T for small investors. The prime example where our assumptions on M are satisfied is the Black-Scholes-Merton model, where $M/\overline{S}_0 = \mathcal{E}(\sigma W)$ is the stochastic exponential of a Brownian motion W scaled by $\sigma > 0$. More generally, $M/\overline{S}_0 = \mathcal{E}(L)$ could be the stochastic exponential of a local martingale L, which is a Lévy process with $\Delta L > -1$ and $\mathbb{E}[M_1^2] < \infty$ and such that $\overline{S}$ is not monotone (see [Kal00, Lemma 4.2] and [CT04, Theorem 9.9]), or one could have $M = \mathcal{E}(\int \sigma_t \,\mathrm{d}W_t)$ for predictable stochastic volatility process $(\sigma_t)_{t\geq 0}$ that is bounded in $[1/c, c]$, for $c > 1$.

The large trader's strategy $(\Theta_t)_{t\geq 0}$ is her position in the risky asset. Herein, $\Theta_{0-} \geq 0$ denotes the initial position, $\Theta_{0-} - \Theta_t$ is the cumulative number of risky assets sold until time t. The process Θ is predictable, càdlàg and non-negative, i.e. short sales are not permitted, like for instance in [KP10, GZ15]. Disallowing short sales is sensible for the control problem with infinite horizon and negative drift to ensure existence of optimizers and finite time to complete liquidation; It is also supported e.g. by [Sch13, Remark 3.1]. At first we do moreover assume Θ to be decreasing, but this will be generalized later in Section 2.3 to non-monotone strategies of bounded variation.

The large trader is faced with illiquidity costs, since trading causes adverse impact on the prices at which orders are executed, as follows. A process Y, the *market impact process*, captures the price impact from strategy Θ, and is defined as the solution to

$$\mathrm{d}Y_t = -h(Y_t)\,\mathrm{d}t + \mathrm{d}\Theta_t \tag{2.2}$$

for some given initial condition $Y_{0-} \in \mathbb{R}$. Let $h : \mathbb{R} \to \mathbb{R}$ be strictly increasing and continuous with $h(0) = 0$. Further conditions will be imposed later in Assumption 2.2.2. The market is resilient in that market impact Y tends back towards its neutral level 0 over time when the large trader is not active. Resilience is transient with resilience rate $h(Y_t)$ that could be non-linear and is specified by the *resilience function h*. For example, the market recovers at exponential rate $\beta > 0$ (as in [OW13]) when $h(y) = \beta y$ is linear. Clearly, Y depends on Θ and occasionally we will emphasize this by writing $Y = Y^\Theta$.

The actual (quoted) risky asset price S is affected by the strategy Θ of the large trader in a multiplicative way through the market impact process Y, and is modeled by

$$S_t := f(Y_t)\overline{S}_t, \tag{2.3}$$

for an increasing function f of the form

$$f(y) = \exp\Big(\int_0^y \lambda(x)\,\mathrm{d}x\Big), \quad y \in \mathbb{R}, \tag{2.4}$$

with $\lambda : \mathbb{R} \to (0, \infty)$ satisfying Assumption 2.2.2 below, in particular being locally integrable. For strategies Θ that are continuous, the process $(S_t)_{t\geq 0}$ can be seen as the evolution of prices at which the trading strategy Θ is executed. That means, if the large trader is selling risky assets according to a continuous strategy Θ^c, then respective (self-financing) variations of her numéraire (cash) account are given by the proceeds (negative costs) $-\int_0^T S_u \,\mathrm{d}\Theta_u^c$ over any period $[0,T]$. To permit also for non-continuous trading involving block trades, the proceeds from a market sell order of size $\Delta\Theta_t \in \mathbb{R}$ at time t, are given by the term

$$-\overline{S}_t \int_0^{\Delta\Theta_t} f(Y_{t-} + x)\,\mathrm{d}x, \tag{2.5}$$

which is explained from executing the block trade within a (shadow) limit order book, see Remark 2.1.1. Mathematically, defining proceeds from block trades in this way ensures good stability properties for proceeds defined by (2.8) as a function of strategies Θ, cf. Chapter 5. In particular, approximating a block trade by a sequence of continuous trades executed over a shorter and shorter time interval yields the term (2.5) in the limit, see Corollary 5.2.10.

Remark 2.1.1 (Limit order book perspective)**.** Multiplicative price impact and the proceeds from block trading can be interpreted by trading in a shadow limit order book (LOB). We now show how the multiplicative price impact function f is related to a LOB shape that is specified in terms of *relative* price perturbations $\rho_t := S_t/\overline{S}_t$, whereas additive impact corresponds to a LOB shape being specified with respect to absolute price perturbations $S_t - \overline{S}_t$ as in [PSS11]. Note that the LOB shape is static (and Section 2.3 considers a two-sided LOB with zero bid-ask spread). Such can be viewed as a low-frequency model for price impact according to a LOB shape which is representative on longer horizons, but not for high frequency trading over short periods.

Let $s = \rho\overline{S}_t$ be some price close to the unaffected price $\overline{S}_t$ and let $q(\rho)\,\mathrm{d}\rho$ denote the density of (bid or ask) offers at price level s, i.e. at the relative price perturbation ρ. This leads to a measure with cumulative distribution function $Q(\rho) := \int_1^\rho q(x)\,\mathrm{d}x$, $\rho \in (0,\infty)$. The total volume of orders at prices corresponding to perturbations ρ from some range $R \subset (0,\infty)$ then is $\int_R q(x)\,\mathrm{d}x$. Selling $-\Delta\Theta_t$ shares at time t shifts the price from $\rho_{t-}\overline{S}_t$ to $\rho_t\overline{S}_t$, while the volume change is $Q(\rho_{t-}) - Q(\rho_t) = -\Delta\Theta_t$. The proceeds from this sale are $\overline{S}_t \int_{\rho_t}^{\rho_{t-}} \rho\,\mathrm{d}Q(\rho)$. In the terminology of [Kyl85], $Q(r_t) - Q(r_{t-})$ reflects the *depth* of the LOB for price changes by a factor of r_t/r_{t-}. Changing variables, with $Y_t := Q(\rho_t)$ and $f := Q^{-1}$, the proceeds can be expressed as in equation (2.5). In this sense, Y from (2.2) can be understood as the *volume effect process* as in [PSS11, Section 2]. By the drift towards zero in (2.2), this effect is persistent over time but not permanent. Its transient nature relates to the liquidity property that [Kyl85] calls resilience. See Figure 2.1 for illustration.

Example 2.1.2. Let the (one- or two-sided) shadow limit order book density be $q(x) := c/x^r$ on $x \in (0,\infty)$ for constants $c, r > 0$. Parameters c and r determine the market depth (LOB volume): If $r < 1$, a trader can sell only finitely many but buy infinitely many assets at any time. In contrast, for $r > 1$ one could sell infinitely many but buy only finitely many assets at any time instant and (by (2.2)) also in any finite time period. Note that [PSS11, p.185] assume infinite market depth in the target trade direction. The case $r = 1$ describes infinite market depth in both directions. The antiderivative Q and its inverse f are determined for $x > 0$ and $(r-1)y \neq c$ as

$$Q(x) = \begin{cases} c\log x, & \text{for } r = 1,\\ \frac{c}{1-r}(x^{1-r} - 1), & \text{otherwise,} \end{cases} \qquad f(y) = \begin{cases} e^{y/c}, & \text{for } r = 1,\\ \big(1 + \frac{1-r}{c}y\big)^{1/(1-r)}, & \text{otherwise.} \end{cases}$$

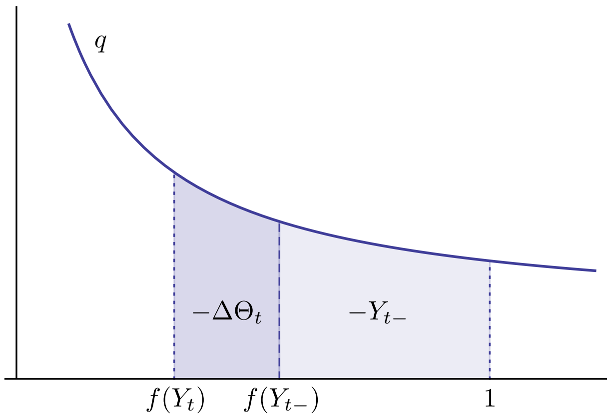


Figure 2.1: Order book density q and behavior of the multiplicative price impact $f(Y)$ when selling a block of size $-\Delta\Theta_t > 0$. Note that $-Y_t = -Y_{t-} - \Delta\Theta_t$.

For the parameter function λ this yields $\lambda(y) = f'(y)/f(y) = (c + (1-r)y)^{-1}$. Note that for $r \neq 1$ the functions f and λ are effectively constrained to the domain $(\frac{c}{r-1}, \infty)$ for $r < 1$ and $(-\infty, \frac{c}{r-1})$ for $r > 1$. In the thesis, we assume that $f > 0$ is defined on the whole real line for simplicity. Yet, let us use this example to explain next how also interesting cases like $r \in (0,\infty) \setminus \{1\}$ can be dealt with by refining the definition of the set of admissible strategies according to f. Indeed, properties of f are only needed within the range of possible values of processes Y^Θ. Hence, the more general case where $I_f := \{y : 0 < f(y) < \infty\}$ is an open interval in $\mathbb{R}$ can be treated by imposing as an additional requirement for admissibility of a strategy Θ (in (2.6), (2.30)) that Y^Θ has to evolve in I_f. For further investigations of this case, see [BBF18b, Example 4.3].

2.2 The problem case for monotone strategies

This section solves the optimal liquidation problem that is central for this chapter. The large investor is facing the task to sell Θ_{0-} risky assets but has the possibility to split it into smaller orders to improve according to some performance criterion. Before Section 2.3, we will restrict ourselves to monotone control strategies that do not allow for intermediate buying. The analysis for this more restricted variant of control policies will be shown later in Section 2.3 to carry over to an alternative problem with a wider set of controls, being of finite variation, admitting also intermediate buy orders.

For an initial position of Θ_{0-} shares, the set of admissible trading strategies is

$$\mathcal{A}_{\text{mon}}(\Theta_{0-}) := \big\{\Theta \;\big|\; \Theta \text{ is decreasing, càdlàg, predictable,} \\ \text{with } \Theta_{0-} \geq \Theta_t \geq 0\big\}. \tag{2.6}$$

Here, the quantity Θ_t represents the number of shares held at time t. Any admissible strategy $\Theta \in \mathcal{A}_{\text{mon}}(\Theta_{0-})$ decomposes into a continuous and a discontinuous part

$$\Theta_t = \Theta_t^c + \sum_{0 \leq s \leq t} \Delta\Theta_s, \tag{2.7}$$

where Θ_t^c is continuous (and decreasing) and $\Delta\Theta_s := \Theta_s - \Theta_{s-} \le 0$. Aiming for an explicit analytic solution, we consider trading on the infinite time horizon $[0,\infty)$ with discounting. The γ-discounted proceeds from strategy Θ up to time $T < \infty$ are

$$L_T(y;\Theta) := -\int_0^T e^{-\gamma t} f(Y_t)\overline{S}_t \,\mathrm{d}\Theta_t^c - \sum_{\substack{0\le t\le T\\ \Delta\Theta_t\neq 0}} e^{-\gamma t}\overline{S}_t \int_0^{\Delta\Theta_t} f(Y_{t-}+x)\,\mathrm{d}x, \tag{2.8}$$

where $y = Y_{0-}$ is the initial state of process Y. Clearly, Y_{0-} and Θ determine Y by (2.2).

Remark 2.2.1. The (possibly) infinite sum in (2.8) has finite expectation. Indeed, for any $\Theta \in \mathcal{A}_{\mathrm{mon}}(\Theta_{0-})$ one has $\sup_{t\le T}|Y_t| < \infty$. Hence, the mean value theorem and properties of f imply for $t \in [0,T]$ that

$$0 \le -\int_0^{\Delta\Theta_t} f(Y_{t-}+x)\,\mathrm{d}x \le -\Delta\Theta_t \sup_{x\in(\Delta\Theta_t,0)} f(Y_{t-}+x) \le -\Delta\Theta_t\cdot\sup_{t\le T} f(Y_t)\,.$$

Thus, by finite variation of Θ the infinite sum in (2.8) a.s. converges absolutely. For $\Theta\in\mathcal{A}_{\mathrm{mon}}(\Theta_{0-})$ the sum is bounded in expectation, because Y and hence $\sup_{t\le T} f(Y_t)$ are bounded, and we have $\mathbb{E}[\sup_{t\in[0,T]}\overline{S}_t] < \infty$ and $0 \le \sum_{t\in[0,T]}(-\Delta\Theta_s) \le \Theta_{0-}$.

Note that the monotone limit $L_\infty(y;\Theta) := \lim_{T\nearrow\infty} L_T(y;\Theta)$ always exists. We consider the control problem to find the optimal strategy that maximizes the expected (discounted) liquidation proceeds over an open (infinite) time horizon

$$\max_{\Theta\in\mathcal{A}_{\mathrm{mon}}(\Theta_{0-})} J(y;\Theta) \qquad \text{for} \quad J(y;\Theta) := \mathbb{E}[L_\infty(y;\Theta)], \tag{2.9}$$

$$\text{with value function} \quad v(y,\theta) := \sup_{\Theta\in\mathcal{A}_{\mathrm{mon}}(\theta)} J(y;\Theta). \tag{2.10}$$

For this problem maximizing over deterministic strategies turns out to be sufficient (see Remark 2.2.6 below). Since expectations $\mathbb{E}[\exp(-\gamma t)\overline{S}_t] = \overline{S}_0\exp(-t(\gamma-\mu))$, $t\ge 0$, depend on μ,γ only through $\delta := \gamma-\mu$, for our optimization problem just the difference δ matters which needs to be positive to have $v(y,\theta) < \infty$ for $\theta > 0$. Thus, regarding γ and μ, only the difference δ will be needed, and it might be interpreted as impatience parameter chosen by the large investor (when choosing γ), specifying her preferences to liquidate earlier rather than later, as a drift rate of the risky asset returns $d\overline{S}/\overline{S}$, or as a combination thereof. The following conditions on δ, f, h are assumed for the remaining Sections 2.2 and 2.3 of this chapter.

Assumption 2.2.2. The map $t\mapsto \mathbb{E}[e^{-\gamma t}\overline{S}_t]$, $t\ge 0$, is decreasing, i.e. $\delta := \gamma-\mu > 0$. The price impact function $f:\mathbb{R}\to(0,\infty)$ satisfies $f(0)=1$, $f\in C^2$ and is strictly increasing such that $\lambda(y) := f'(y)/f(y) > 0$ everywhere.
The resilience function $h:\mathbb{R}\to\mathbb{R}$ from (2.2) is C^2 with $h(0)=0$ and $h'>0$.
Resilience and market impact satisfy $(h\lambda)' > 0$ and $(h\lambda + h')' > 0$.
There exist solutions y_0 to $h(y_0)\lambda(y_0)+\delta = 0$ and y_∞ to $h(y_\infty)\lambda(y_\infty)+h'(y_\infty)+\delta = 0$.
(Uniqueness of y_0 and y_∞ holds by the other conditions.)

Remark 2.2.3 (Interplay of impact and resilience functions)**.** The two assumptions $(h\lambda)' > 0$ and $(h\lambda+h')' > 0$ are technical requirements for our verification of optimality.

Already from the shadow limit order book (LOB) perspective (cf. Remark 2.1.1), some sort of condition connecting both resilience speed h and price impact f, thus LOB shape, appears natural. Examples satisfying Assumption 2.2.2 with $f(y) := e^{\lambda y}$ for constant $\lambda > 0$ are e.g. linear resilience speed, $h(y) = \beta y$ with $\beta > 0$ and any discounting $\delta > 0$; or for instance bounded resilience, $h(y) = \alpha \arctan(\beta y)$, for $\alpha, \beta > 0$, with $\beta < \lambda$ and not too large discounting $0 < \delta < \frac{1}{2}\alpha\lambda\pi$ (a larger δ would give that the trivial strategy to sell everything initially at time 0 is optimal).

The main results Theorems 2.2.4 and 2.3.2 of this chapter solve the optimal liquidation problem for one- respectively two- sided limit order books in infinite time horizon. The proof of Theorem 2.2.4 is developed in [BBF18b, Sect. 4] using smooth pasting and calculus of variations approaches to obtain a candidate solution, together with direct verification of the variational (in-)equalities (2.17)–(2.20).

Theorem 2.2.4. *Let the model parameters h, λ, δ satisfy Assumption 2.2.2 and $\Theta_{0-} \geq 0$ be given. Define $y_\infty < y_0 < 0$ as the unique solutions of $h(y_\infty)\lambda(y_\infty) + h'(y_\infty) + \delta = 0$ and $h(y_0)\lambda(y_0) + \delta = 0$, respectively, and let*

$$\tau(y) := -\frac{1}{\delta} \log\left(\frac{f(y)}{f(y_0)} \frac{h(y)\lambda(y) + h'(y) + \delta}{h'(y)} \right), \tag{2.11}$$

for $y \in (y_\infty, y_0]$ with inverse function $\tau \mapsto \bar{y}(\tau) : [0, \infty) \to (y_\infty, y_0]$. Moreover, consider the decreasing function $\theta : (y_\infty, y_0] \to [0, \infty)$ given by

$$\theta(y) := \int_{y_0}^{y} \left(1 + \frac{h(z)\lambda(z)}{\delta} - \frac{h(z)h''(z)}{\delta h'(z)} + \frac{h(z)\big(h\lambda + h' + \delta\big)'(z)}{\delta\big(h\lambda + h' + \delta\big)(z)} \right) \mathrm{d}z \tag{2.12}$$

and denote its inverse by $\theta \mapsto \mathrm{y}(\theta)$, $\theta \geq 0$. For given $\Theta_{0-} \geq 0$ and $Y_{0-} \in \mathbb{R}$

$$\begin{aligned} \Delta_0 &:= \inf \{ d \in [0, \Theta_{0-}] \mid Y_{0-} - d \leq \mathrm{y}(\Theta_{0-} - d) \} \wedge \Theta_{0-} \,, \\ T_w &:= \inf \{ t > 0 \mid y_w(t) > \mathrm{y}(\Theta_{0-}) \} \,, \\ T &:= T_w + \tau\big(\mathrm{y}(\Theta_{0-} - \Delta_0)\big) \,, \end{aligned}$$

where $y_w \in C^1([0, \infty))$ solves $y_w'(t) = -h(y_w(t))$, for $t \geq 0$, with $y_w(0) = Y_{0-}$. Define the process Θ^{opt} by

$$\Theta_t^{opt} := (\Theta_{0-} - \Delta_0)\mathbb{1}_{[0,T_w)}(t) + \theta\big(\bar{y}(T - t)\big)\mathbb{1}_{[T_w,T)}(t) \qquad \text{for } t \geq 0. \tag{2.13}$$

Then the strategy Θ^{opt} is the unique maximizer to the problem (2.9) of optimal liquidation $\max_{\Theta \in \mathcal{A}_{mon}(\Theta_{0-})} \mathbb{E}[L_\infty(y; \Theta)]$ for Θ_{0-} assets with initial market impact being $Y_{0-} = y$.

Note that the optimal liquidation strategy does not depend on the particular form of the martingale M (what has been noted as a robust property in related literature). Since $T < \infty$ is finite, the open horizon control from Theorem 2.2.4 is clearly optimal for the problem on any finite horizon $T' \geq T$; cf. Remark 2.2.10 and Chapter 3 for $T' < T$.

Remark 2.2.5 (The optimal sell-only strategy)**.** The strategy Θ^{opt} from Theorem 2.2.4 acts as follows.

1. If $Y_{0-} \geq y_0 + \Theta_{0-}$, sell all assets at once: $\Theta_0 = 0$.

2. If $\mathrm{y}(\Theta_{0-}) < Y_{0-} < y_0 + \Theta_{0-}$, then sell a block of size $\Delta_0 \equiv \Theta_{0-} - \Theta_0$ such that $\Theta_0 > 0$ and $Y_0 \equiv Y_{0-} + \Delta_0 = \mathrm{y}(\Theta_0)$.

3. If $Y_{0-} < \mathrm{y}(\Theta_{0-})$, wait until time T_w. That is, set $\Theta_t = \Theta_{0-}$ for $t < T_w$. This leads to $Y_t = y_w(t)$ for $t < T_w$.

4. As soon as step 2 or 3 lead to the state $Y_s = \mathrm{y}(\Theta_s)$ for some time $s \geq 0$, sell continuously: $\Theta_t = \theta(\bar{y}(T-t))$, $s \leq t \leq T$, until time $T = s + \tau(\mathrm{y}(\Theta_s))$.

5. Stop when all assets are sold at some time $T < \infty$: $\Theta_t = 0$, $t \in [T, \infty)$.

Remark 2.2.6 (On deterministic optimal controls)**.** The optimal liquidation strategy is deterministic and the value function turns out to be continuous (even differentiable). More precisely, the value function is

$$V(y,\theta) = \begin{cases} V_{\mathrm{bdry}}(\theta - \Delta(y,\theta)) + \int_{y-\Delta(y,\theta)}^{y} f(x)\,\mathrm{d}x, & \text{for } y \geq \mathrm{y}(\theta) \\ V_{\mathrm{bdry}}(\theta - \Delta(y,\theta)) \cdot \exp\left(\int_{\mathrm{y}(\theta)}^{y} \frac{-\delta}{h(x)}\,\mathrm{d}x\right) & \text{for } y \leq \mathrm{y}(\theta) \end{cases} \tag{2.14}$$

where $V_{\mathrm{bdry}}(\theta) = \left(fh\frac{h\lambda+\delta}{\delta h'}\right)(\mathrm{y}(\theta))$ and $\Delta(y,\theta) \geq 0$ such that $\mathrm{y}(\theta - \Delta(y,\theta)) = y - \Delta(y,\theta)$ for $\mathrm{y}(\theta) \leq y \leq y_0 + \theta$, whereas $\Delta(y,\theta) = \theta$ for $y > y_0 + \theta$ with $\mathrm{y}(\theta)$ denoting the boundary solution of Theorem 2.2.4, cf. [BBF18b, Lemma 4.2]. This is shown in [BBF18b, Sect. 4] while proving Theorem 2.2.4. Here, we show directly why non-deterministic strategies are suboptimal for (2.9) and optimizing over deterministic admissible controls is sufficient. Yet, finding explicit solutions here still requires to construct candidate solutions and prove optimality, as in the sequel.

If one considers optimization just over strategies that are to be executed until a time $T < \infty$, then the value function will be the same as if we were optimizing over the subset of deterministic strategies. Indeed, by optional projection (see [DM82, VI.57]) we have

$$\mathbb{E}[L_T(y;\Theta)] = -\mathbb{E}\Big[M_T \int_0^T e^{-\delta t} f(Y_{t-})\,\mathrm{d}\Theta_t^c\Big] - \mathbb{E}\Big[M_T \sum_{\substack{0 \leq t \leq T \\ \Delta\Theta_t \neq 0}} e^{-\delta t} \int_0^{\Delta\Theta_t} f(Y_{t-} + x)\,\mathrm{d}x\Big].$$

For any $T \in [0,\infty)$, letting $\mathrm{d}\widetilde{\mathbb{P}} = M_T/M_0\,\mathrm{d}\mathbb{P}$ on $\mathcal{F}_T$ yields $\mathbb{E}[L_T(y;\Theta)] = \mathbb{E}^{\widetilde{\mathbb{P}}}[\ell_T(\Theta)]$ for $\ell_T(\Theta) := -M_0 \int_0^T e^{-\delta t} f(Y_{t-})\,\mathrm{d}\Theta_t^c - M_0 \sum_{0 \leq t \leq T,\, \Delta\Theta_t \neq 0} e^{-\delta t} \int_0^{\Delta\Theta_t} f(Y_{t-} + x)\,\mathrm{d}x$. Note that ℓ is a deterministic functional of Θ, and that the measure $\widetilde{\mathbb{P}}$ does not depend on Θ. Thus, optimization for any finite horizon T can be done ω-wise, i.e. for the finite-horizon problem optimizing over the subset of deterministic strategies gives the same value function. Note that this is similar to [Løk12, Prop. 7.2]. Using monotonicity of L_T in T, we have $\mathbb{E}[L_\infty(y;\Theta)] = \sup_{T\in[0,\infty)} \mathbb{E}[L_T(y;\Theta)]$, hence the change of measure argument above yields that $v(y,\theta) = \sup_{T\in[0,\infty)} \sup_{\Theta\in\mathcal{A}_{\mathrm{mon}}(\theta)} \mathbb{E}[L_T(y;\Theta)]$ is equal to

$$\sup_{T\in[0,\infty)} \sup_{\substack{\Theta\in\mathcal{A}_{\mathrm{mon}}(\theta) \\ \text{deterministic}}} \ell_T(\Theta) = \sup_{\substack{\Theta\in\mathcal{A}_{\mathrm{mon}}(\theta) \\ \text{deterministic}}} \ell_\infty(\Theta). \tag{2.15}$$

Moreover, one can check that any deterministic maximizer $\Theta^* \in \mathcal{A}_{\mathrm{mon}}(\theta)$ to (2.15) is also optimal for the original problem (2.9), where $v(y,\theta) < \infty$ thanks to $\delta < 0$.

Remark 2.2.7. In [PSS11], Predoiu, Shaikhet and Shreve consider a similar optimal execution problem, with an additive price impact ψ such that $S_t = \overline{S}_t + \psi(Y_t)$ with *volume effect process* Y_t as in (2.2). They study the case of martingale $\overline{S}_t$ on a finite time horizon $[0, T]$. The execution costs, which they seek to minimize in expectation, are equal to the negative liquidation proceeds $-L_T$ in our model (for $\gamma, \mu = 0$) with fixed $Y_{0-} := 0$. See also Remark 2.2.10 below.

The next result provides sufficient conditions for optimality to the problem (2.9) for each possible initial state $Y_{0-} = y \in \mathbb{R}$ of the impact process, by the martingale optimality principle, for proof see [BBF18b, Prop. 3.6]. In contrast, in the related additive model in [PSS11] the optimal buying strategy for finite time horizon without drift ($\delta = 0$), and impact process starting at zero was characterized using an elegant convexity argument; cf. Remark 2.2.10.

Proposition 2.2.8. *Let $V : \mathbb{R} \times [0, \infty) \to [0, \infty)$ be a continuous function such that $G_t(y; \Theta) := L_t(y; \Theta) + e^{-\gamma t}\overline{S}_t \cdot V(Y_t, \Theta_t)$, with $Y = Y^\Theta$ and $y = Y_{0-}$, is a supermartingale for each $\Theta \in \mathcal{A}_{mon}(\Theta_{0-})$ and additionally $G_0(y; \Theta) \le G_{0-}(y; \Theta) := \overline{S}_0 \cdot V(Y_{0-}, \Theta_{0-})$. Then*

$$\overline{S}_0 \cdot V(y, \theta) \ge v(y, \theta)$$

with $\theta = \Theta_{0-}$. Moreover, if there exists $\Theta^* \in \mathcal{A}_{mon}(\Theta_{0-})$ such that $G(y; \Theta^*)$ is a martingale and it holds $G_0(y; \Theta^*) = G_{0-}(y; \Theta^*)$, then $\overline{S}_0 \cdot V(y, \theta) = v(y, \theta)$ and $v(y, \theta) = J(y; \Theta^*)$.

Remark 2.2.9. The additional condition on G_0 and G_{0-} can be regarded as extending the (super-)martingale property from time intervals $[0, T]$ to time „$0-$“.

In order to make use of Proposition 2.2.8, one applies Itô's formula to G, assuming that V is smooth enough and using the fact that $[\overline{S}_\cdot, e^{-\gamma \cdot} V(Y_\cdot, \Theta_\cdot)] = 0$ because $\overline{S}$ is quasi-left-continuous and $e^{-\gamma \cdot} V(Y_\cdot, \Theta_\cdot)$ is predictable and of bounded variation, to get

$$\begin{aligned} \mathrm{d}G_t = {} & e^{-\delta t} V(Y_{t-}, \Theta_{t-})\, \mathrm{d}M_t \\ & + e^{-\delta t} M_{t-} \Big(\big(-\delta V - h V_y\big)(Y_{t-}, \Theta_{t-})\, \mathrm{d}t \\ & \qquad + \big(V_y + V_\theta - f\big)(Y_{t-}, \Theta_{t-})\, \mathrm{d}\Theta_t^c \\ & \qquad + \int_0^{\Delta\Theta_t} \big(V_y + V_\theta - f\big)(Y_{t-} + x, \Theta_{t-} + x)\, \mathrm{d}x \Big) \end{aligned} \tag{2.16}$$

with the abbreviating conventions $\big(-\delta V - h V_y\big)(a, b) := -\delta V(a, b) - h(a) V_y(a, b)$ and $\big(V_y + V_\theta - f\big)(a, b) := V_y(a, b) + V_\theta(a, b) - f(a)$. The martingale optimality principle now suggests equations for regions where the optimal strategy should sell or wait, in that the $\mathrm{d}\Theta$-integrands should be zero when there is selling and the $\mathrm{d}t$-integrand must vanish when only time passes (waiting). We will construct a classical solution to the *variational inequality* $\max\{-\delta V - h V_y\,,\ f - V_y - V_\theta\} = 0$, that is a function V in $C^{1,1}(\mathbb{R} \times [0, \infty), \mathbb{R})$

and a strictly decreasing *free boundary* function $\mathrm{y}(\cdot) \in C^2([0,\infty),\mathbb{R})$, such that

$$
\begin{aligned}
-\delta V - h(y)V_y &= 0 && \text{in } \overline{\mathcal{W}} && (2.17)\\
-\delta V - h(y)V_y &< 0 && \text{in } \mathcal{S} && (2.18)\\
V_y + V_\theta &= f(y) && \text{in } \overline{\mathcal{S}} && (2.19)\\
V_y + V_\theta &> f(y) && \text{in } \mathcal{W} && (2.20)\\
V(y,0) &= 0 && \forall y \in \mathbb{R} && (2.21)
\end{aligned}
$$

for wait region $\mathcal{W}$ and sell region $\mathcal{S}$ (cf. Figure 2.2) defined as

$$
\begin{aligned}
\mathcal{W} &:= \{(y,\theta) \in \mathbb{R}\times[0,\infty) \mid y < \mathrm{y}(\theta)\},\\
\mathcal{S} &:= \{(y,\theta) \in \mathbb{R}\times[0,\infty) \mid y > \mathrm{y}(\theta)\}.
\end{aligned} \tag{2.22}
$$

The optimal liquidation studied here belongs to the class of finite-fuel control problems,

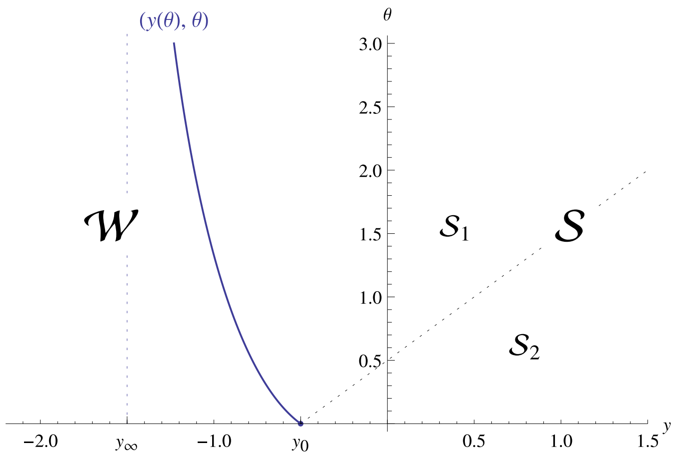


Figure 2.2: The division of the state space, for $\delta = 0.5$, $h(y) = y$ and $\lambda(y) \equiv 1$.

which often lead to *free boundary problems* similar to the one derived above. See [KS86] for an explicit solution of the finite-fuel monotone follower problem, and [JJZ08] for further examples and an extensive list of references. The proof of Theorem 2.2.4 consists of an explicit construction of $\mathrm{y}(\theta)$ and the value function V as in (2.14) by means of smooth pasting (or alternatively calculus of variations) and direct verification of the variational (in-)equalities (2.17)–(2.21). For details, see [BBF18b, Sect. 4].

Remark 2.2.10 (A first look on the finite time horizon problem)**.** For a given finite horizon $T < \infty$, the execution problem with general order book shape has been solved by [PSS11] for additive price impact and no drift ($\delta = 0$). The problem with multiplicative impact could be transformed to the additive situation using intricate state-dependent order book shapes, cf. [Løk12]. Let us show how a convexity argument as in [PSS11] can be applied also directly to solve the finite horizon case in the multiplicative setup when the drift δ is zero, but not for $\delta \neq 0$. We will solve the general case $\delta \in \mathbb{R}$ in Chapter 3 using calculus of variations.

By Remark 2.2.6 it suffices to consider deterministic strategies $\Theta \in \mathcal{A}_{\mathrm{mon}}(\Theta_{0-})$. Let $F(y) = \int_0^y f(x)\,\mathrm{d}x$. For deterministic Θ and $g(x) := f(h^{-1}(x))x + \delta F(h^{-1}(x))$ we have

$$\mathbb{E}[L_T(\Theta)] = F(Y_{0-}) - e^{-\delta T} F(Y_T) - \int_0^T e^{-\delta t} g(h(Y_t))\,\mathrm{d}t\,. \tag{2.23}$$

Since $g'(x) = \big(f\,(h\lambda + h' + \delta)/h'\big)(h^{-1}(x))$, the function g obtains a global minimum at $h(y_\infty)$ and is decreasing on the left and increasing on the right. So its convex hull $\hat{g}(x) = \sup\{\ell(x) \mid \ell \text{ is affine with } \ell \le g\}$ exists. With $C_{\delta,T} := \int_0^T e^{-\delta t}\,\mathrm{d}t$, Jensen's inequality yields

$$\mathbb{E}[L_T(\Theta)] \le F(Y_{0-}) - e^{-\delta T} F(Y_T) - C_{\delta,T} \cdot \hat{g}\bigg(\int_0^T h(Y_t) \frac{e^{-\delta t}}{C_{\delta,T}}\,\mathrm{d}t\bigg) \tag{2.24}$$

and we have equality in (2.24) if and only if $h(Y_t)$ stays constant in the interval where $\hat{g}$ and g coincide for almost all $t \in (0,T)$. Such impact-fixing strategies are analogous to the *type A strategies* of [PSS11].

The integral in (2.24) can be solved in general only for $\delta = 0$. Now let $\delta = 0$ but keep the other parts of Assumption 2.2.2. Note that $y_0 = 0$ subsequently. In that case $C_{\delta,T} = T$ and $\int_0^T h(Y_t)\,\mathrm{d}t = Y_{0-} - \Theta_{0-} - Y_T$ for any strategy that liquidates until time T, so

$$\mathbb{E}[L_T(\Theta)] \le F(Y_{0-}) - F(Y_T) - T\hat{g}\bigg(\frac{Y_{0-} - \Theta_{0-} - Y_T}{T}\bigg) =: \hat{G}(Y_T)\,. \tag{2.25}$$

In order to identify the convex hull $\hat{g}$ explicitly for $\delta = 0$, consider the second derivative

$$\begin{aligned} g''(x) &= 2f'\big(h^{-1}(x)\big)\big(h^{-1}\big)'(x) + f''\big(h^{-1}(x)\big)\big((h^{-1})'(x)\big)^2 x + f'\big(h^{-1}(x)\big)\big(h^{-1}\big)''(x)x \\ &= \bigg(2\frac{f'}{h'} + \frac{f''h}{(h')^2} - \frac{f'hh''}{(h')^3}\bigg)\big(h^{-1}(x)\big) \\ &= \bigg(\frac{f}{(h')^3}\Big(2\lambda(h')^2 + (\lambda' + \lambda^2)hh' - \lambda hh''\Big)\bigg)\big(h^{-1}(x)\big) \\ &= \bigg(\frac{f}{(h')^3}\Big((h\lambda + h')\lambda h' - h\lambda(h\lambda + h')' + (h\lambda + h')(h\lambda)'\Big)\bigg)\big(h^{-1}(x)\big). \end{aligned}$$

Since $h\lambda + h' > 0$ on (y_∞, ∞) and $h\lambda < 0$ on $(-\infty, y_0)$, we find that g is strictly convex on an open interval covering $\big[h(y_\infty), h(y_0)\big]$. Moreover, by $y_0 = 0$ we have $g'(0) = g'(h(y_0)) = f(0) = 1$ and for every $x > 0$, $g'(x) > f(h^{-1}(x)) \ge f(0) = 1$. Hence, we found a convex function that is dominated by g and therefore also by $\hat{g}$:

$$g(x) \ge \hat{g}(x) \ge \begin{cases} g(h(y_\infty)) & \text{for } x \le h(y_\infty), \\ g(x) & \text{for } x \in [h(y_\infty), 0], \\ g(0) + x & \text{for } x \ge 0. \end{cases} \tag{2.26}$$

Let $e(y) := (Y_{0-} - \Theta_{0-} - y)/T$, so that we have $\hat{G}'(y) = -f(y) + \hat{g}'\big(e(y)\big)$ and $\hat{G}''(y) = -f'(y) - \frac{1}{T}\hat{g}''\big(e(y)\big) \le -f'(y) < 0$, i.e. strict concavity of $\hat{G}$. Moreover, at $r := e^{-1}(h(y_\infty))$ we have $\hat{G}'(r) = -f(r) < 0$ and at $\ell := e^{-1}(0) = Y_{0-} - \Theta_{0-}$ we have

$\hat{G}'(\ell) = -f(\ell) + \hat{g}'(0) = -f(\ell) + 1$. Hence, if $Y_{0-} \le \Theta_{0-}$ we find $\hat{G}'(\ell) \ge 0$. So $\hat{G}$ obtains a global maximum at some $y^* \in [\ell, r]$ in that case and therefore $e(y^*) \in [h(y_\infty), 0]$, where $g = \hat{g}$, if $Y_{0-} \le \Theta_{0-}$.

Consider the so-called (as in [PSS11]) *type A strategy* Θ^* that performs an initial block trade $\Delta\Theta_0^* = Y_0 - Y_{0-}$ to reach impact level $Y_0 := h^{-1}\big(e(y^*)\big)$, then trades continuously until time T at constant rate $\mathrm{d}\Theta_t^*/\mathrm{d}t = h(Y_0) = e(y^*)$, and finishes with a block trade of size $\Delta\Theta_T^* = y^* - Y_0$, reaching impact level $Y_T = y^*$. By construction, we have $\mathbb{E}[L_T(\Theta^*)] = \hat{G}(y^*)$, so Θ^* is optimal (if $Y_{0-} \le \Theta_{0-}$).

Remark 2.2.11 (Optimal execution)**.** How to optimally acquire an asset position, minimizing the expected costs, is the natural counterpart to the previous liquidation problem; cf. [PSS11]. To this end, if we represent the admissible strategies by increasing càdlàg processes Θ starting at 0 (describing the cumulative number of shares purchased over time), then the discounted costs (negative proceeds) of an admissible (purchase) strategy Θ takes the form

$$\int_0^\infty e^{\eta t} f(Y_{t-}) M_t \,\mathrm{d}\Theta_t^c + \sum_{\substack{t \ge 0 \\ \Delta\Theta_t \neq 0}} e^{\eta t} M_t \int_0^{\Delta\Theta_t} f(Y_{t-} + x)\,\mathrm{d}x, \tag{2.27}$$

with discounted unaffected price process $e^{-\gamma t}\overline{S}_t = e^{\eta t} M_t$ for $\eta := \mu - \gamma = -\delta$. To have a well-posed minimization problem for infinite horizon, one needs to assume that the price process increases in expectation, i.e. $\eta > 0$, and thus the trader aims to buy an asset with rising (in expectation) price.

In this case, the value function of the optimization problem will be described by the variational inequality $\min\{f + V_y - V_\theta \ , \ \eta V - hV_y\} = 0$. An approach as taken previously for the optimal liquidation problem permits again to construct the classical solution to this free boundary problem explicitly. Thereby, the state space is divided into a wait region and a buy region by the free boundary, that is described by

$$\theta'(y) = -1 + \frac{h(y)\lambda(y)}{\eta} - \frac{h(y)h''(y)}{\eta h'(y)} + \frac{h(y)(h\lambda + h' - \eta)'(y)}{\eta(h\lambda + h' - \eta)(y)}, \qquad y \ge y_0, \tag{2.28}$$

with initial condition $\theta(y_0) = 0$, where y_0 is the unique root of $h(y)\lambda(y) = \eta$ (similar to (2.12) from the optimal liquidation problem). Verification of optimality will go though under the assumptions $\eta > 0$, $f \in C^2$ with $f(0) = 1$, $\lambda(y) := f'(y)/f(y) > 0$, resilience $h \in C^2$ with $h(0) = 0$, $h' > 0$, the technical condition $(h')^2 > hh''$ and such that $(h\lambda)' > 0$ and $(h\lambda + h')' > 0$. Note that apart from $\eta > 0$ and $(h')^2 > hh''$, these match Assumption 2.2.2. Examples satisfying $(h')^2 > hh''$ are $h(y) = \beta y$ and $h(y) = \alpha \arctan(\beta y)$ for $\alpha, \beta > 0$.

It may be interesting to note that the boundary defined by (2.28) does not have a vertical asymptote, because such an asymptote could only occur at a root y_∞ of the denominator $h\lambda + h' - \eta$, but $y_\infty < y_0$ and $\theta'(y_0) = \big(h\,(h\lambda)'/(\eta h')\big)(y_0) > 0$. The technical condition $(h')^2 > hh''$ guarantees that the boundary is strictly increasing for all $y \ge y_0$.

2.3 The problem case for non-monotone strategies

In this section, we solve under Assumption 2.2.2 the optimal liquidation problem when the admissible liquidation strategies allow for intermediate buying. To focus again on transient price impact and explicit analytical results, we keep other model aspects simple and consider the problem in a two-sided order book model with zero bid-ask spread. This is an idealization of the predominant one-tick-spread that is observed for common relatively liquid risky assets [CdL13]. See Remark 2.3.6 though. We show that the optimal trading strategy is monotone when Y_{0-} is not too small (see Remark 2.3.5). More precisely, the two-dimensional state space decomposes into a buy region and a sell region with a non-constant interface, that coincides with the free boundary from Theorem 2.2.4.

In previous sections, we considered pure selling strategies and specified the model for such, i.e. in the sense of Remark 2.1.1 we specified only the bid side of the LOB. Now, we extend the model to allow for buying as well. In this case, a large investor's trading strategy may be described by a pair of increasing càdlàg processes (A^+, A^-) with $A^{\pm}_{0-} = 0$, where A^+_t (resp. A^-_t) describes the cumulative number of assets sold (resp. bought) up to time t. Her risky asset position is $\Theta_t = \Theta_{0-} - (A^+_t - A^-_t)$ at time $t \geq 0$. We assume that the price impact process $Y = Y^\Theta$ is given by (2.2) with $\Theta = \Theta_{0-} - (A^+ - A^-)$, and that the best bid and ask prices evolve according to the same process $S = f(Y^\Theta)\overline{S}$, i.e. the bid-ask spread is taken as zero. The proceeds from executing a market buy order at time t of size $\Delta A^-_t > 0$ are given again by (2.5) with $\Delta\Theta_t = \Delta A^-_t$. Proceeds being negative means that the trader pays for acquired assets. Thus, the γ-discounted (cumulative) proceeds from trading strategy (A^+, A^-) over time period $[0, T]$ are

$$L_T = -\int_0^T e^{-\gamma t} f(Y_t)\overline{S}_t \,\mathrm{d}\Theta^c_t - \sum_{\substack{\Delta\Theta_t \neq 0\\ t\leq T}} e^{-\gamma t}\overline{S}_t \int_0^{\Delta\Theta_t} f(Y_{t-} + x)\,\mathrm{d}x\,. \tag{2.29}$$

For finite variation strategies Θ the sum in (2.29) converges absolutely, cf. Remark 2.2.1.

We consider the optimization problem over the set of admissible trading strategies

$$\mathcal{A}_{\mathrm{bv}}(\Theta_{0-}) := \big\{\Theta = \Theta_{0-} - (A^+ - A^-) \mid A^{\pm} \text{ is increasing, càdlàg, predictable, bounded, with } A^{\pm}_{0-} = 0 \text{ and } \Theta_t \geq 0 \text{ for } t \geq 0\big\}, \tag{2.30}$$

where $A = A^+ - A^-$ denotes the minimal decomposition for a process A of finite (here even bounded) variation; the last condition means that short-selling is not allowed.

For an admissible strategy $\Theta \in \mathcal{A}_{\mathrm{bv}}(\Theta_{0-})$, $L_T(y;\Theta)$ as defined in (2.8), but extended to general bounded variation strategies by (2.29), describes the proceeds from strategy Θ until time T. These proceeds are a.s. finite for every $T \geq 0$, see Remark 2.2.1. To show that $\lim_{T\to\infty} L_T(y;\Theta)$ exists in L^1, let $L(y;\Theta) = L^+(y;\Theta) - L^-(y,\Theta)$ be the minimal decomposition of the (finite variation) process $L(y;\Theta)$, i.e. $L^+(y;\Theta)$ are the proceeds from selling (according to A^+), while $L^-(y;\Theta)$ are the expenses for buying (according to A^-). For Θ in $\mathcal{A}_{\mathrm{bv}}(\Theta_{0-})$, the processes $A^{\pm}$ and $f(Y)$ are bounded by some constant C. By a change of measure argument, as in Remark 2.2.6, we obtain

$$\mathbb{E}\big[|L^{\pm}_T(y;\Theta) - L^{\pm}_t(y;\Theta)|\big] \leq e^{-\delta t} C^2 M_0 \quad \text{for all } t \leq T < \infty,$$

with $\delta = \gamma - \mu > 0$ and M_0 in $(0, +\infty)$. Hence $\big(L^{\pm}_T(y;\Theta)\big)_{T\geq 0}$ are Cauchy sequences in L^1, so they converge in L^1 for $T \to \infty$ to some limits $L^{\pm}_\infty(y;\Theta) \in L^1$, and also almost surely (limits being monotone and finite). In particular, the difference $\lim_{T\to\infty} L_T(y;\Theta) = L^+_\infty(y;\Theta) - L^-_\infty(y;\Theta) =: L_\infty(y;\Theta)$ exists in L^1.

Remark 2.3.1. The functional $\Theta \mapsto \mathbb{E}[L_T(y;\Theta)]$ is not concave. For example, let $f(y) = e^{\lambda y}$, $h(y) = \beta y$ and Θ^1 be the strategy that linearly interpolates between $\Theta^1_0 = 0$, $\Theta^1_{t_1} = \theta$, and $\Theta^1_T = 0$. It is not difficult to find parameters $t_1, \theta, \mu, \lambda, \beta, Y_{0-}$ for which already a plot reveals that the interpolation $\alpha \mapsto \mathbb{E}[L_T(\alpha\Theta^1 + (1-\alpha)\Theta^0)]$, $0 \le \alpha \le 1$, between Θ^1 and the trivial strategy $\Theta^0 = 0$ is clearly non-concave.

So, the gain functional $J(y;\Theta)$ for the optimal liquidation problem with possible intermediate buying,

$$\max_{\Theta \in \mathcal{A}_{\mathrm{bv}}(\Theta_{0-})} J(y;\Theta) \qquad \text{for} \qquad J(y;\Theta) := \mathbb{E}[L_\infty(y;\Theta)]\,, \tag{2.31}$$

is well-defined. By arguments as in Section 2.2 (cf. Proposition 2.2.8 and (2.16)) one sees that in this case it suffices to find a classical solution to the following problem

$$V_y + V_\theta = f \qquad \text{on } \mathbb{R} \times [0,\infty), \tag{2.32}$$

$$-\delta V - h(y) V_y \le 0 \qquad \text{on } \mathbb{R} \times [0,\infty), \tag{2.33}$$

with suitable boundary conditions, ensuring that a classical solution exists and that the (super-)martingale properties from Proposition 2.2.8 extend to $[0-,T]$, cf. Remark 2.2.9. The optimal liquidation strategy then can be described by a sell region and a buy region, divided by a boundary.

The sell region turns out to be the same as for the problem without intermediate buying in Section 2.2, i.e. the region $\mathcal{S}$, while the wait region $\mathcal{W}$ there becomes a buy region $\mathcal{B} := \mathbb{R} \times [0,\infty) \setminus \overline{\mathcal{S}}$ here. Similarly to Remark 2.2.6, we extend the definition of $\Delta(y,\theta)$ to $\mathcal{B}$. For $(y,\theta) \in \mathbb{R} \times [0,\infty)$, let $\Delta(y,\theta)$ be the signed $\|\cdot\|_\infty$ distance in direction $(-1,-1)$ of the point (y,θ) to the boundary $\partial\mathcal{S} = \{(\mathrm{y}(\theta),\theta) \mid \theta \ge 0\} \cup \{(y,0) \mid y \ge y_0\}$, i.e. $(y-\Delta, \theta-\Delta) \in \partial S$. Recall the definition of $V^{\mathcal{S}} := V$ in (2.14) and let

$$V^{\mathcal{B}}(y,\theta) := V_{\mathrm{bdry}}(\theta - \Delta(y,\theta)) - \int_y^{y-\Delta(y,\theta)} f(x)\,\mathrm{d}x, \quad \text{for } (y,\theta) \in \mathcal{B}.$$

The discussion so far suggests that the following function would be a classical solution to the problem (2.32) – (2.33) describing the value function of the optimization problem (2.31):

$$V^{\mathcal{B},\mathcal{S}}(y,\theta) := \begin{cases} V^{\mathcal{S}}(y,\theta), & \text{if } (y,\theta) \in \overline{\mathcal{S}}, \\ V^{\mathcal{B}}(y,\theta), & \text{if } (y,\theta) \in \mathcal{B}, \end{cases} \tag{2.34}$$

up to the multiplicative constant $\overline{S}_0$. Note that both cases in (2.34) can be combined to

$$V^{\mathcal{B},\mathcal{S}}(y,\theta) = V_{\mathrm{bdry}}(\theta - \Delta(y,\theta)) + \int_{y-\Delta(y,\theta)}^{y} f(x)\,\mathrm{d}x, \quad \text{for all } (y,\theta).$$

The next theorem proves the conjectures already stated in this section for solving the optimal liquidation problem with possible intermediate buying.

Theorem 2.3.2. *Let the model parameters h, λ, δ satisfy Assumption 2.2.2. Consider functions τ, y and $\boldsymbol{\theta}$ from Theorem 2.2.4 and let*

$$\Delta_0 := \inf\{d \in [-\Theta_{0-},\infty) \mid Y_{0-} + d = \mathrm{y}(\Theta_{0-} + d)\} \vee -\Theta_{0-}\,,$$

$$T := \tau\big(\mathrm{y}(\Theta_{0-} + \Delta_0)\big)\,.$$

For given number of shares $\Theta_{0-} \geq 0$ to liquidate and initial state of the market impact process $Y_{0-} = y$, the unique optimal strategy Θ^{opt} of problem (2.9) *is given by* $\Theta_t^{opt} = \mathrm{y}(\bar{y}(T-t))\mathbb{1}_{[0,T)}(t)$ *for* $t \geq 0$.

Moreover, the function $V^{\mathcal{B},\mathcal{S}}$ is in $C^1(\mathbb{R} \times [0,\infty))$ and solves (2.32) *and* (2.33) *and the value function of the optimization problem* (2.31) *is given by* $\overline{S}_0 \cdot V^{\mathcal{B},\mathcal{S}}$.

For the proof of Theorem 2.3.2, see [BBF18b, Thm. 5.1]. By continuity arguments as in Chapter 5, one could show that the optimal strategy of Theorem 2.3.2 is even optimal in a set of bounded semimartingale strategies (to which the definition of proceeds can be extended continuously in certain topologies on the càdlàg space, see [BBF19, Example 5.2]).

Remark 2.3.3 (The optimal buy-and-sell strategy)**.** If $(y, \Theta_{0-}) \in \overline{\mathcal{S}}$, Θ^{opt} is the liquidation strategy for Θ_{0-} shares and impact process starting at y as described in Theorem 2.2.4. If $(y, \Theta_{0-}) \in \mathcal{B}$, Θ^{opt} consists of an initial buy order of Δ_0 shares (so that the state process (Y, Θ) jumps at time 0 to the boundary between $\mathcal{B}$ and $\mathcal{S}$) and then Θ^{opt} continues according to the liquidation strategy for $\Theta_{0-} + \Delta_0$ shares and impact process starting at $y + \Delta_0$ as described in Theorem 2.2.4.

Remark 2.3.4 (Transient impact is essential)**.** As already noted in [GZ15, Prop. 3.5(III)], a multiplicative order book with permanent instead of transient impact, i.e. $h \equiv 0$, leads to a trivial optimal control with complete initial liquidation at time 0, in absence of transaction costs. This can also be seen directly as follows. If $h \equiv 0$ we have $Y_t^{\Theta} = Y_{0-} - \Theta_{0-} + \Theta_t$ and proceeds (2.8) may be written as

$$\begin{aligned} L_T(\Theta) &= \int_0^T F(Y_t^{\Theta})\,\mathrm{d}\big(e^{-\delta t} M_t\big)_t - \big(e^{-\delta T} M_T F(Y_T^{\Theta}) - M_0 F(Y_{0-})\big) \qquad (2.35)\\ &= -\delta \int_0^T e^{-\delta t} M_t F(Y_t^{\Theta})\,\mathrm{d}t + \int_0^T e^{-\delta t} F(Y_t^{\Theta})\,\mathrm{d}M_t - \big(e^{-\delta T} M_T F(Y_T^{\Theta}) - M_0 F(Y_{0-})\big) \end{aligned}$$

with the antiderivative $F(y) := \int_{-\infty}^{y} f(x)\,\mathrm{d}x \geq 0$ of f, assuming $F(0) < \infty$. So we get for any two strategies Θ and $\hat{\Theta}$ with $\Theta_t \geq \hat{\Theta}_t$ for all $t \geq 0$, that $\mathbb{E}[L_T(\Theta)] \leq \mathbb{E}[L_T(\hat{\Theta})]$. Thus it is optimal to liquidate all assets at time 0, because $\hat{\Theta}_t := 0 \leq \Theta_t$ for all $t \geq 0$ and $\Theta \in \mathcal{A}_{\mathrm{bv}}(\Theta_{0-})$. Equation (2.35) moreover shows that in the case of no drift ($\delta = 0$) and permanent impact, *every* strategy that liquidates until time T is optimal. This was already observed in [GZ15, comment before Prop. 3.5] and shows a remarkable difference of effects from permanent and transient impact; cf. also Remark 2.2.10.

Remark 2.3.5 (Price manipulation)**.** The results show that when the initial level of market impact is sufficiently small, i.e. $Y_{0-} < y_0$, so that the market price is sufficiently depressed and has a strong upwards trend by (2.2), then the optimal liquidation strategy may comprise an initial block buy, followed by continuous selling of the risky asset position. In this sense our model admits *transaction-triggered price manipulation* in the spirit of [ASS12, Definition 1] for sufficiently small $Y_{0-} < y_0$. Let us note that [LS13, p. 745] emphasize the particular relevance of the martingale case (zero drift) when analyzing (non)existence of price manipulation strategies, and that it seems natural to buy an asset whose price tends to rise. The case $Y_{0-} < 0$ could be considered as adding an exogenous but non-transaction triggered upward component to the drift. In any case, buying could only occur at initial time $t = 0$ and afterwards the optimal strategy is just selling. Nonetheless, for typical choices of the unperturbed price process $\overline{S}$ (e.g. exponential Brownian motion) one can show that our model does not offer arbitrage opportunities (in the usual sense) for the large trader, and so strategies, whose expected proceeds are strictly positive, have to admit negative proceeds (i.e. losses) with positive probability, see [BBF19, Section 4].

On the other hand, if the level of market impact is not overly depressed, i.e. $Y_{0-} \geq y_0$, then an optimal liquidation strategy will never involve intermediate buying. This includes in particular the case of a neutral initial impact $Y_{0-} = 0$ (as in [PSS11]), or of an only mildly depressed initial impact $Y_{0-} \in [y_0, \infty)$. Monotonicity of the optimal strategy would extend to cases with non-zero bid-ask spread, as explained below.

Remark 2.3.6 (On non-zero bid-ask spread)**.** The results in this section also have implications for models with non-zero bid-ask spread. Indeed, if the initial market impact is not too small ($Y_{0-} \geq y_0$) and the LOB bid side is described as in our model, the optimal liquidation strategy in a model with non-zero bid-ask spread would still be monotone (so relate only to the LOB bid side) and would be described by Theorem 2.3.2, since

$$\sup_{\Theta \in \mathcal{A}_{\mathrm{mon}}(\Theta_{0-})} J(Y_{0-};\Theta) = \sup_{\Theta \in \mathcal{A}_{\mathrm{bv}}(\Theta_{0-})} J(Y_{0-};\Theta) \geq \sup_{\Theta \in \mathcal{A}_{\mathrm{bv}}(\Theta_{0-})} J^{\mathrm{spr}}(Y_{0-};\Theta),$$

with $J^{\mathrm{spr}}(Y_{0-};\Theta)$ denoting the cost functional for the non-zero spread model, as $J(Y_{0-},\cdot)$ and $J^{\mathrm{spr}}(Y_{0-},\cdot)$ coincide on $\mathcal{A}_{\mathrm{mon}}(\Theta_{0-})$ and the inequality is due to the spread.

3 Optimal execution with price trends – a three-dimensional free boundary problem

In this chapter, we solve an optimal liquidation problem in finite time horizon in an environment where the fundamental price has increasing or decreasing trend by means of a nonzero drift factor. The martingale case could be solved with convexity arguments à la [PSS11], as explained in Remark 2.2.10. That argument does not work for nonzero drift, where the objective functional $\Theta \mapsto \mathbb{E}[L_T(\Theta)]$ is non-concave (cf. Remark 2.3.1) and I cannot see another convexity structure to exploit for easier verification. So we prove optimality through a different method in two steps in Section 3.3. First, calculus of variations gives a candidate optimal strategy that is characterized by a smooth boundary surface separating buy and sell regions. This candidate satisfies a local optimality criterion in the sense of Theorem 3.3.8 among strategies characterized by smooth boundary surfaces. This local result implies validity of the variational inequality in a neighborhood of our candidate surface, cf. Lemma 3.3.10, which we can extend in a second step to the whole state space and thereby prove global optimality among all bounded variation strategies.

In Section 3.1, we formulate the financial market model and optimization objective. We discuss our main result Theorem 3.2.1 in Section 3.2. Section 3.3 is devoted to the proof, from construction of the candidate solution and local optimality in Theorem 3.3.8, via a necessary reparametrization of the state space in Section 3.3.1, to the verification in Section 3.3.3. As a consequence of this result, suitably combined with the optimal sell-only liquidation strategy for infinite time horizon, cf. Theorem 2.2.4, we obtain in Section 3.4 the optimal sell-only liquidation strategy in finite time horizon, when prices are generally decreasing, see Theorem 3.4.1.

3.1 The model and optimization objective

In absence of the large trader, the unaffected (fundamental) price is like in Chapter 2 of the form

$$\overline{S}_t = e^{\mu t} M_t, \qquad \overline{S}_0 \in (0, \infty), \tag{3.1}$$

for constant $\mu \in \mathbb{R}$ and M being a non-negative square integrable martingale on $[0, T]$ that quasi-left continuous. The drift factor μ allows to model beliefs about the short time price dynamics. The prime example being a geometric Brownian motion, $M = \overline{S}_0 \mathcal{E}(\sigma W)$ for a Brownian motion W and volatility $\sigma > 0$, as in the Black-Scholes model.

The large investor's strategy is her position Θ_t in the risky asset, starting with an exogenously given amount $\Theta_{0-} \in \mathbb{R}$. The predictable càdlàg process $(\Theta_t)_{t \in [0,T]}$ is of bounded variation. Her trading activity causes transient market impact, which we denote by the process $Y = Y^{\Theta}$. It follows the dynamics

$$\mathrm{d}Y_t = -h(Y_t)\,\mathrm{d}t + \mathrm{d}\Theta_t \tag{3.2}$$

starting from some given initial impact $Y_{0-} \in \mathbb{R}$. Let the *resilience function* $h : \mathbb{R} \to \mathbb{R}$ be strictly increasing with $h(0) = 0$. This way, the large investors impact reverts back towards zero with resilience rate $h(Y_t)$ whenever she does not trade. For example, with $h(y) = \beta y$, the market recovers at exponential rate $\beta > 0$.

The actual risky asset price is affected by the large investors trading activity through the impact process Y as

$$S_t = f(Y_t)\overline{S}_t\,, \tag{3.3}$$

with an increasing *impact function* $f : \mathbb{R} \to (0,\infty)$. We consider a market with zero bid-ask spread. Proceeds from a market sell order of size $\Delta\Theta_t \in \mathbb{R}$ at time t are given by

$$-\overline{S}_t \int_0^{\Delta\Theta_t} f(Y_{t-} + x)\,\mathrm{d}x, \tag{3.4}$$

which can be motivated by a limit order book perspective, cf. Remark 2.1.1, or by stability considerations, cf. Chapter 5. For general bounded variation strategies (Θ_t) with decomposition into continuous and discontinuous part $\Theta_t = \Theta_t^c + \sum_{u\le t:\Delta\Theta_u\neq 0} \Delta\Theta_u$, proceeds from trading are

$$L_T(\Theta) := -\int_0^T f(Y_t^{\Theta})\overline{S}_t\,\mathrm{d}\Theta_t^c - \sum_{\substack{0\le t\le T\\ \Delta\Theta_t\neq 0}} \overline{S}_t \int_0^{\Delta\Theta_t} f(Y_{t-}^{\Theta} + x)\,\mathrm{d}x\,. \tag{3.5}$$

The large investors seeks to maximize expected proceeds $\mathbb{E}[L_T(\Theta)]$ while liquidating her position, $\Theta_T = 0$. The set of admissible strategies is

$$\mathcal{A}_T := \big\{\Theta \;\big|\; (\Theta_t)_{t\in[0,T]} \text{ is predictable càdlàg, has bounded variation and } \Theta_T = 0\big\}. \tag{3.6}$$

So for given $Y_0- = y$ and $\Theta_{0-} = \theta$, our objective reads

$$\max_{\Theta\in\mathcal{A}_T} J(T,y,\theta;\Theta) \quad \text{for} \quad J(T,y,\theta;\Theta) := \mathbb{E}[L_T(\Theta) \mid Y_0- = y, \Theta_{0-} = \theta], \tag{3.7}$$

$$\text{with value function } v(T,y,\theta) := \max_{\Theta\in\mathcal{A}_T} J(T,y,\theta;\Theta). \tag{3.8}$$

Note that predictability of Θ guarantees that $\overline{S}$ and Θ have no common jumps, as the large investor could exploit such to her favor. If we take $\overline{S}$ to be continuous, adaptability of Θ would suffice. To make our model assumptions concrete we have the following standing assumptions.

Assumption 3.1.1. The resilience function $h : \mathbb{R} \to \mathbb{R}$ is in C^3 with $h(0) = 0$ and $h' > 0$.
The impact function $f : \mathbb{R} \to (0,\infty)$ is in C^3 with $\lambda(y) := f'(y)/f(y) > 0$ everywhere

and such that $\lim_{y\to-\infty} f(y) = 0$ and $\lim_{y\to\infty} f(y) = \infty$.
There exist y_∞ and y_0 with $(h\lambda + h' - \mu)(y_\infty) = 0$ and $(h\lambda - \mu)(y_0) = 0$, respectively. Resilience and market impact satisfy $(h\lambda)' > 0$ and $(h\lambda + h')' > 0$ everywhere. Moreover, we require $h'' < (h\lambda)'h'/(h\lambda - \mu)$ on (y_0, ∞).

An example satisfying Assumption 3.1.1 is $f(y) = e^{\lambda y}$, $h(y) = \beta y$ for constant $\lambda, \beta > 0$. Note that the upper bound for h'' on (y_0, ∞) is equivalent to $q(y) := (h\lambda + h' - \mu)(y)/h'(y)$ being increasing on $[y_\infty, \infty)$, see Lemma 3.1.3. It is needed in Lemma 3.3.1. In comparison to Assumption 2.2.2 for the infinite horizon problem, we require surjectivity and more regularity for $f : \mathbb{R} \to (0, \infty)$ and h here in order to calculate the second variation of the proceeds functional in Theorem 3.3.8.

Remark 3.1.2. As in Remark 2.2.1, the possibly infinite sum in (3.5) has finite expectation, since $\mathbb{E}[\sup_{t\in[0,T]} \overline{S}_t] < \infty$ and Y^Θ is bounded for any (bounded variation) Θ.

Lemma 3.1.3. *The function $q(y) := (h\lambda + h' - \mu)(y)/h'(y)$ satisfies $q(y) < 0$ for $y < y_\infty$, $q(y_\infty) = 0$, $q(y_0) = 1$, and $q' > 0$ on $[y_\infty, \infty)$.*

Proof. Since $h' > 0$ and $(h\lambda + h')' > 0$, the definition of y_∞ gives $q(y) \lesseqgtr 0$ for $y \lesseqgtr y_\infty$ and the definition of y_0, together with $(h\lambda)' > 0$ gives $q(y) \lesseqgtr 1$ for $y \lesseqgtr y_0$. For $y > y_0$, we have $(h')^2 q' = (h\lambda)'h' - (h\lambda - \mu)h'' > 0$. It remains to show $q' > 0$ on $[y_\infty, y_0]$. On that interval, we have

$$(h')^2 q' = (h\lambda + h' - \mu)'h' - (h\lambda + h' - \mu)h''.$$

So if $h''(y) < 0$, then $q'(y) > 0$ for $y \in [y_\infty, y_0]$, because $h\lambda + h' - \mu \geq 0$ there. On the other hand, we have

$$(h')^2 q' = (h\lambda - \mu)'h' + \cancel{h''h'} - (h\lambda - \mu)h'' - \cancel{h'h''}.$$

So $h''(y) \geq 0$ also implies $q'(y) > 0$ for $y \in [y_\infty, y_0]$, because $h\lambda - \mu \leq 0$ there. □

3.2 The optimal liquidation strategy for finite horizon

In this section, we will describe the main result of the chapter and discuss the basic road map for verification. We will consider time backwards, as remaining time time to liquidation $\tau = T - t$. It turns out that the state space $\{(\tau, y, \theta) \in [0, T] \times \mathbb{R}^2\}$ of remaining time τ, current impact level y and current asset position θ consists of a buy region $\mathcal{B}$ and a sell region $\mathcal{S}$, separated by a smooth boundary surface $\mathcal{I}$. The optimal strategy will perform an initial block trade to reach $\mathcal{I}$, then trade continuously in rares along $\mathcal{I}$ until time T. At terminal time, $\tau = 0$, there is no choice but to clear the position, $\Theta_T = 0$, with a block trade. Yet, the optimal strategy will reach a state $(0, Y_{T-}, \Theta_{T-}) \in \mathcal{I}$ that satisfies $Y_{T-} > y_\infty$ and $\Theta_{T-} = g(Y_{T-})$ with *terminal impact function*

$$g(y) := y - f^{-1}\left(\left(f\frac{h\lambda + h' - \mu}{h'}\right)(y)\right), \tag{3.9}$$

for $y > y_\infty$. Now we can represent $\mathcal{I}$ as the union of orbits of suitable curves $\tau \mapsto (\tau, \bar{y}(\tau; z), \bar{\theta}(\tau; z))$, $\tau \in [0, T]$, that reach $(0, z, g(z))$, $z \in (y_\infty, \infty)$. These curves satisfy differential equations $\bar{y}_\tau = D(\bar{y})$ and, in order to stay on $\mathcal{I}$ by trading in rates, $\bar{\theta}_\tau = D(\bar{y}) - h(\bar{y})$, where

$$D(y) := \mu\left(\frac{f\,(h\lambda + h' - \mu)/h'}{(f\,(h\lambda + h' - \mu)/h')'}\right)(y), \tag{3.10}$$

for $y > y_\infty$. Such representation $\mathcal{I} = \{(\tau, \bar{y}(\tau; z), \bar{\theta}(\tau; z)) \mid \tau \in [0,T], z > y_\infty\}$ leads to a reparametrization of the whole state space $(\tau, y, \theta) = P(\tau, z, d)$ in terms of terminal impact (before the last block trade) z and initial jump $d \in \mathbb{R}$.

Moreover, the boundary surfaces $\mathcal{I}_{T_1}$ and $\mathcal{I}_{T_2}$ for different time horizons $T_1 < T_2$ coincide for $\tau < T_1$ in the sense that $\mathcal{I}_{T_1} = \mathcal{I}_{T_2} \cap [0, T_1] \times \mathbb{R}^2$.

To summarize this description, our main result for this chapter is as follows.

Theorem 3.2.1. *Let f, λ, h, μ satisfy Assumption 3.1.1. Define y_∞ and y_0 as the unique solutions to $h(y_\infty)\lambda(y_\infty) + h'(y_\infty) - \mu = 0$ and $h(y_0)\lambda(y_0) - \mu = 0$, respectively. Consider the unique solutions $\bar{y}(\cdot; z), \bar{\theta}(\cdot; z)$ on $[0,T]$ to the system of differential equations*

$$\begin{cases} \bar{y}_\tau(\tau; z) = D(\bar{y}(\tau; z)), & \text{for } \tau \in [0,T], \\ \bar{\theta}_\tau(\tau; z) = (D - h)(\bar{y}(\tau; z)), & \text{for } \tau \in [0,T], \\ \bar{y}(0; z) = z, \\ \bar{\theta}(0; z) = g(z), \end{cases}$$

for $z > y_\infty$, with g from (3.9) and D from (3.10).

Then $P(\tau, z, d) := (\tau, \bar{y}(\tau; z) + d, \bar{\theta}(\tau; z) + d)$ is a bijection from $[0,T] \times (y_\infty, \infty) \times \mathbb{R}$ to $[0,T] \times \mathbb{R}^2$. Moreover, let $(T, z^*, d^*) := P^{-1}(T, Y_{0-}, \Theta_{0-})$ and define $\Theta^*_t := \bar{\theta}(T - t, z^*)$, for $t \in [0,T)$, and $\Theta^*_T := 0$. Then Θ^* is the maximizer of our objective problem (3.7),

$$\mathbb{E}[L_T(\Theta^*)] = \max_{\Theta \in \mathcal{A}_T} \mathbb{E}[L_T(\Theta)].$$

We defer the proof to the end of Section 3.3 on page 47. It amounts to an application of martingale optimality (Proposition 3.2.5) by proving variational (in)equalities (3.12) – (3.14). In Section 3.3, we first restrict our attention to strategies Θ that are continuously differentiable on $(0,T)$ with possible jumps at 0 and T. Such strategies can be described by the dynamics of Y^Θ_t alone, i.e. by $\bar{y}$ instead of $\bar{\theta}$. Rewriting our objective as a calculus of variations problem, we obtain $\bar{y}$ as a local optimizer, cf. Theorem 3.3.8. Global existence and uniqueness of $\bar{y}$ is shown in Lemma 3.3.2. The proceeds for the corresponding strategy (Θ^*) define a candidate value function V for our original control problem. By construction, V then already satisfies the variational equality (3.14). After proving in Section 3.3.1 that our reparametrization P of the state space is indeed bijective, we can prove the variational equality (3.12) on the boundary $\mathcal{I}$ directly, cf. Lemma 3.3.9, and show that the local optimality result for $\bar{y}$ implies the variational inequality (3.13) in a neighborhood of $\mathcal{I}$, see Lemma 3.3.10. Validity of (3.13) in the whole state space follows as a consequence, cf. Lemma 3.3.9.

Let us now comment on some properties of the optimal strategy Θ^* from Theorem 3.2.1 and afterwards work out the details of the variational equalities and inequality.

Remark 3.2.2. Buy and sell regions are determined by the sign of d^*, i.e. he have $\mathcal{B} = P([0,T] \times (y_\infty, \infty) \times (-\infty, 0))$ and $\mathcal{S} = P([0,T] \times (y_\infty, \infty) \times (0, \infty))$, respectively. The optimal control Θ^* acts as follows:

1. If $(T, Y_{0-}, \Theta_{0-}) \in \mathcal{B}$, perform a block buy of size $|d|$ to the boundary surface $\mathcal{I} := P([0,T] \times (y_\infty, \infty) \times \{0\})$. In case $(T, Y_{0-}, \Theta_{0-}) \in \mathcal{S}$, do a block sale of size $|d|$ to the surface $\mathcal{I}$.
2. Now that $(T, Y_0^{\Theta^*}, \Theta_0^*) = P(T, z, 0)$, trade in rates $\mathrm{d}\Theta_t^* = -\bar{\theta}_\tau(T-t; z)\,\mathrm{d}t$, so $(T-t, Y_t^{\Theta^*}, \Theta_t^*) = P(T-t, z, 0)$ stays on the boundary $\mathcal{I}$ for all $t \in [0, T)$.
3. At terminal time T, when $Y_{T-}^{\Theta^*} = z$ and $\Theta_{T-}^* = g(z)$, perform a block trade (buy or sale) of size $-g(z)$ to clear the position, $\Theta_T^* = 0$.

Remark 3.2.3 (Large time horizons)**.** For large enough horizon T and $\mu < 0$, i.e. decreasing prices in expectation, the optimal non-short-selling strategy is the infinite horizon solution of Theorem 2.3.2 which terminates in finite time $\tau(Y_{0-}, \Theta_{0-}) \le T$ (Note that Assumption 2.2.2 follows from Assumption 3.1.1 if $\mu < 0$; we require more smoothness on f and h here to apply the second variation in Theorem 3.3.8 below). If $T = \tau(Y_{0-}, \Theta_{0-})$, the above solution $\bar{y}, \bar{\theta}$ equals the infinite horizon solution, but for $T > \tau(Y_{0-}, \Theta_{0-})$, it is optimal to go short temporarily and buy back the assets at the end.

Remark 3.2.4 (Intermediate buying)**.** Depending on the model parameters, the optimal trading strategy may also require continuous buying in rates before time T. For $\mu > 0$, i.e. increasing prices in expectation, this may even happen without any short sales, see e.g. Figure 3.1a: to liquidate $\Theta_{0-} = 1$ asset with no initial impact, $Y_{0-} = 0$, and $f(y) = e^y$, $h(y) = y$ in $T = 5$ time when prices are generally increasing ($\mu = 0.2$, top red line) it is optimal to perform an initial block buy of size $\Delta\Theta_0^* \approx 0.2$, then slowly buy more assets in decreasing rates until time $t \approx 1$, subsequently selling assets in increasing rates until terminal time T, when the remaining 0.6 assets are sold en bloc.

However, for decreasing prices, i.e. $\mu < 0$, buying in rates, i.e. $-\bar{\theta}'(T-t) \ge 0$ always implies a short position $\bar{\theta}(T-t) \le 0$: Indeed, on the interval $(y_\infty, y_0]$ we have $(h\lambda - \mu)(h\lambda + h' - \mu)/h' + ((h\lambda + h' - \mu)/h')'h < 0$, since $h\lambda < \mu$ on this interval, $y_0 < 0$ for $\mu < 0$, and $(h\lambda + h' - \mu)/h'$ is positive and increasing on (y_∞, ∞). Hence buying, $-\bar{\theta}_\tau(T-t; z) \ge 0$, implies $\bar{y}(T-t; z) \ge y_0$ which leads to a short position by Lemmas 3.3.1 and 3.3.2.

Key for the proof of Theorem 3.2.1 is the following principle.

Proposition 3.2.5 (Martingale optimality principle)**.** *Let $V : [0,T] \times \mathbb{R}^2 \to \mathbb{R}$ satisfy $V(0, \cdot, 0) = 0$, such that for each $\Theta \in \mathcal{A}_T$, $G_t(\Theta) := L_t(\Theta) + \bar{S}_t V(T-t, Y_t^\Theta, \Theta_t)$ is a supermartingale and additionally $G_0(\Theta) \le G_{0-}(\Theta) := \bar{S}_0 V(T, y, \theta)$ where $Y_{0-} = y$, $\Theta_{0-} = \theta$. Then*

$$\bar{S}_0 V(T, y, \theta) \ge v(T, y, \theta).$$

Moreover, if there exists a strategy $\Theta^* \in \mathcal{A}_T$ such that $G_\cdot(\Theta^*)$ is a martingale with $G_0(\Theta^*) = G_{0-}(\Theta^*)$, then $\bar{S}_0 V(T, y, \theta) = v(T, y, \theta)$ and $v(T, y, \theta) = J(T, y, \theta; \Theta^*)$.

Proof. Since $\Theta_T = 0$, we have $\mathbb{E}[G_T(\Theta)] = \mathbb{E}[L_T] + \mathbb{E}[\bar{S}_T V(0, Y_T, 0)] = \mathbb{E}[L_T(\Theta)]$. So the supermartingale property immediately gives

$$\bar{S}_0 V(T, y, \theta) = \mathbb{E}[G_{0-}(\Theta)] \ge \mathbb{E}[G_0(\Theta)] \ge \mathbb{E}[G_T(\Theta)] = \mathbb{E}[L_T(\Theta)]$$

for all $\Theta \in \mathcal{A}_T$. Hence $\bar{S}_0 V(T, y, \theta) \ge v(T, y, \theta)$. The second part follows similarly from the martingale property. $\square$

To utilize Proposition 3.2.5, assume $V \in C^1$ and apply Itô's formula to G:

$$\begin{aligned} \mathrm{d}G_t = V(T-t, Y_{t-}, \Theta_{t-})\,\mathrm{d}M_t \\ + \overline{S}_t\Big(\big(-V_\tau - hV_y + \mu V\big)(T-t, Y_{t-}, \Theta_{t-})\,\mathrm{d}t \\ + \big(V_y + V_\theta - f\big)(T-t, Y_{t-}, \Theta_{t-})\,\mathrm{d}\Theta_t^c \\ + \int_0^{\Delta\Theta_t} \big(V_y + V_\theta - f\big)(T-t, Y_{t-}+x, \Theta_{t-}+x)\,\mathrm{d}x\Big)\,, \end{aligned} \tag{3.11}$$

with abbreviations $(-V_\tau - hV_y + \mu V)(u,a,b) := -V_\tau(u,a,b) - h(a)V_y(u,a,b) + \mu V(u,a,b)$ and $(V_y + V_\theta - f)(u,a,b) := V_y(u,a,b) + V_\theta(u,a,b) - f(a)$. Since Y^Θ and Θ are bounded and $\overline{S}$ has integrable second moments, the local martingale part of $\mathrm{d}G_t$ is a true martingale. To apply Proposition 3.2.5, we will construct a classical solution to the *variational inequality* $\min\{\mathcal{L}V,\, V_y + V_\theta - f\} = 0$ with the differential operator $\mathcal{L}V(\tau,y,\theta) := V_\tau(\tau,y,\theta) + h(y)V_y(\tau,y,\theta) - \mu V(\tau,y,\theta)$, that is, a function $V \in C^1([0,T]\times\mathbb{R}^2;\mathbb{R})$ and a function $d : [0,T]\times\mathbb{R}^2 \to \mathbb{R}$ parametrizing the *free boundary surface* $\mathcal{I} = \{d(\tau,y,\theta) = 0\}$ such that

$$\mathcal{L}V(\tau,y,\theta) = 0 \qquad \text{for } (\tau,y,\theta) \in \mathcal{I}, \tag{3.12}$$

$$\mathcal{L}V(\tau,y,\theta) > 0 \qquad \text{for } (\tau,y,\theta) \notin \mathcal{I}, \tag{3.13}$$

$$V_y(\tau,y,\theta) + V_\theta(\tau,y,\theta) = f(y) \qquad \text{for } (\tau,y,\theta) \in [0,T]\times\mathbb{R}^2, \tag{3.14}$$

$$V(0,y,0) = 0 \qquad \text{for all } y \in \mathbb{R}. \tag{3.15}$$

The optimal strategy will be to buy whenever $d(T-t, Y_t, \Theta_t) < 0$ and sell whenever $d(T-t, Y_t, \Theta_t) > 0$ in order to keep $d(T-t, Y_t, \Theta_t) = 0$ until time $t = T$ when the position is cleared with a block market order of size $\Delta\Theta_T = -\Theta_T$. It will turn out that $d(\tau,y,\theta) < 0$ for all $y \le y_\infty$, i.e. for $Y_{t-} \le y_\infty$, a block buy will cause $Y_t > y_\infty$ immediately. Let $\mathcal{B}$ denote the *buy region* and $\mathcal{S}$ denote the *sell region*,

$$\mathcal{B} = \{(\tau,y,\theta) \in [0,T)\times\mathbb{R}^2 \mid d(\tau,y,\theta) < 0\}, \tag{3.16}$$

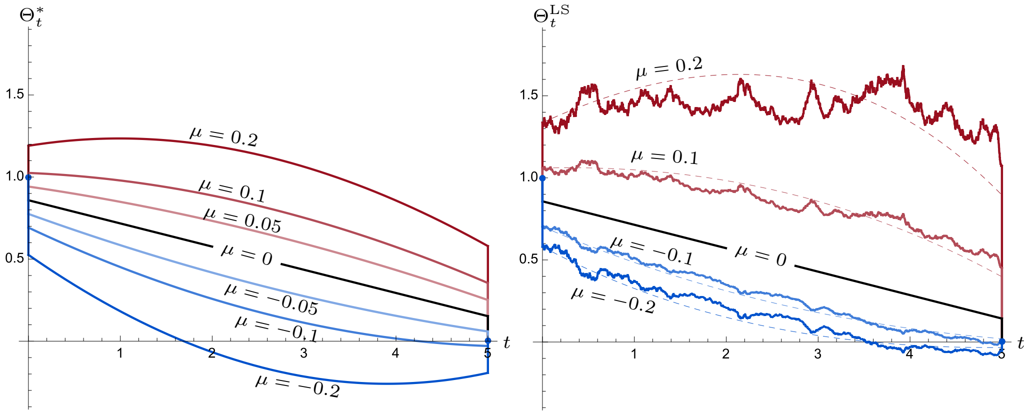
$$\mathcal{S} = \{(\tau,y,\theta) \in [0,T)\times\mathbb{R}^2 \mid d(\tau,y,\theta) > 0\}. \tag{3.17}$$

Remark 3.2.6. The optimal control is deterministic. Indeed, by optional projection [DM82, Theorem VI.57] we have

$$\begin{aligned} \mathbb{E}[L_T(\Theta)] &= \mathbb{E}\Big[M_T\Big(-\int_0^T e^{\mu t} f(Y_t)\,\mathrm{d}\Theta_t^c - \sum_{0\le t\le T} e^{\mu t}\int_0^{\Delta\Theta_t} f(Y_{t-}+x)\,\mathrm{d}x\Big)\Big] \\ &= \mathbb{E}_{\mathbb{Q}}\Big[-\int_0^T e^{\mu t} f(Y_t)\,\mathrm{d}\Theta_t^c - \sum_{0\le t\le T} e^{\mu t}\int_0^{\Delta\Theta_t} f(Y_{t-}+x)\,\mathrm{d}x\Big] =: \mathbb{E}_{\mathbb{Q}}[\ell_T(\Theta)] \end{aligned}$$

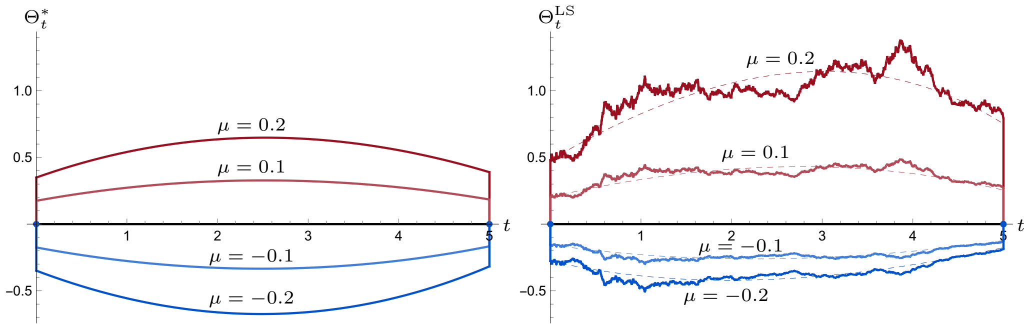
for $\mathrm{d}\mathbb{Q} = M_T\,\mathrm{d}\mathbb{P}$. Since $\mathbb{Q}$ does not depend on Θ and ℓ_T is a deterministic functional, the optimization can be done ω-wise and optimizing over deterministic controls will yield the same value function.

Remark 3.2.7 (Comparing multiplicative and additive impact)**.** Let us highlight some differences between optimal liquidation strategies for our multiplicative transient price



(a) Liquidation strategies Θ^* in our multiplicative price impact model $S_t = \overline{S}_t \exp(Y^\Theta_t)$.

(b) Liquidation strategies $\Theta^{\mathrm{LS}}_t(\omega)$ (solid lines) and expected positions $\mathbb{E}[\Theta^{\mathrm{LS}}_t]$ (dashed) in the additive price impact model $S_t = \overline{S}_t + Y^\Theta_t$.



(c) Round-trip strategies Θ^* in our model.

(d) Round-trip strategies $\Theta^{\mathrm{LS}}_t(\omega)$ (solid lines) and their expected positions $\mathbb{E}[\Theta^{\mathrm{LS}}_t]$ (dashed) in the LS-model.

Figure 3.1: Optimal strategies Θ^* in our model (left) and Θ^{LS} the LS-model (right, cf. Remark 3.2.7) to liquidate $\Theta_{0-} = 1$ asset (top row) or perform a round-trip ($\Theta_{0-} = 0$, bottom row) in $T = 5$ time when prices are generally increasing (red), generally decreasing (blue) or have no trend (black). Fundamental price is $\mathrm{d}\overline{S}_t = \mu \overline{S}_t \,\mathrm{d}t + \sigma \overline{S}_t \,\mathrm{d}W_t$ with $\overline{S}_0 = 1$, $\sigma = 0.5$ and different μ, initial impact is $Y_{0-} = 0$ and resilience is linear with speed $\beta = 1$, i.e. $\mathrm{d}Y^\Theta_t = -Y^\Theta_t \,\mathrm{d}t + \mathrm{d}\Theta_t$ in both models.

impact model and the additive transient price impact model of [LS13], which generalizes the continuous time model as in [OW13] by permitting non-zero drift for the unaffected price process. Let us call these the mLOB- and the LS-model. We take the unaffected price process in both models to be a geometric Brownian motion with drift, $\overline{S}_t = \overline{S}_0 e^{\mu t} \mathcal{E}(\sigma W)_t$ with Brownian motion W, volatility $\sigma > 0$, drift factor $\mu \in \mathbb{R}$ and initial price $\overline{S}_0 \in (0, \infty)$. The martingale case $\mu = 0$ is solved in [PSS11] via a convexity argument for additive impact, that can be adapted for multiplicative impact, cf. Remark 2.2.10. For both models, a constant rate of trading is optimal when $\mu = 0$. Consider the decomposition $\overline{S}_t = \overline{S}_0 + N_t + K_t$ into martingale part $N_t := \int_0^t \sigma \overline{S}_s \,\mathrm{d}W_s$ and finite variation part $K_t := \int_0^t \mu \overline{S}_s \,\mathrm{d}s$.

[LS13] consider zero initial impact and linear resilience, i.e. $\mathrm{d}Y_t^\Theta = -\beta Y_t^\Theta \,\mathrm{d}t + \mathrm{d}\Theta_t$ with resilience factor $\beta > 0$ and $Y_{0-} = 0$ in our notation. For a bounded semimartingale strategy $(\Theta_t)_{t\in[0,T]}$ with initial position $\Theta_{0-} = \theta$ and $\Theta_T = 0$, the price at which trading occurs in the LS-model is $S_t^\Theta := S_t^0 + \eta Y_{t-}^\Theta$, motivated by a (additive) block-shaped limit order book with height $1/\eta \in (0, \infty)$.

In a similar fashion, as described in Remark 2.1.1, we can relate our multiplicative price impact model to a multiplicative limit order book (mLOB). To compare both models, their underlying limit order books should have similar features. In particular, both order books should admit infinite market depth (LOB volume) for buy and for sell orders; and prices should initially be similar for small volume impact y, i.e. $\overline{S}_0 + \eta y \approx \overline{S}_0 f(y)$. Taking $f(y) = e^{\lambda y}$ with constant $\lambda := \eta/\overline{S}_0$ satisfies these requirements, cf. Remark 2.1.1. We take w.l.o.g. $\eta = 1$. The liquidation costs to be minimized in expectation in the LS-model are given in [LS13, Lemma 2.5] as

$$\mathcal{C}(\Theta) := \int_{[0,T]} S_{t-}^0 \,\mathrm{d}\Theta_t + [S^0, \Theta]_T + \int_{[0,T]} Y_{t-}^\Theta \,\mathrm{d}\Theta_t + \frac{1}{2}[\Theta]_T .$$

According to [LS13, Thm 2.6], the corresponding optimal semimartingale strategy which minimizes $\mathbb{E}[\mathcal{C}(\Theta)]$ is

$$\begin{aligned}\Theta_t^{\mathrm{LS}} = {} & \frac{(1+\beta(T-t))\theta - \frac{1}{2}(1+\beta t)Z_0}{2+\beta T} - \frac{1}{2}\int_{(0,t]} \varphi(s)\,\mathrm{d}Z_s + \frac{1}{2\beta}K_t' \\ & - \beta \int_0^t \left(\frac{1}{2}\int_{(0,s]} \varphi(r)\,\mathrm{d}Z_r + \frac{1}{2}K_s \right) \mathrm{d}s\,, \quad t \in [0,T),\end{aligned}$$

with $\varphi(t) = (2 + \beta\,(T-t))^{-1}$, derivative $K_t' := \mathrm{d}K_t/\mathrm{d}t = \mu \overline{S}_t$ and martingale $Z_t = \mathbb{E}\big[K_T + \beta \int_0^T K_s \,\mathrm{d}s \;\big|\; \mathcal{F}_t\big]$. With geometric Brownian motion for the unaffected price, as in the Black-Scholes model, this gives $Z_0 = \big((1-e^{\mu T})(1+\frac{\beta}{\mu}) + \beta T\big)\overline{S}_0$ and $\mathrm{d}Z_t = \big(\big(1-e^{\mu(T-t)}\big)\big(1+\frac{\beta}{\mu}\big) + \beta\,(T-t)\big)\sigma \overline{S}_t \,\mathrm{d}W_t$ if $\mu \neq 0$ and $Z_t = 0$ if $\mu = 0$. Hence, for $\mu \neq 0$, the optimal liquidation strategy Θ^{LS} in the LS-model is a non-deterministic adapted semimartingale. As noted in [LS13], it is not of finite variation. In contrast, cf. Remark 3.2.6, the optimal strategy in our mLOB-model is deterministic and of finite variation.

Consider for example the regime of generally decreasing prices, $\mu < 0$: Apart from a possible initial block buy, as long as the large investor has a long position $\Theta_t > 0$ in the asset at time $t > 0$, intermediate buying would be suboptimal in our mLOB-model, cf. Remark 3.2.4. In comparison, the semimartingale nature of Θ^{LS} requires perpetual buying, a rather counter-intuitive outcome regarding that the postulated order book shape is invariant over time and (unaffected) returns $\mathrm{d}\overline{S}/\overline{S}$ are i.i.d. In this sense, the optimal strategy in the LS-model exhibits transaction-triggered price manipulation à la [ASS12, Def. 1] (in continuous time) also for negative price trend $\mu < 0$, whereas such is not the case in our mLOB-model for moderate parameter choices where a short position is not reached in time $t < T$. For generally increasing prices, $\mu > 0$, it is natural to expect intermediate buying to possibly be optimal also at times of long position.

Figure 3.1b displays common realizations of optimal strategies in the LS-model for different fundamental price trends μ for comparison to the corresponding optimizers of the mLOB-model in Figure 3.1a.

From the point of view of an investor trading for herself, it is instructive to consider optimal round-trips, i.e. $\Theta_{0-} = 0$. The LS-model seems to suggest a generally higher intermediate buy position for positive fundamental price trends ($\mu > 0$) than the mLOB-model, cf. Figures 3.1c and 3.1d. Interestingly, the optimal trading activity also seems to be more symmetric between $\mu > 0$ and $\mu < 0$ cases in the mLOB-model than in the LS-model.

Let us note that the LS-model would give a deterministic optimal strategy (of finite variation) if the unaffected price process would be an arithmetic Brownian motion, $\mathrm{d}S_t^0 = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$. This indicates that additive impact models are better suited for additive (Bachelier) price dynamics, while a multiplicative impact model suits multiplicative (Black-Scholes) price dynamics. Note that stochastic optimal strategies would naturally occur if relevant state variables are stochastic, as we will investigate in Chapter 4.

While additive models for asset prices and price impact have the benefit of easier analysis, in particular for the martingale case without drift, we believe that multiplicative models offer benefits from a conceptual point of view. For example, shares of company are a relative concept, corresponding to a proportional amount of the companies value. It is reasonable to follow early contributions like [Jar94, Fre98] and model price impact in relative terms.

3.3 Free boundary construction and verification via calculus of variations

We will now derive the functions g and D stated in (3.9) and (3.10) that characterize the optimizer through a calculus of variations ansatz. Therefore, assume for a moment that $g \in C^1$ and $D \in C^0$ are unknown functions on to be determined domains. We verify assumed properties of g in Lemma 3.3.1 and of D in Lemma 3.3.2. For the general theory of calculus of variations methods, see e.g. the book [GF00].

We use necessary conditions from calculus of variations to identify g and D. Then, after showing in Section 3.3.1 that the reparametrization P of Theorem 3.2.1 is indeed a bijection, we apply sufficient conditions for a local optimality result in Section 3.3.2, which is necessary for the verification of global optimality in Section 3.3.3.

Heuristically, the optimal strategy Θ should consist of an initial jump from within the buy or sell region to the (yet unknown) boundary $\mathcal{I}$ the separates both, followed by continuous trading in rates $\mathrm{d}\Theta_t = -\bar{\theta}'(T-t)\,\mathrm{d}t$ for $t \in [0,T)$ until terminal time T, and a terminal jump to reach $\Theta_T = 0$. We will first restrict our focus to such strategies with possible initial and terminal jumps, and continuous rate of trading $-\bar{\theta}'$ during time $(0,T)$. If $\Theta_t = \bar{\theta}(T-t)$, $t \in [0,T)$, for some $\bar{\theta} \in C^1$, then by (3.2) the impact process $Y = Y^\Theta$ is of the form $Y_t = \bar{y}(T-t)$ for $\bar{y} \in C^1$ such that $\bar{y}'(\tau) = h(\bar{y}(\tau)) + \bar{\theta}'(\tau)$. Let us assume, that we can parametrize the boundary at terminal time ($\tau = 0$) by a C^1 function g on a to-be-determined interval such that $\Theta_{T-} = g(Y_{T-})$, i.e. $\bar{\theta}(0) = g(\bar{y}(0))$. Now, we can express the expected proceeds $\mathbb{E}[L_T(\Theta)]$ of the strategy Θ that corresponds to $\bar{\theta} \in C^1([0,T])$ solely in terms of the corresponding $\bar{y} \in C^1([0,T])$. Then we have

$\mathbb{E}[L_T(\Theta)] = J_{T,Y_{0-}}(\bar{y})$ with

$$\begin{aligned} J_{T,Y_{0-}}(\bar{y}) := F(Y_{0-}) - F\big(\bar{y}(T)\big) + e^{\mu T}\big(F(y) - F\big(y - g(y)\big)\big)\big|_{y=\bar{y}(0)} \\ + e^{\mu T}\int_0^T e^{-\mu\tau} f\big(\bar{y}(\tau)\big)\big(\bar{y}'(\tau) - h(\bar{y}(\tau))\big)\,\mathrm{d}\tau, \end{aligned} \tag{3.18}$$

where $F : \mathbb{R} \to \mathbb{R}$ is an antiderivative of f. For notational convenience, we will abbreviate $J(\bar{y}) = J_{T,Y_{0-}}(\bar{y})$ and usually skip the argument τ of $\bar{y}$, $\bar{y}'$. We will maximize $J(\bar{y})$ not for all $\bar{y} \in C^1$ but only for those which correspond to a strategy $\Theta_t = \bar{\theta}(T-t)$, $t \in [0, T)$, that can be reached from (Y_{0-}, Θ_{0-}) by a suitable jump, $(Y_0, \Theta_0) = (Y_{0-} + \Delta, \Theta_{0-} + \Delta)$. That is, we need to consider those $\bar{y}$ for which $\bar{\theta}(T) - \bar{y}(T) = \Theta_{0-} - Y_{0-}$. Using the assumed connection $\bar{\theta}(0) = g(\bar{y}(0))$ of $\bar{y}$ and $\bar{\theta}$ at $\tau = 0$ through a function g, this condition reads

$$\Theta_{0-} - Y_{0-} = K_T(\bar{y}) := g\big(\bar{y}(0)\big) + \int_0^T \big(\bar{y}' - h(\bar{y})\big)\,\mathrm{d}\tau - \bar{y}(T)\,. \tag{3.19}$$

We will abbreviate $K = K_T$. In the language of calculus of variations, we now maximize $J(\bar{y})$ subject to the *isoperimetric condition* $K(\bar{y}) \equiv \Theta_{0-} - Y_{0-}$. This is equivalent to maximizing $(J + m_T K)(\bar{y})$ for an unknown constant (in $\bar{y}$) *Lagrange multiplier* $m_T \in \mathbb{R}$, cf. [GF00, Sect. 2.12.1]. We have

$$\begin{aligned} (J + m_T K)(\bar{y}) = F(Y_{0-}) - F\big(\bar{y}(T)\big) - \bar{y}(T)m_T \\ + e^{\mu T}\big(F(y) - F\big(y - g(y)\big) + e^{-\mu T} m_T g(y)\big)\big|_{y=\bar{y}(0)} \\ + e^{\mu T}\int_0^T G\big(\tau, \bar{y}(\tau), \bar{y}'(\tau)\big)\,\mathrm{d}\tau\,, \end{aligned} \tag{3.20}$$

where $G(\tau, y, p) := \big(e^{-\mu\tau} f(y) + e^{-\mu T} m_T\big)\big(p - h(y)\big)$. A necessary condition for $\bar{y} \in C^1$ maximizing $J + m_T K$ is that the first variation vanishes, cf. [GF00, Sect. 1.4], i.e. $\boldsymbol{\delta}(J + m_T K)(\bar{y})[\zeta] = 0$ for every C^1 perturbation ζ of $\bar{y}$. Note that $\bar{y}(T)$ fixed is on the (to be found) boundary surface, so $\zeta(T) = 0$. The first variation of $J + m_T K$ at $\bar{y}$ in direction ζ with $\zeta(T) = 0$ is

$$\begin{aligned} &\boldsymbol{\delta}(J + m_T K)(\bar{y})[\zeta] \\ &\quad = e^{\mu T}\Big(f\big(y - g(y)\big) + e^{-\mu T} m_T\Big)\big(g'(y) - 1\big)\Big|_{y=\bar{y}(0)}\zeta(0) \\ &\qquad + e^{\mu T}\int_0^T \Big(G_y(\tau, \bar{y}(\tau), \bar{y}'(\tau)) - \tfrac{\mathrm{d}}{\mathrm{d}\tau} G_p(\tau, \bar{y}(\tau), \bar{y}'(\tau))\Big)\zeta(\tau)\,\mathrm{d}\tau. \end{aligned} \tag{3.21}$$

By first considering perturbations with $\zeta(0) = 0$ we find that $\bar{y}(\tau)$ must satisfy the Euler equation $G_y - \frac{\mathrm{d}}{\mathrm{d}\tau} G_p = 0$, i.e.

$$-e^{-\mu\tau} f(y)(h\lambda + h' - \mu)(y) - e^{-\mu T} m_T h'(y) = 0 \quad \text{for all } \tau, \text{ at } y = \bar{y}(\tau), \tag{3.22}$$

or equivalently

$$e^{-\mu T} m_T = -e^{-\mu\tau} f(y)\left(\frac{h\lambda + h' - \mu}{h'}\right)(y) \quad \text{for all } \tau, \text{ at } y = \bar{y}(\tau). \tag{3.23}$$

Hence the integral term in (3.21) vanishes and we obtain the boundary condition

$$e^{\mu T}\Big(f\big(y-g(y)\big)+e^{-\mu T}m_T\Big)\big(g'(y)-1\big)\Big|_{y=\bar{y}(0)}=0\,. \tag{3.24}$$

Now assume $g'\neq 1$. Then (3.23) and (3.24) yield at $\tau=0$, $y=\bar{y}(0)$ that

$$f\big(y-g(y)\big)h'(y)=f(y)(h\lambda+h'-\mu)(y)\,. \tag{3.25}$$

Since the left-hand side is positive and the right-hand side is only positive for $y>y_\infty$, we find that g is to be defined on (y_∞,∞) only and we would necessarily have for the terminal impact $Y_{T-}>y_\infty$. Since $f:\mathbb{R}\to(0,\infty)$ is invertible by assumption, solving (3.25) for $g(y)$, we find the representation (3.9). Let us now summarize some properties of g.

Lemma 3.3.1. *The C^1 function $g:(y_\infty)\to\mathbb{R}$ from (3.9) satisfies $g(y)\to\infty$ as $y\searrow y_\infty$, $g>0$ on (y_∞,y_0), $g(y_0)=0$, $g<0$ on (y_0,∞) and $g'(y)<1$ for all $y\in(y_\infty,\infty)$.*

Note that for constant λ, we even have $g(y)=-\frac{1}{\lambda}\log\big((h\lambda+h'-\mu)(y)/h'(y)\big)$ and therefore $g'<0$ everywhere.

Proof. Since $f'>0$, we have $f^{-1}\in C^1$ and thus also $g\in C^1$. The sign change at y_0 follows from $(h\lambda-\mu)(y_0)=0$ and monotonicity and positivity of $(h\lambda+h'-\mu)/h'$ on (y_∞,∞). The limit as $y\searrow y_\infty$ follows similarly. We have $g'<1$ by direct calculation since $f'>0$ and $(f\,(h\lambda+h'-\mu)/h')'>0$ on (y_∞,∞). □

Differentiating both sides of the Euler equation (3.22) w.r.t. τ, we find

$$0=\mu e^{-\mu\tau}f(\bar{y})(h\lambda+h'-\mu)(\bar{y})-e^{-\mu\tau}\big(f\cdot(h\lambda+h'-\mu)\big)'(\bar{y})\bar{y}'-e^{-\mu T}m_T h''(\bar{y})\bar{y}',$$

which, together with (3.23), gives and ODE for $\bar{y}$.

$$\begin{aligned}\bar{y}'&=\left(\frac{\mu h'(h\lambda+h'-\mu)}{(h\lambda+h'-\mu)\big(\lambda h'+(h\lambda)'\big)-(h\lambda-\mu)(h\lambda+h')'}\right)(\bar{y})\\&=\mu\left(\frac{f\,(h\lambda+h'-\mu)/h'}{(f\,(h\lambda+h'-\mu)/h')'}\right)(\bar{y})=D(\bar{y})\,,\end{aligned} \tag{3.26}$$

where we recognize D from (3.10). Note that this is the same derivative as in the infinite time horizon problem as can be seen from differentiating the right-hand side of (2.11). With the initial (terminal time) condition $g(\bar{y}(0))=\bar{\theta}(0)$ and $\bar{\theta}'(\tau)=\bar{y}'(\tau)-h(\bar{y}(\tau))$, the candidate boundary surface $\mathcal{I}=\{(\tau,\bar{y}(\tau),\bar{\theta}(\tau))\mid\bar{y}(0)\in(y_\infty,\infty),\tau\in[0,T]\}$ is now uniquely determined. Let us summarize some properties of $\bar{y}$ and D.

Lemma 3.3.2. *The function $D:[y_\infty,\infty)\to\mathbb{R}$ from (3.10) is locally Lipschitz continuous, satisfies $D(y_\infty)=0$ and $\operatorname{sgn}(D(y))=\operatorname{sgn}(\mu)$ for all $y>y_\infty$. Moreover, for each initial value $\bar{y}(0)\in(y_\infty,\infty)$, the ODE (3.26), $\bar{y}'=D(\bar{y})$, has a unique global solution $\bar{y}:[0,\infty)\to(y_\infty,\infty)$.*

In particular, the boundary surface $\mathcal{I}$ is contained in $[0,T]\times(y_\infty,\infty)\times\mathbb{R}$.

Proof. The sign property of D and local Lipschitz-continuity on $[y_\infty, \infty)$ follow from the properties of $q = (h\lambda + h' - \mu)/h'$ in Lemma 3.1.3 by observing $D(y) = \mu f(y)q(y)/(fq)'(y)$.

The constant function $\tau \mapsto y_\infty$ is another solution of the autonomous ODE $y' = D(y)$. Local Lipschitz continuity gives local existence and uniqueness for $y' = D(y)$ with fixed $y(0) = [y_\infty, \infty)$. So that trajectories cannot cross and thus $\bar{y}(\cdot) > y_\infty$ whenever $\bar{y}(0) > y_\infty$.

To show global existence for initial value $\bar{y}(0) = z \in (y_\infty, \infty)$, consider the inverse, $\tau(y) := \int_z^y D(x)^{-1}\,\mathrm{d}x$. It suffices to prove $\tau(y) \to \infty$ for $y \to \infty$. We have $\tau(y) = \frac{1}{\mu}\big(\log f(y) - \log f(z) + \log q(y) - \log q(z)\big)$. Hence $f(y) \to \infty$ implies $\tau(y) \to \infty$ for $y \to \infty$. □

3.3.1 Reparametrizing the state space

Since $\bar{y}$ from Lemma 3.3.2 is defined via the terminal impact $Y_{T-} = \bar{y}(0) \in (y_\infty, \infty)$, it proves useful to reparametrize the whole state space in terms of terminal impact $z \in (y_\infty, \infty)$, time to liquidation $\tau \in [0, T]$, and initial block size $d \in \mathbb{R}$, instead of τ and initial impact and position. We now prove that such reparametrization covers the whole state space, cf. Corollary 3.3.5. In subsequent Section 3.3.2, we utilize this reparametrization to formulate in (3.29) our candidate V for the value function v of problem (3.7) and the corresponding free boundary surface $\mathcal{I}$ in Theorem 3.3.8.

To stress the dependence of $\bar{y}$ and $\bar{\theta}$ on z, let us write

$$\begin{cases} \bar{y}_\tau(\tau; z) = D\big(\bar{y}(\tau; z)\big), & \text{for } \tau \in [0, T], \\ \bar{\theta}_\tau(\tau; z) = \bar{y}_\tau(\tau; z) - h\big(\bar{y}(\tau; z)\big), & \text{for } \tau \in [0, T], \\ \bar{y}(0; z) = z, \\ \bar{\theta}(0; z) = g(z), \end{cases} \tag{3.27}$$

where g is given by (3.9) and D is given by (3.10). Existence and uniqueness of $\bar{y}(\cdot; z), \bar{\theta}(\cdot; z) \in C^1([0, T])$ follow from Lemmas 3.3.1 and 3.3.2. Our candidate boundary surface will be $\mathcal{I} = \big\{\big(\tau, \bar{y}(\tau; z), \bar{\theta}(\tau; z)\big) \mid \tau \in [0, T], z \in (y_\infty, \infty)\big\}$, see also Theorem 3.3.8.

Intuitively, it is clear that given a point (τ, y, θ) on the boundary, we can follow the path $(\bar{y}, \bar{\theta})$ through that point backwards to reach $(0, z, g(z))$. To this end, we need to invert $z \mapsto \bar{y}(\tau; z)$. By [Wal98, Theorem III.13.X], the map $z \mapsto \bar{y}(\tau; z)$ is continuously differentiable. By the following lemma, it is moreover increasing and thus injective.

Lemma 3.3.3. *We have $\bar{y}_z(\tau; z) > 0$ for all $\tau \in [0, T]$ and $z \in (y_\infty, \infty)$.*

Proof. Fix an arbitrary z. The pair $(\bar{y}, \bar{y}_z) := (\bar{y}(\cdot; z), \bar{y}_z(\cdot; z))$ solves the *autonomous* ODE $\partial_\tau(\bar{y}, \bar{y}_z) = (D(\bar{y}), D'(\bar{y})\bar{y}_z)$ with initial value $(z, 1)$. But the pair $(\bar{y}(\cdot; z), 0)$ also solves this ODE and since two solution trajectories of autonomous ODEs cannot cross, i.e. $\{(\bar{y}(\tilde{\tau}; z), 0) \mid \tilde{\tau}\} \cap \{(\bar{y}(\tau; z), \bar{y}_z(\tau; z)) \mid \tau\} = \varnothing$, we must have $\bar{y}_z(\tau; z) \neq 0$ for all τ. By continuity of $\bar{y}_z(\cdot; z)$ and the initial value 1 it follows that $\bar{y}_z(\tau; z) > 0$ for all τ. □

By Lemma 3.3.2, $\bar{y}(\tau, \cdot)$ has the range (y_∞, ∞), because $\tau \mapsto \bar{y}(\tau; z)$ is continuous and thus bounded on any compact interval $[0, \tau]$ so that at τ we can reach any point in (y_∞, ∞) by starting from an appropriate $z = \bar{y}(0, z) \in (y_\infty, \infty)$. We will write $z(\tau; y) := (\bar{y}(\tau; \cdot))^{-1}(y)$, $y > y_\infty$, for the inverse. It solves the differential equation

$$z_\tau(\tau; y) = \frac{-\bar{y}_\tau\big(\tau; z(\tau; y)\big)}{\bar{y}_z\big(\tau; z(\tau; y)\big)}, \qquad \text{with} \qquad z(0; y) = y\,.$$

Let $g_\tau(y) = \bar\theta(\tau; z(\tau; y))$, so that $\{(y, g_\tau(y)) \mid y > y_\infty\}$ is the τ-slice of the (candidate) boundary surface. In particular, we have $g_0(y) = g(y)$ and

$$g'_\tau(y) = \frac{\bar\theta_z\big(\tau; z(\tau; y)\big)}{\bar y_z\big(\tau; z(\tau; y)\big)}. \tag{3.28}$$

To show that the whole representation $P(\tau, z, d) := (\tau, \bar y(\tau; z) + d, \bar\theta(\tau; z) + d)$ is injective, we investigate its Jacobian.

Lemma 3.3.4. *The z derivatives of $\bar y$ and $\bar\theta$ have the following representations*

$$\begin{aligned}
\bar y_z(\tau; z) &= 1 + \int_0^\tau D'(\bar y(s; z))\bar y_z(s; z)\,\mathrm{d}s\,,\\
\bar\theta_z(\tau; z) &= g'(z) + \int_0^\tau (D' - h')(\bar y(s; z))\bar y_z(s; z)\,\mathrm{d}s\,.
\end{aligned}$$

Moreover, we have $(\bar y_z - \bar\theta_z)(\tau; z) > 0$ for all $\tau \in [0, T]$ and $z \in (y_\infty, \infty)$.

Proof. The above representation of $\bar y_z$ and $\bar\theta_z$ follows from [Wal98, Theorem III.13.X eq. (14)]. Now fix an arbitrary z. By Lemma 3.3.3, the difference $\bar y_z - \bar\theta_z$ is increasing in τ, because $\partial_\tau(\bar y_z - \bar\theta_z) = h'(\bar y)\bar y_z$. Moreover, since $(\bar y_z - \bar\theta_z)(0; z) = 1 - g'(z) > 0$ by Lemma 3.3.1, we get $\bar y_z - \bar\theta_z > 0$ everywhere. □

The desired bijective reparametrization P is provided by the next result.

Corollary 3.3.5. *The state space representation $P : [0, T] \times (y_\infty, \infty) \times \mathbb{R} \to [0, T] \times \mathbb{R}^2$ defined by $P(\tau, z, d) := (\tau, \bar y(\tau; z) + d, \bar\theta(\tau; z) + d)$ is bijective.*

Proof. For injectivity, note that $z \mapsto \bar y(\tau; z) - \bar\theta(\tau; z)$ is injective by Lemma 3.3.4 and since $y - \theta = (y - d) - (\theta - d)$, we can reconstruct z from τ and $y - \theta$ for $(\tau, y, \theta) \in P([0, T] \times \mathbb{R}^2)$. Now with $d = y - \bar y(\tau; z)$, we have uniquely identified (τ, z, d) satisfying $(\tau, y, \theta) = P(\tau, z, d)$.

For surjectivity, note that the τ-slice of the (candidate) boundary surface has an asymptote, $g_\tau(y) \to +\infty$ for $y \searrow y_\infty$, so it suffices to show that e.g. $g_\tau(y) \le 0$ for all y large enough. By Lemma 3.3.1, we have $g_0(y) \le 0$ for all $y \ge y_0$.

First consider the case $\mu \le 0$, so that $D(y) \le 0$ by Lemma 3.3.2. Direct calculations with (3.28) and Lemma 3.3.4 yield that $\partial_\tau g_\tau(y) = D(y)(1 - g'_\tau(y)) - h(y) \le -h(y) \le 0$ for $y \ge 0$. Hence, $g_\tau(y) \le g_0(y)$ whenever $\mu \le 0$, for all τ and $y > 0$.

Now for $\mu > 0$ we have $y_0 > 0$ and thus $(h\lambda-\mu)(h\lambda+h'-\mu)/h' + \big((h\lambda+h'-\mu)/h'\big)'h > 0$ on (y_0, ∞) by Lemma 3.1.3. This implies $\bar\theta_\tau(\tau; z) = \big(D - h\big)(\bar y(\tau; z)) < 0$ whenever $\bar y(\tau; z) > y_0$. Since moreover $\bar y_\tau = D(\bar y) > 0$ for $\mu > 0$, we find in particular that $\bar\theta_\tau(\tau; z) < 0$ for all $z > y_0$. Since already $g_0(y) < 0$ for y large enough, we hence must also have $g_\tau(y) < 0$ for all $y > \bar y(\tau; y_0)$ large enough. □

3.3.2 Local optimality for smooth strategies

We are now ready to formulate our candidate for the value function and optimal strategy Θ and to prove a local optimality result for the boundary $\mathcal{I}$ that characterizes Θ in Theorem 3.3.8 using the second variation of the functional $J + m_T K$.

For $\mathcal{I} := P([0,T] \times (y_\infty, \infty) \times \{0\})$ let Θ be the strategy which performs a possible initial jump to reach $\mathcal{I}$ from $Y_{0-} = y$, $\Theta_{0-} = \theta$, then trades in rates $-\bar{\theta}'$ such as to stay on $\mathcal{I}$, and finishes with a terminal jump to reach $\Theta_T = 0$. Then $V(T, y, \theta) = \mathbb{E}[L_T(\Theta)]$ is given by

$$V(\tau, y, \theta) := J_{\tau,y}\big(\bar{y}(\cdot; z(\tau, y, \theta))\big), \tag{3.29}$$

where $z(\tau, y, \theta)$ denotes the second component of $P^{-1}(\tau, y, \theta)$. The function V is our candidate for the value function v of the original problem (3.7). We can rewrite V in terms of the new coordinates, $V(P(\tau, z, 0)) = \bar{V}(\tau; z)$ with $\bar{V}(\tau; z) = V\big(\tau, \bar{y}(\tau; z), \bar{\theta}(\tau; z)\big)$ given by

$$\bar{V}(\tau; z) = e^{\mu\tau}\big(F(z) - F(z - g(z))\big) - e^{\mu\tau}\int_0^\tau e^{-\mu s} f\big(\bar{y}(s; z)\big)\big(h\big(\bar{y}(s; z)\big) - \bar{y}_\tau(s; z)\big)\,\mathrm{d}s. \tag{3.30}$$

By construction, V already satisfies (3.14).

Lemma 3.3.6. *The function V from* (3.29) *satisfies $V(0, \cdot, 0) = 0$ and the variational equality* (3.14)*, $V_y(\tau, y, \theta) + V_\theta(\tau, y, \theta) = f(y)$, everywhere.*

Proof. Note that $J = J_{\tau,y}$ and $K = K_\tau$ do not depend on θ directly, but $\bar{y}$ does. Therefore, denote this function by $\bar{y}_{y,\theta}$. We immediately see $J_{0,y}(\bar{y}_{y,0}) = 0$. Now, consider the diagonal $(y + d, \theta + d)$ for arbitrary displacement $d \in \mathbb{R}$. Since $K_\tau(\bar{y}_{y+d,\theta+d}) = K_\tau(\bar{y}_{y,\theta})$ and $J_{\tau,y+d}(\cdot) = J_{\tau,y}(\cdot) + F(y + d) - F(y)$, we find $\bar{y}_{y+d,\theta+d} = \bar{y}_{y,\theta}$, and thus

$$V(\tau, y + d, \theta + d) = V(\tau, y, \theta) + F(y + d) - F(y)\,.$$

The variational equality (3.14), $V_y + V_\theta = f$, follows. □

By utilizing the second variation $\boldsymbol{\delta}^2(J + m_T K)(\bar{y})[\zeta]$ we get optimality of $\bar{y}$ in Theorem 3.3.8 below. The second variation is the second order term in the Taylor expansion of $(J + m_T K)(\bar{y} + \zeta)$ around $\bar{y}$, cf. [GF00, Ch. 5]. The general Taylor expansion reads as follows.

Proposition 3.3.7 (Taylor)**.** *Let $\varphi \in C^{n+1}(\mathbb{R}^k; \mathbb{R})$, $x \in \mathbb{R}^k$ and $h \in \mathbb{R}^k$. Then we have*

$$\varphi(x + h) = \sum_{k=0}^{n} \sum_{|\boldsymbol{\alpha}|=k} \frac{h^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}\partial^{\boldsymbol{\alpha}}\varphi(x) + \sum_{|\boldsymbol{\alpha}|=n+1} \frac{(n+1)}{\boldsymbol{\alpha}!} h^{\boldsymbol{\alpha}} \int_0^1 (1 - t)^n \partial^{\boldsymbol{\alpha}}\varphi(x + th)\,\mathrm{d}t,$$

with multi-index $\boldsymbol{\alpha} \in \mathbb{N}_0^k$.

Theorem 3.3.8. *The candidate boundary*

$$\mathcal{I} = P([0,T] \times (y_\infty, \infty) \times \{0\}) = \big\{(\tau, \bar{y}(\tau; z), \bar{\theta}(\tau; z) \;\big|\; \tau \in [0,T], z \in (y_\infty, \infty)\big\} \tag{3.31}$$

with $\bar{y}, \bar{\theta}$ given by (3.27) *is locally optimal in the following sense:*

Let $\bar{y} := \bar{y}(\cdot; z)$ for $(T, z, d) := P(T, Y_{0-}, \Theta_{0-})$. Then there exists $\varepsilon > 0$ such that for all $\hat{y} \in C^1([0,T])$ which satisfy $\|\bar{y} - \hat{y}\|_{W^{1,\infty}} := \|\bar{y} - \hat{y}\|_\infty \vee \|\bar{y}' - \hat{y}'\|_\infty \in (0, \varepsilon)$ and $\hat{y}(T) = \bar{y}(T)$ we have $(J + m_T K)(\bar{y}) > (J + m_T K)(\hat{y})$.

The condition $\hat{y}(T) = \bar{y}(T)$ is necessary for $\hat{y}$ to correspond to a strategy which starts in (Y_{0-}, Θ_{0-}).

Proof. The Taylor expansion of $(J+m_TK)(\bar{y}+\zeta)$ with $\zeta(T)=0$ gives

$$(J+m_TK)(\bar{y}+\zeta) = (J+m_TK)(\bar{y}) + \boldsymbol{\delta}(J+m_TK)(\bar{y})[\zeta] + \boldsymbol{\delta}^2(J+m_TK)(\bar{y})[\zeta] + \boldsymbol{\varepsilon}(\zeta)$$

where the first variation $\boldsymbol{\delta}(J+m_TK)(\bar{y})[\zeta]$ given in (3.21) equals zero at $\bar{y}$ by construction and the second variation $\boldsymbol{\delta}^2(J+m_TK)(\bar{y})[\zeta]$ and error term $\boldsymbol{\varepsilon}(\zeta)$ are

$$\begin{aligned}
&\boldsymbol{\delta}^2(J+m_TK)(\bar{y})[\zeta] \\
&\quad= \frac{e^{\mu T}}{2}\Big(f\big(y-g(y)\big)g''(y) + e^{-\mu T}m_T g''(y) - f'\big(y-g(y)\big)\big(1-g'(y)\big)^2\Big)\Big|_{y=\bar{y}(0)}\zeta(0)^2 \\
&\qquad + \frac{e^{\mu T}}{2}\int_0^T \Big(G_{yy}(\tau,\bar{y}(\tau),\bar{y}'(\tau)) - \tfrac{\mathrm{d}}{\mathrm{d}\tau}G_{yp}(\tau,\bar{y}(\tau),\bar{y}'(\tau))\Big)\zeta(\tau)^2\,\mathrm{d}\tau\,, \qquad (3.32)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\varepsilon}(\zeta) = \frac{1}{2}e^{\mu T}\bigg(&\int_0^1 (1-\eta)^2 f''(\bar{y}+\eta\zeta)\,\mathrm{d}\eta \\
&- \int_0^1 (1-\eta)^2\big(F\circ(\mathrm{id}-g)\big)'''(\bar{y}+\eta\zeta)\,\mathrm{d}\eta \\
&- e^{-\mu T}m_T\int_0^1 (1-\eta)^2 g'''(\bar{y}+\eta\zeta)\,\mathrm{d}\eta\bigg)\bigg|_{\tau=0}\zeta(0)^3 \\
+ e^{\mu T}\int_0^T\!\!\int_0^1 &(1-\eta)^2\Big(\tfrac{1}{2}\zeta^3 G_{yyy} + \tfrac{3}{2}\zeta^2\zeta' G_{yyp}\Big)(\tau,\bar{y}+\eta\zeta,\bar{y}'+\eta\zeta')\,\mathrm{d}\eta\,\mathrm{d}\tau.
\end{aligned}$$

Since f, g, F, G are all C^3, continuity of $\bar{y}, \bar{y}'$ and compactness of $[0,T]$ give that we can bound the error term by $|\boldsymbol{\varepsilon}(\zeta)| \le C\,\|\zeta\|^3_{W^{1,\infty}}$ for some constant C. For perturbations $\zeta \in C^1([0,T])$, which additionally satisfy $\zeta(0)=0$, we get

$$\boldsymbol{\delta}^2(J+m_TK)(\bar{y})[\zeta] = \frac{1}{2}e^{\mu T}\int_0^T \big(G_{yy} - \tfrac{\mathrm{d}}{\mathrm{d}\tau}G_{yp}\big)\zeta^2\,\mathrm{d}\tau\,,$$

with $G_{yy} - \frac{\mathrm{d}}{\mathrm{d}\tau}G_{yp} = -e^{-\mu\tau}f'\,(h\lambda + h' - \mu) - e^{-\mu\tau}f\,(h\lambda+h')' - e^{-\mu T}m_T h''$, evaluated at $\bar{y}(\tau)$. By (3.23) this simplifies to

$$G_{yy} - \tfrac{\mathrm{d}}{\mathrm{d}\tau}G_{yp} = -e^{-\mu\tau}h'(\bar{y}(\tau))\bigg(f\frac{h\lambda+h'-\mu}{h'}\bigg)'(\bar{y}(\tau)),$$

which is negative because $\bar{y} > y_\infty$ by Lemma 3.3.2 and h and $f\,(h\lambda+h'-\mu)/h'$ are increasing on (y_∞,∞) by Assumption 3.1.1. Hence for all perturbations $\zeta \not\equiv 0$ with $\zeta(0)=\zeta(T)=0$ we have $\boldsymbol{\delta}^2(J+m_TK)(\bar{y})[\zeta] < 0$. Now considering $\zeta(0)\neq 0$, by (3.24) and $g' < 1$ we get

$$\boldsymbol{\delta}^2(J+m_TK)(\bar{y})[\zeta] < \frac{e^{\mu T}}{2}g''(y)\big(f\big(y-g(y)\big) + e^{-\mu T}m_T\big)\Big|_{y=\bar{y}(0)}\zeta(0)^2 = 0.$$

So for $\|\hat{y}-\bar{y}\|_{W^{1,\infty}}$ small enough, we get $(J+m_TK)(\hat{y}) < (J+m_TK)(\bar{y})$. □

Note that given a C^1 path $\bar{y}:[0,T]\to\mathbb{R}$ of $Y_t = \bar{y}(T-t)$, $t\in[0,T)$, the asset position $\Theta_t = \bar{\theta}(T-t)$ is uniquely determined by $\bar{\theta}(T) = \Theta_{0-} + \bar{y}(T) - Y_{0-}$ and $\bar{\theta}'(\tau) = \bar{y}'(\tau) - h(\bar{y}(\tau))$, subject to the condition that we can reach $(\bar{y}(T),\bar{\theta}(T))$ with an initial jump, i.e. that $K(\bar{y}) = \Theta_{0-} - Y_{0-}$. The variational equality (3.14) and (3.15) for the value of the candidate optimal strategy Θ given by $\bar{y}$ is straight forward.

Theorem 3.3.8 is key to verify the variational inequality $\mathcal{L}V < 0$, cf. Lemma 3.3.10.

3.3.3 Proving the variational inequality

We will now prove the variational inequality $\mathcal{L}V \leq 0$ for our candidate value function V from (3.29). First, we rewrite $\mathcal{L}V$ in terms of the new variables τ, terminal impact z, and initial jump size d. We prove equality $\mathcal{L}V = 0$ on the boundary surface $\mathcal{I}$ in Lemma 3.3.9. The strict inequality $\mathcal{L}V < 0$ in $[0,T] \times \mathbb{R}^2 \setminus \mathcal{I}$ then follows from Theorem 3.3.8, which guaranties $\mathcal{L}V < 0$ in a neighborhood of $\mathcal{I}$, cf. Lemma 3.3.10.

Let $d(\tau, y, \theta)$ be the $\|\cdot\|_\infty$-distance in direction $(-1,-1)$ of the point $(\tau, y, \theta) \in [0,T] \times \mathbb{R}^2$ to the boundary $\mathcal{I} = P([0,T] \times (y_\infty, \infty) \times \{0\})$, i.e. third component of $P^{-1}(\tau, y, \theta)$. For brevity, we will usually abbreviate $d = d(\tau, y, \theta)$, and its partial derivatives $d_\tau = d_\tau(\tau, y, \theta)$ etc. Since $(\tau, y - d, \theta - d)$ lies on the boundary, we have $\theta - d = \bar{\theta}(y - d)$. This yields all partial derivatives of d

$$\begin{aligned} d_y &= \Big(\frac{g'_\tau}{g'_\tau - 1}\Big)(y-d)\,, \\ d_\theta &= \Big(\frac{1}{1-g'_\tau}\Big)(y-d) = 1 - d_y\,, \\ d_\tau &= \Big(-D + \frac{h}{1-g'_\tau}\Big)(y-d) = -D(y-d) + h(y-d)d_\theta\,. \end{aligned} \tag{3.33}$$

Using (3.30), we can now calculate the partial derivatives of the candidate value function $V(\tau, y, \theta) = \bar{V}(\tau; z(\tau; y-d)) + F(y) - F(y-d)$. For $\bar{V}$ we get

$$\begin{aligned} \bar{V}_\tau(\tau; z) &= \mu \bar{V}(\tau; z) - \big(f\,(h-D)\big)(\bar{y}(\tau; z))\,, \\ \bar{V}_z(\tau; z) &= e^{\mu\tau}\big(f(z) - f(z - g_0(z))(1 - g'_0(z))\big) \\ &\qquad - e^{\mu\tau}\int_0^\tau e^{-\mu s}\big(f'\,(h-D) + f\,(h' - D')\big)(\bar{y}(s;z))\,\bar{y}_z(s;z)\,\mathrm{d}s\,. \end{aligned}$$

For brevity, we will omit the arguments of $\bar{V}(\tau; z(\tau; y-d))$ and af the partial derivatives of $z(\tau; y-d)$, as we do it with d. We have

$$\begin{aligned} V_\tau(\tau, y, \theta) &= \bar{V}_\tau + \bar{V}_z\,\big(z_\tau + z_y\,(-d_\tau)\big) - f(y-d)(-d_\tau)\,, \\ V_y(\tau, y, \theta) &= \bar{V}_z\, z_y\,(1 - d_y) + f(y) - f(y-d)(1-d_y)\,, \\ V_\theta(\tau, y, \theta) &= \bar{V}_z\, z_y\,(-d_\theta) + f(y-d)d_\theta\,. \end{aligned}$$

Now checking the variational equality (3.12) is straight-forward.

Lemma 3.3.9. *The candidate buy-sell value function V from Lemma 3.3.6 satisfies the variational equality* (3.12), *$\mathcal{L}V = 0$, on the boundary $\mathcal{I}$.*

Proof. By direct calculation, using $d = 0$ in $\mathcal{I}$,

$$\begin{aligned} \mathcal{L}V &\equiv -\mu V + V_\tau + hV_y \\ &= -\mu\bar{V} + \bar{V}_\tau + \bar{V}_z z_\tau - \bar{V}_z z_y d_\tau + f d_\tau + h\bar{V}_z z_y d_\theta + hf - hf d_\theta \\ &= -f\cdot(h-D) + \bar{V}_z z_\tau - \bar{V}_z z_y\,(-D + hd_\theta) \\ &\qquad + f\cdot(-D + hd_\theta) + h\bar{V}_z z_y d_\theta + hf - hf d_\theta \\ &= \bar{V}_z z_\tau + \bar{V}_z z_y D = \bar{V}_z\,(-\bar{y}_\tau z_y - D z_y) = 0\,. \end{aligned}$$

□

Now, to investigate the variational inequality (3.13) for general (τ, y, θ), fix a point $(\tau, y_b, \theta_b) \in \mathcal{I}$ on the boundary and vary the distance d. Since $d(\tau, y_b + x, \theta_b + x) = x$, we find that $d_\theta(\tau, y_b + x, \theta_b + x) = d_\theta(\tau, y_b, \theta_b)$, $d_y(\tau, y_b + x, \theta_b + x) = d_y(\tau, y_b, \theta_b)$, and $d_\tau(\tau, y_b + x, \theta_b + x) = d_\tau(\tau, y_b, \theta_b)$, i.e. d_θ, d_y and d_τ are constant when (τ, y_b, θ_b) is fixed. Let

$$\begin{aligned} k(d) &:= \mathcal{L}V(\tau, y_b + d, \theta_b + d) \qquad (3.34) \\ &= -\mu V(\tau, y_b+d, \theta_b+d) + V_\tau(\tau, y_b+d, \theta_b+d) + h(y_b+d) V_y(\tau, y_b+d, \theta_b+d) \\ &= -\mu\big(\bar{V}(\tau, z(\tau; y_b)) + F(y_b + d) - F(y_b)\big) \\ &\quad + \bar{V}_\tau(\tau, z(\tau; y_b)) + \bar{V}_z(\tau, z(\tau; y_b))\big(z_\tau(\tau; y_b) - z_y(\tau; y_b) d_\tau\big) + f(y_b) d_\tau \\ &\quad + h(y_b + d)\big(\bar{V}_z(\tau, z(\tau; y_b)) z_y(\tau; y_b) d_\theta + f(y_b + d) - f(y_b) d_\theta\big)\,. \end{aligned}$$

By Lemma 3.3.9, we already know $k(0) = 0$. To prove $k(d) > 0$ for $d \neq 0$, it suffices to show $k'(d) > 0$ for $d > 0$ and $k'(d) < 0$ for $d < 0$. With now fixed $z_y = z_y(\tau; y_b)$ and $\bar{V}_z = \bar{V}_z(\tau; z_y(\tau; y_b))$, we have

$$\begin{aligned} k'(d) &= -\mu f(y_b + d) + h'(y_b + d)\big(\bar{V}_z z_y d_\theta + f(y_b + d) - f(y_b) d_\theta\big) \\ &\quad + h(y_b + d) f'(y_b + d) \\ &= \big(f\,(h\lambda + h' - \mu)\big)(y_b + d) + h'(y_b + d)\big(\bar{V}_z z_y - f(y_b)\big) d_\theta\,. \qquad (3.35) \end{aligned}$$

In particular, at $d = 0\pm$, meaning the one-sided derivatives $k'(0+)$ and $k'(0-)$, respectively, we get with $q := (h\lambda + h' - \mu)/h'$ that

$$\big(\bar{V}_z(\tau, z(\tau; y_b)) z_y(\tau; y_b) - f(y_b)\big) d_\theta = \frac{k'(0\pm)}{h'(y_b)} - f(y_b) q(y_b)\,.$$

Note that the left-hand side does not depend on the value of d. Using our local optimality result from Theorem 3.3.8, we will show $k'(0-) \le 0 \le k'(0+)$ in Lemma 3.3.10 below. This implies

$$-V_\theta(\tau, y_b, \theta_b) = \big(\bar{V}_z(\tau, z(\tau; y_b)) z_y(\tau; y_b) - f(y_b)\big) d_\theta = -f(y_b) q(y_b)\,, \qquad (3.36)$$

so that (3.35) simplifies to $k'(d)/h'(y_b+d) = \big(fq\big)(y_b+d) - \big(fq\big)(y_b)$. By Assumption 3.1.1 we have $fq < 0$ on $(-\infty, y_\infty)$ and fq is positive and increasing on (y_∞, ∞). Since $y_b > y_\infty$ this implies $k'(d) < 0$ for $d < 0$ and $k'(d) > 0$ for $d > 0$, so that

$$k(d) > 0 \quad \text{for all } d \neq 0. \qquad (3.37)$$

The above derivation depends on the correct sign of $k'(0\pm)$. The idea is that a wrong sign of $k'(0\pm)$, would cause a local violation of the variational inequality (3.13) which we could exploit to construct a strategy near our candidate $\bar{\theta}$ that would generate strictly larger proceeds, contradicting Theorem 3.3.8.

Lemma 3.3.10. *We have $k'(0+) \ge 0$ and $k'(0-) \le 0$ for the function k from* (3.34)*.*

Proof. We will prove $k'(0+) \ge 0$ by contradiction using Theorem 3.3.8. The proof of $k'(0-) \le 0$ is analogous. To make the dependence of k and k' on the boundary point $(\tau, y_b, \theta_b) = P(\tau, z_b, 0) \in \mathcal{I}$ explicit, we use subscript notation $k_{\tau,z_b}(d)$ and $k'_{\tau,z_b}(d)$, respectively. By continuity of k and k' in τ, we can assume $\tau < T$.

Assume $k'_{\tau_b,z_b}(0+) < 0$ for some τ_b, z_b. By continuity of $(\tau,z,d) \mapsto k'_{\tau,z}(d)$ there exists a neighborhood $U = (\tau_1,\tau_2) \times (z_b - \varepsilon, z_b + \varepsilon) \times (0,\varepsilon)$ of "$(\tau_b, z_b, 0+)$" such that $k'_{\tau,z}(d) < 0$ for all $(\tau,z,d) \in U$, or "$k' < 0$ in U" for short. Since moreover $k_{\tau,z}(0) = 0$ by Lemma 3.3.9, we also have $k < 0$ in U.

We now need to construct a path $(\tau, \hat{y}(\tau), \hat{\theta}(\tau)) = P(\tau, \hat{z}(\tau), \hat{d}(\tau))$ that remains on the boundary surface $\mathcal{I}$ for $\tau \notin (\tau_1,\tau_2)$ and runs inside U for $\tau \in (\tau_1,\tau_2)$, and such that moreover $\hat{\theta}' = \hat{y}' - h(\hat{y})$ everywhere. Take a smooth function $\varphi : [0,T] \to [0,1]$ with support $[\tau_1,\tau_2]$ and let $\hat{d}(\tau) := \hat{\varepsilon}\varphi(\tau)$ for some (to be determined) constant $\hat{\varepsilon} \in (0,\varepsilon)$. It remains to find an appropriate $\hat{z}$. By fixing $\hat{z}(\tau)$ constant with value $z_1 \in (z_b - \varepsilon, z_b + \varepsilon)$ on $[0,\tau_1]$ and constant with value $z_2 \in (z_b - \varepsilon, z_b + \varepsilon)$ on $[\tau_1, \infty)$, we have $\hat{\theta}' = \hat{y}' - h(\hat{y})$ outside of $[\tau_1,\tau_2]$. We know that

$$\begin{aligned}
\hat{y}'(\tau) &= \hat{\varepsilon}\varphi'(\tau) + \bar{y}_\tau(\tau;\hat{z}(\tau)) + \bar{y}_z(\tau;\hat{z}(\tau))\hat{z}'(\tau),\\
\hat{\theta}'(\tau) &= \hat{\varepsilon}\varphi'(\tau) + \bar{\theta}_\tau(\tau;\hat{z}(\tau)) + \bar{\theta}_z(\tau;\hat{z}(\tau))\hat{z}'(\tau).
\end{aligned}$$

Since $\bar{\theta}_\tau = \bar{y}_\tau - h(\bar{y})$, the condition $\hat{\theta}' = \hat{y}' - h(\hat{y})$ reduces to the non-autonomous ODE

$$\hat{z}' = \frac{h\big(\bar{y}(\tau;\hat{z}) + \hat{\varepsilon}\varphi(\tau)\big) - h\big(\bar{y}(\tau;\hat{z})\big)}{\big(\bar{y}_z - \bar{\theta}_z\big)(\tau;\hat{z})} =: R(\tau,\hat{z},\hat{\varepsilon})\,, \tag{3.38}$$

with abbreviation $\hat{z} = \hat{z}(\tau)$. By monotonicity of h and Lemma 3.3.4, we have $R(\tau,\hat{z},\hat{\varepsilon}) > 0$ for $\tau_1 < \tau < \tau_2$, since $\hat{\varepsilon}\varphi > 0$ there, and $R(\tau,\hat{z},\hat{\varepsilon}) = 0$ for $\tau \notin (\tau_1,\tau_2)$. Since $R(\tau,\cdot,\hat{\varepsilon})$ is locally Lipschitz, there exists a unique local solution $\hat{z} = \hat{z}_{\hat{\varepsilon}}$ for (3.38) with $\hat{z}(\tau_b) = z_b$ on a maximal interval $I_{\hat{\varepsilon}}$ with $\tau_b \in I_{\hat{\varepsilon}}$ such that $\hat{z}_{\hat{\varepsilon}}(\cdot) \in (z_b - \varepsilon, z_b + \varepsilon)$ everywhere. In the limiting case $\hat{\varepsilon} \to 0$ we have the constant solution $\hat{z}_0(\cdot) \equiv z_b$ globally. Hence there exists $\hat{\varepsilon} > 0$ small enough, such that $I_{\hat{\varepsilon}} \supset [\tau_1,\tau_2]$. Now, we immediately see that $I_{\hat{\varepsilon}} \supset [0,T]$ for $T := \tau_2$ and that $\hat{z} = \hat{z}_{\hat{\varepsilon}}$ is constant on $[0,\tau_1]$. Moreover, since $\hat{y}(\cdot) - \bar{y}(\cdot;\hat{z}_{\hat{\varepsilon}}(T)) \to 0$ in $\|\cdot\|_{W^{1,\infty}}$ on $[0,T]$ as $\hat{\varepsilon} \to 0$, we can assume that $\hat{\varepsilon} > 0$ is small enough such that Theorem 3.3.8 applies. Note that $\bar{y}(T;\hat{z}_{\hat{\varepsilon}}(T)) = \hat{y}(T)$ and $\bar{\theta}(T;\hat{z}_{\hat{\varepsilon}}(T)) = \hat{\theta}(T)$, so

$$K\big(\bar{y}(\cdot;\hat{z}_{\hat{\varepsilon}}(T))\big) = \bar{\theta}(T;\hat{z}_{\hat{\varepsilon}}(T)) - \bar{y}(T;\hat{z}_{\hat{\varepsilon}}(T)) = \hat{\theta}(T) - \hat{y}(T) = K(\hat{y})\,.$$

Hence the corresponding strategy $\hat{\Theta}$ with $\hat{\Theta}_t = \hat{\theta}(T-t)$ for $t \in [0,T)$ and $\hat{\Theta}_T = 0$ which starts at impact $\hat{Y}_{0-} = Y_{0-}$ with $\hat{\Theta}_{0-} = \Theta_{0-}$ generates strictly less expected proceeds $\mathbb{E}[L_T(\hat{\Theta})] = J(\hat{y})$ than the strategy Θ with $\Theta_t = \bar{\theta}(T-t)$ for $t \in [0,T)$ and $\Theta_T = 0$ which starts from the same point. On the other hand, we have $L_T = G_T$ (after the terminal block trade) since $V(0,\cdot,0) = 0$, so (3.11) gives

$$\begin{aligned}
L_T(\hat{\Theta}) - L_T(\Theta) &= G_{0-}(\hat{\Theta}) - G_{0-}(\Theta) + \int_0^T \mathrm{d}\big(G_.(\hat{\Theta}) - G_.(\Theta)\big)_t\\
&= \overline{S}_0 V(T,Y_{0-},\Theta_{0-}) - \overline{S}_0 V(T,Y_{0-},\Theta_{0-}) + \int_0^T \mathrm{d}\big(G_.(\hat{\Theta}) - G_.(\Theta)\big)_t\\
&= \int_0^T e^{-\gamma t} V(T-t,\hat{Y}_t,\hat{\Theta}_t)\,\mathrm{d}M_t - \int_0^T e^{-\gamma t} V(T-t,Y_t,\Theta_t)\,\mathrm{d}M_t\\
&\qquad - \int_0^T e^{-\gamma t}\overline{S}_t\,\mathcal{L}V(T-t,\hat{Y}_t,\hat{\Theta}_t)\,\mathrm{d}t + \int_0^T e^{-\gamma t}\overline{S}_t\,\mathcal{L}V(T-t,Y_t,\Theta_t)\,\mathrm{d}t,
\end{aligned}$$

using Lemma 3.3.6. Since V is continuous and $\hat{Y}, \hat{\Theta}, Y, \Theta$ are bounded, the local martingales are true martingales, so in expectation we get

$$\begin{aligned}\mathbb{E}[L_T(\hat{\Theta}) - L_T(\Theta)] &= \int_0^T \mathbb{E}\big[e^{-\gamma t}\overline{S}_t\big(\mathcal{L}V(T-t,\hat{Y}_t,\hat{\Theta}_t) - \mathcal{L}V(T-t,Y_t,\Theta_t)\big)\big]\,\mathrm{d}t \\ &= -\int_0^T e^{\mu(T-\tau)}\overline{S}_0\big(k_{\tau,\hat{z}_{\hat{\varepsilon}}(\tau)}(\hat{d}(\tau)) - k_{\tau,\hat{z}_{\hat{\varepsilon}}(0)}(0)\big)\,\mathrm{d}\tau \\ &= -\overline{S}_0\int_{\tau_1}^{\tau_2} e^{\mu(T-\tau)}k_{\tau,\hat{z}_{\hat{\varepsilon}}(\tau)}(\hat{\varepsilon}\varphi(\tau))\,\mathrm{d}\tau > 0\,,\end{aligned}$$

which contradicts $\mathbb{E}[L_T(\hat{\Theta})] < \mathbb{E}[L_T(\Theta)]$. □

Now we have all ingredients for the proof of our main result.

Proof of Theorem 3.2.1. By construction, Θ^* is deterministic and thus predictable. Because jumps only occur at initial and terminal time and Θ^* is absolutely continuous in between, it is right-continuous and of bounded variation.

Since an admissible strategy Θ is of bounded variation, the state process $(T-t, Y_t^{\Theta}, \Theta_t)$ is bounded. More precisely, if $C > 0$ is a bound on the total variation of $\Theta = \Theta^+ - \Theta^-$, i.e. $\Theta_T^+(\omega)+\Theta_T^-(\omega) \le C$, then $|\Theta_t| \le C$ and $Y_t \in [\min\{Y_{0-},0\}-C, \max\{Y_{0-},0\}+C]$. Hence, continuity of V gives boundedness of $V(T-t, Y_t, \Theta_t)$ and so the local martingale part in equation (3.11) is a true martingale for every strategy. By Lemmas 3.3.9 and 3.3.10 and thus (3.37) we have that V satisfies the variational (in-)equality (3.12)–(3.13). Together with Lemma 3.3.6 we have that G is a supermartingale for every Θ and a martingale for Θ^*, so by Proposition 3.2.5, up to the factor $\overline{S}_0$, V is indeed the value function of problem (3.7). □

3.4 Solving the problem for monotone strategies

Let us now consider the monotone case of a large investor who cannot perform intermediate buying or short selling. The optimal free boundary $\mathcal{I}_1 \cup \mathcal{I}_2$ that separates wait and sell regions in this case is a mixture of a part $\mathcal{I}_2$ of the free boundary from Theorem 3.2.1 and a suitable enlargement $\mathcal{I}_1$ of the free boundary for infinite horizon trading from Chapter 2.

We will restrict ourselves to a regime of decreasing prices (in expectation), i.e. $\mu < 0$ for the rest of this section. Moreover, in addition to Assumption 3.1.1, we will assume that impact function f and resilience h are such that the terminal position function g from Lemma 3.3.1 is strictly decreasing on $(y_\infty, y_0]$, which is the case for e.g. $f(y) = e^{\lambda y}$ with constant λ. Admissible strategies are

$$\mathcal{A}_{\mathrm{mon}}(\theta) := \big\{\Theta \;\big|\; (\Theta_t)_{t\in[0,T]} \text{ is adapted non-increasing with } \theta \ge \Theta_t \ge \Theta_T = 0\big\}. \quad (3.39)$$

By Remark 3.2.3, the orbit $\mathcal{I}_0 := P([0,T] \times \{y_0\} \times \{0\})$ characterizes the optimal solution to the corresponding infinite horizon problem whenever it liquidates it time T (the optimal infinite horizon boundaries for the problems with buying but no short sales and without buying coincide).

Denote by $\mathcal{I}_1 := \{(\tau + s, y, \theta) \mid (\tau, y, \theta) \in \mathcal{I}_0, s \ge 0\}$ the elongation of $\mathcal{I}_0$ backwards in time. The orbit $\mathcal{I}_0$ splits our surface $\mathcal{I}$ from Theorem 3.2.1 into two regions: by

Lemma 3.3.1, we have that the subset $\mathcal{I}_2 := P([0,T] \times (y_\infty, y_0) \times \{0\}) \subset \mathcal{I}$ contains all orbits $(T-t, Y_t, \Theta_t)_{t\in[0,T)}$ that finish at positive position $\Theta_{T-} = g(Y_{T-}) > 0$ with a block sale, while the other part $\mathcal{I} \setminus (\mathcal{I}_0 \cup \mathcal{I}_2)$ contains all orbits which finish with a terminal block buy, $\Theta_{T-} = g(Y_{T-}) < 0$.

Let $\bar{w}(\tau; y)$ solve $\bar{w}(0; y) = y$ and $\partial_\tau \bar{w}(\tau; y) = h(\bar{w}(\tau; y))$. For $i = 0, 1, 2$ denote

$$\begin{aligned} \mathcal{S}_i &:= \{(\tau, y+d, \theta+d) \mid (\tau, y, \theta) \in \mathcal{I}_i, d \geq 0\}, \\ \mathcal{W}_i &:= \{(\tau+s, \bar{w}(s; y), \theta) \mid (\tau, y, \theta) \in \mathcal{I}_i, s \geq 0\}, \\ \mathcal{W}_3 &:= \{(s, \bar{w}(s; y), \theta) \mid \theta \geq 0, s \geq 0, y < g^{-1}(\theta)\}. \end{aligned} \tag{3.40}$$

Now the optimal sell only problem $\max_{\Theta \in \mathcal{A}_{\mathrm{mon}}(\theta)} \mathbb{E}[L_T(\Theta) \mid Y_{0-} = y, \Theta_{0-} = \theta]$ is solved as follows.

Theorem 3.4.1. *Let f, λ, h, μ satisfy Assumption 3.1.1 and such that additionally $\mu < 0$. Let $w(\cdot; y)$ solve the differential equation $\partial_t w(t; y) = -h(w(t; y))$ with $w(0; y) = y$ and define*

$$\begin{aligned} T_w &:= 0 \vee \sup\{t \in [0,T] \mid (T-t, w(t; Y_{0-}), \Theta_{0-}) \in \mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3\}, \\ \Delta_0 &:= 0 \vee \sup\{d \geq 0 \mid (T, Y_{0-} - d, \Theta_{0-} - d) \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3\}. \end{aligned}$$

Moreover, let $y(t; x)$ and $\theta(t; x, \theta_0)$ solve the differential equations $\partial_t y = -D(y)$ and $\partial_t \theta = \partial_t y + h(y)$ for $t \in [0,T]$, with $y(0; x) = x$ and $\theta(0; x, \theta_0) = \theta_0$. Define the process Θ^* on $[0,T]$ by

$$\Theta_t^* := (\Theta_{0-} - \Delta_0)\mathbb{1}_{[0,T_w)}(t) + \theta\big(t - T_w; w(T_w; Y_{0-}), \Theta_{0-} - \Delta_0\big)\mathbb{1}_{[T_w,T)}(t).$$

Then $\Theta^* \in \mathcal{A}_{mon}(\Theta_{0-})$ maximizes expected proceeds among all $\mathcal{A}_{mon}(\Theta_{0-})$, i.e.

$$\mathbb{E}[L_T(\Theta^*)] = \max_{\Theta \in \mathcal{A}_{mon}(\Theta_{0-})} \mathbb{E}[L_T(\Theta)].$$

Remark 3.4.2. The specification of Θ^* in Theorem 3.4.1 is just a formalization of the following procedure:

1. If $(T, Y_{0-}, \Theta_{0-}) \in \mathcal{W}_1 \cup \mathcal{W}_2$, wait until $(T-t, Y_t^{\Theta^*}, \Theta_t^*) \in \mathcal{I}_1 \cup \mathcal{I}_2$ at time $t = T_w \in [0,T]$, i.e. with $\Theta_t^* = \Theta_{0-}$ on $[0, T_w)$ so that $Y_t^{\Theta^*} = w(t; Y_{0-})$ for $t \in [0, T_w]$.

2. If $(T, Y_{0-}, \Theta_{0-}) \in \mathcal{W}_3$, wait until the end, $T_w = T$, and finish with a single block sale of size $\Delta\Theta_T = -\Theta_{0-}$.

3. If $(T, Y_{0-}, \Theta_{0-}) \in \mathcal{S}_1 \cup \mathcal{S}_2$, do not wait ($T_w = 0$), but perform a block sale of size $|\Delta_0|$ to reach the boundary $\mathcal{I}_1 \cup \mathcal{I}_2$ immediately.

4. If after waiting or initial jump $(T-t, Y_t^{\Theta^*}, \Theta_t^*) \in \mathcal{I}_1$ at time $t = T_w$, trade continuously in rates $\mathrm{d}\Theta_t^*/\mathrm{d}t = (h - D)(Y_t^{\Theta^*})$ along the boundary, keeping $(T-t, Y_t^{\Theta^*}, \Theta_t^*) \in \mathcal{I}_1$, until $\Theta_s^* = 0$ at some time $s \in [T_w, T]$ and stop trading.

5. If after waiting or initial jump $(T-t, Y_t^{\Theta^*}, \Theta_t^*) \in \mathcal{I}_2$ at time $t = T_w$, trade continuously in rates $\mathrm{d}\Theta_t^*/\mathrm{d}t = (h - D)(Y_t^{\Theta^*})$ along the boundary, keeping $(T-t, Y_t^{\Theta^*}, \Theta_t^*) \in \mathcal{I}_2$ for $t \in [T_w, T)$, until time T and perform a final block sale of size $\Delta\Theta_T^* = -\Theta_{T-}^*$.

Proof of Theorem 3.4.1. The proof in $\mathcal{S}_2 \cup \mathcal{S}_1 \cup \mathcal{W}_1$ reduces to existing optimality results. This will be done in Step 1. We will then handle the remaining cases $(T, Y_{0-}, \Theta_{0-}) \in \mathcal{W}_2 \cup \mathcal{W}_3$ by proving the corresponding variational (in)equalities directly. Similarly to Proposition 3.2.5 in the buy-and-sell case, it suffices to prove $\mathcal{L}V = 0$ and $V_y + V_\theta > f$ in $(\mathcal{W}_2 \setminus \mathcal{I}_2) \cup \mathcal{W}_3$ with equality at the boundary surface $\mathcal{I}_2$. This is done in step 2 for $\mathcal{W}_2$ and in step 3 for $\mathcal{W}_3$.

Step 1. Form initial state $(T, Y_{0-}, \Theta_{0-}) \in \mathcal{S}_2$, the solution to the maximization problem over $\mathcal{A}_T \supset \mathcal{A}_{\text{mon}}(\Theta_{0-})$ from Theorem 3.2.1 is monotone, as described in Remark 3.2.4, and coincides with Θ^*.

If $(T, Y_{0-}, \Theta_{0-}) \in \mathcal{S}_1 \cup \mathcal{W}_1$, the solution coincides with the optimal liquidation strategy Θ^* in infinite horizon from Theorem 2.2.4, which liquidates before time T.

Step 2. Consider $(T, Y_{0-}, \Theta_{0-}) \in \mathcal{W}_2$, so that $Y_t^{\Theta^*} = w(t; Y_{0-})$ during the initial waiting period $[0, T_w]$. Note that $w(t; \bar{w}(t; y)) = y$ and the derivative $w_y(t; y)$ of w w.r.t. its initial condition, $w_y(t; y) := \partial_y w(t; y)$, satisfies the differential equation $\frac{\mathrm{d}}{\mathrm{d}t} w_y(t; \bar{w}(t; y)) = -h'(\bar{w}(t; y)) w_y(t; \bar{w}(t; y))$ and $w_y(0; y) = 1$.

Since the boundary surface part $\mathcal{I}_2$ coincides with the boundary surface from the buy-and-sell problem Theorem 3.2.1, the initial waiting time is given by

$$s(\tau, y, \theta) := \inf \{t \geq 0 \mid d(\tau - t, w(t; y), \theta) = 0\}$$

and satisfies $s(\tau, y, \theta) \in [0, T]$. Using (3.33) we find at the boundary $\mathcal{I}_2$, that

$$\Big(\frac{\mathrm{d}}{\mathrm{d}t} d(\tau - t, w(t; y), \theta)\Big)\Big|_{t = s(\tau, y, \theta)} = (D - h')(y_b) = \bar{\theta}'(\tau - s) > 0,$$

where $y_b = w(s(\tau, y, \theta); y)$ and the trading rate $-\bar{\theta}'(\tau - s)$ along the boundary $\mathcal{I}_2$ is negative, as pointed out in Remark 3.2.4. Now, since the partial derivative of d along the trajectory $(\tau - t, w(t; y), \theta)$ is non-zero at $t = s(\tau, y, \theta)$, the implicit function theorem gives $s \in C^1(\mathcal{W}_2)$ (with one-sided derivatives at the boundary $\mathcal{I}_2$).

Let $V^{\mathcal{S}}$ denote the value function in $\mathcal{S}_2$, where it coincides with the buy-and-wait value function from Theorem 3.2.1. Then the value of Θ^* is given in $\mathcal{W}_2$ as

$$V^{\mathcal{W}}(\tau, y, \theta) = e^{\mu s(\tau, y, \theta)} V^{\mathcal{S}}(\tau - s(\tau, y, \theta), w(s(\tau, y, \theta); y), \theta) =: e^{\mu s} V^{\mathcal{S}}.$$

Using that $\mathcal{L}V^{\mathcal{S}}(\tau_b, y_b, \theta_b) \equiv \big(V^{\mathcal{S}}_\tau + h(y_b) V^{\mathcal{S}}_y - \mu V^{\mathcal{S}}\big)(\tau_b, y_b, \theta_b) = 0$ at the boundary $(\tau_b, y_b, \theta_b) \in \mathcal{I}_2$, we get

$$\begin{aligned}
V^{\mathcal{W}}_\tau &= \mu e^{\mu s} V^{\mathcal{S}} s_\tau + e^{\mu s} V^{\mathcal{S}}_\tau \cdot (1 - s_\tau) - h(w) e^{\mu s} V^{\mathcal{S}}_y s_\tau &&= e^{\mu s} V^{\mathcal{S}}_\tau\,, \\
V^{\mathcal{W}}_y &= \mu e^{\mu s} V^{\mathcal{S}} s_y + e^{\mu s} V^{\mathcal{S}}_\tau \cdot (-s_y) - h(w) e^{\mu s} V^{\mathcal{S}}_y s_y + e^{\mu s} V^{\mathcal{S}}_y w_y &&= e^{\mu s} V^{\mathcal{S}}_y w_y\,, \\
V^{\mathcal{W}}_\theta &= \mu e^{\mu s} V^{\mathcal{S}} s_\theta + e^{\mu s} V^{\mathcal{S}}_\tau \cdot (-s_\theta) - h(w) e^{\mu s} V^{\mathcal{S}}_y s_\theta + e^{\mu s} V^{\mathcal{S}}_\theta &&= e^{\mu s} V^{\mathcal{S}}_\theta\,.
\end{aligned}$$

Now fix $(\tau - s(\tau, y, \theta)), w(s(\tau, y, \theta); y), \theta) = (\tau_b, y_b, \theta_b) \in \mathcal{I}_2$ on the boundary and vary $s = s(\tau, y, \theta)$, so that $(\tau, y, \theta) = (\tau_b + s, \bar{w}(s; y_b), \theta_b)$. Let

$$\begin{aligned}
\ell(s) &:= e^{-\mu s} \mathcal{L} V^{\mathcal{W}}(\tau_b + s, \bar{w}(s; y_b), \theta_b) \\
&= V^{\mathcal{S}}_\tau(\tau_b, y_b, \theta_b) + h(\bar{w}(s; y_b)) V^{\mathcal{S}}_y(\tau_b, y_b, \theta_b) w_y(s; \bar{w}(s; y_b)) - \mu V^{\mathcal{S}}(\tau_b, y_b, \theta_b).
\end{aligned}$$

Since $w_y(0; y_b) = 1$, we immediately get $\ell(0) = \mathcal{L}V^{\mathcal{S}}(\tau_b, y_b, \theta_b) = 0$. Now, using that $\frac{\mathrm{d}}{\mathrm{d}s} w_y(s; \bar{w}(s; y_b)) = -h'(\bar{w}(s; y_b)) w_y(s; \bar{w}(s; y_b))$, we get $\ell'(s) = 0$. Hence $\mathcal{L}V^{\mathcal{W}} = 0$ in $\mathcal{W}_2$.

For proving the variational inequality $V_y + V_\theta > f$ in $\mathcal{W}_2 \setminus \mathcal{I}_2$, remember that we have $V_\theta^{\mathcal{S}}(\tau_b, y_b, \theta_b) = \big(f\frac{h\lambda+h'-\mu}{h'}\big)(y_b)$ on the boundary by (3.36) and consider

$$r(s) := \frac{V_y^{\mathcal{W}} + V_\theta^{\mathcal{W}}}{f} = e^{\mu s}\frac{-\big(f\frac{h\lambda-\mu}{h'}\big)(y_b)w_y(s;\bar{w}(s;y_b)) + \big(f\frac{h\lambda+h'-\mu}{h'}\big)(y_b)}{f(\bar{w}(s;y_b))}.$$

Since $w_y(0;y) = 1$ and $\bar{w}(0;y_b) = y_b$, we find $r(0) = 1$. Thus, to show that $r(s) > 1$ for all $s > 0$ it suffices to prove $r'(s) > 0$. Direct calculations give

$$\begin{aligned} &e^{-\mu s}h'(y_b)\frac{f(\bar{w}(s;y_b))}{f(y_b)}r'(s) \\ &\quad = \big(h\lambda-\mu\big)(y_b)\big(h\lambda+h'-\mu\big)(y)w_y - \big(h\lambda-\mu\big)(y)\big(h\lambda+h'-\mu\big)(y_b) =: R(y) \end{aligned}$$

where $y = \bar{w}(s;y_b)$ and $w_y = w_y(s;y)$. Now it suffices to prove $R(y) > 0$ for all $y < y_b$. Since $y_b \in (y_\infty, y_0)$, we have $(h\lambda-\mu)(y) < (h\lambda-\mu)(y_b) \le 0$ and $\big(h\lambda+h'-\mu\big)(y_b) > 0$. Hence, for $y \le y_\infty$, we immediately find $R(y) > \big(h\lambda-\mu\big)(y_b)\big(h\lambda+h'-\mu\big)(y)w_y > 0$. Now consider $y \in (y_\infty, y_b)$. Note that $w_y \in (0,1]$ for all $s \ge 0$, since $\partial_s w_y = -h'w_y$. Hence, we get $R(y) \ge Q(y,y_b) := \big(h\lambda-\mu\big)(y_b)\big(h\lambda+h'-\mu\big)(y) - \big(h\lambda-\mu\big)(y)\big(h\lambda+h'-\mu\big)(y_b)$. However, $y \mapsto Q(y,y_b)$ is decreasing in $[y_\infty, y_0]$, because

$$\partial_y Q(y,y_b) = \big(h\lambda-\mu\big)(y_b)\,\big(h\lambda+h'\big)'(y) - \big(h\lambda\big)'(y)\,\big(h\lambda+h'-\mu\big)(y_b) < 0.$$

So $R(y) > Q(y,y_b) > Q(y_b,y_b) = 0$ for $y_\infty < y < y_b$ and therefore $r'(s) > 0$ for all $s > 0$, which gives $V_y^{\mathcal{W}} + V_\theta^{\mathcal{W}} > f$ in $\mathcal{W}_2 \setminus \mathcal{I}_2$.

Step 3. It remains to consider $\mathcal{W}_3$. In this region, the expected proceeds from Θ^* are given by $V^{\mathcal{W}_3}(\tau,y,\theta) = e^{\mu\tau}\int_0^\theta f(w(\tau;y)-z)\,\mathrm{d}z$ for w as in step 2 above. We have partial derivatives

$$\begin{aligned} V_\tau^{\mathcal{W}_3} &= \mu V^{\mathcal{W}_3} - h(w(\tau;y))e^{\mu\tau}\big(f(w(\tau;y)) - f(w(\tau;y)-\theta)\big)\,, \\ V_y^{\mathcal{W}_3} &= e^{\mu\tau}w_y(\tau;y)\big(f(w(\tau;y)) - f(w(\tau;y)-\theta)\big)\,, \\ V_\theta^{\mathcal{W}_3} &= e^{\mu\tau}f(w(\tau;y)-\theta). \end{aligned}$$

Fix $y_b := w(\tau;y) < g^{-1}(\theta) \le y_0$ and vary τ so that $y = \bar{w}(\tau;y_b)$. Consider

$$\begin{aligned} \ell(\tau) &:= \mathcal{L}V^{\mathcal{W}_3}(\tau,\bar{w}(\tau;y_b),\theta) \\ &= e^{\mu\tau}\underbrace{\big(h(\bar{w}(\tau;y_b))w_y(\tau;\bar{w}(\tau;y_b)) - h(y_b)\big)}_{=:\tilde{\ell}(\tau)}\big(f(w(\tau;y)) - f(w(\tau;y)-\theta)\big). \end{aligned}$$

By $\bar{w}(0;y_b) = y_b$ and $w_y(0;\cdot) = 1$, we find $\tilde{\ell}(0) = 0$. Moreover, differentiation gives $\tilde{\ell}'(\tau) = h'(\bar{w}(\tau;y_b))w_y(\tau;\bar{w}(\tau;y_b)) - \frac{\mathrm{d}}{\mathrm{d}\tau}w_y(\tau;\bar{w}(\tau;y_b)) = 0$ and hence $\ell(\tau) = 0$ for all τ.

For the variational inequality, consider

$$r(\tau) := \frac{V_y^{\mathcal{W}_3} + V_\theta^{\mathcal{W}_3}}{f} = e^{\mu\tau}\frac{\big(f(y_b) - f(y_b-\theta)\big)w_y + f(y_b-\theta)}{f(\bar{w}(\tau;y_b))}.$$

As in step 2, $r(0) = 1$ is immediately clear and we will show $r'(\tau) > 0$ for $\tau > 0$. We have

$$e^{-\mu\tau}f(y)r'(\tau) = -\big(f(y_b) - f(y_b-\theta)\big)\cdot\big(h\lambda+h'-\mu\big)(y)\cdot w_y - f(y_b-\theta)\cdot\big(h\lambda-\mu\big)(y),$$

so for τ large enough such that $y = \bar{w}(\tau; y_b) \leq y_\infty$, it immediately follows that $r'(\tau) > 0$. Now consider the case of $y = \bar{w}(\tau; y_b) > y_\infty$. Since $(h\lambda + h' - \mu)(y) > 0$ in this case and $w_y \leq 1$, we find $e^{-\mu\tau} f(y) r'(\tau)/h'(y) \geq -f(y_b)\big(\frac{h\lambda+h'-\mu}{h'}\big)(y) + f(y_b - \theta)$. Since $y_\infty < y_b < g^{-1}(\theta)$ and g is strictly decreasing on (y_∞, y_0) by assumption, we have $f(y_b - \theta) > f(y_b - g(y_b))$. Hence

$$\begin{aligned}
\frac{e^{-\mu\tau} f(y)}{f(y_b) h'(y)} r'(\tau) &\geq \frac{f(y_b - g(y_b))}{f(y_b)} - \left(\frac{h\lambda + h' - \mu}{h'}\right)(y) \\
&= \left(\frac{h\lambda + h' - \mu}{h'}\right)(y_b) - \left(\frac{h\lambda + h' - \mu}{h'}\right)(y) > 0,
\end{aligned}$$

since $(h\lambda + h' - \mu)/h'$ is increasing on (y_∞, ∞) as noted after Assumption 3.1.1. So we have $r'(\tau) > 0$ and thus $r(\tau) > 1$ for all $\tau > 0$. □

Remark 3.4.3. When time horizon T is short, one might a priori attempt to just follow the infinite horizon solution as long as possible and stop early with a terminal block trade. Such a strategy would be optimal only in special cases like [Kar85] where the controlled diffusion is a controlled Brownian motion. This is not the case here. If such strategy would be optimal, the boundary surface would be constant in τ, i.e. $\mathcal{I}_0 \cup \mathcal{I}_1 \cup \mathcal{I}_2 = \{(\tau, y, \theta) \mid (s, y, \theta) \in \mathcal{I}_0 \text{ for some } s\}$. In particular, in $\mathcal{I}_2$, we would have $g_\tau(y) = g_0(y)$ for the τ-slice $\{(\tau, y, g_\tau(y)) \mid y \in (y_\infty, y_0]\}$ of the boundary given by g_τ from (3.28). However, by Lemma 3.3.4 this is not the case.

Part II Transient price impact with stochastic liquidity

4 Optimal liquidation under stochastic liquidity

This chapter presents an explicit solution of a two-dimensional singular control problem of finite fuel type for infinite time horizon. The problem stems from a modification of the optimal liquidation setup of Chapter 2 considering stochastic liquidity in the sense that the volume effect process, which determines the inter-temporal resilience of the market in spirit of [PSS11], is taken to be stochastic, being driven by own random noise. The optimal control is obtained as the local time of a diffusion process reflected at a non-constant free boundary. To solve the HJB variational inequality and prove optimality, we need a combination of probabilistic arguments and calculus of variations methods, involving Laplace transforms of inverse local times for diffusions reflected at elastic boundaries.

The present chapter is based on the article [BBF18c]. Section 4.2 states the solution for the singular stochastic control problem posed in Section 4.1, and outlines the general course of arguments to come. In Section 4.3, a calculus of variations problem is posed, by restricting to strategies given by diffusions reflected at smooth boundaries. The free boundary is thereby constructed in Section 4.4. By solving the HJB variational inequality (4.9), we prove optimality and derive the value function and the optimal control in Section 4.5. As an extension of the underlying article [BBF18c], Section 4.6 draws the link to optimal stopping and Section 4.7 provides an in-depth comparison to the deterministic liquidity limit (Chapter 2).

4.1 The model and the optimal control problem

We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with two correlated Brownian motions W and B with correlation coefficient $\rho \in [-1, 1]$, such that

$$[W, B]_t = \rho t\,, \quad t \geq 0.$$

for the quadratic co-variation of W and B. The filtration $(\mathcal{F}_t)_{t\geq 0}$ is assumed to satisfy the usual conditions of completeness and right continuity, so we can take càdlàg versions for semimartingales. For notions from stochastic analysis we refer to [JS03].

We consider a market with a risky asset, in addition to the riskless numéraire asset whose (discounted) price is constant at 1. The large investor holds $\Theta_t \geq 0$ shares of the risky asset at time t. She may liquidate her initial position of Θ_{0-} shares by trading according to

$$\Theta_t := \Theta_{0-} - A_t\,,$$

where A is a predictable, càdlàg, monotone process, describing the cumulative number of assets sold up to time t. We define the set of admissible strategies as

$$\mathcal{A}(\Theta_{0-}) := \{A : A \text{ non-decreasing, càdlàg, predictable,} \\ \text{with } 0 =: A_{0-} \leq A_t \leq \Theta_{0-}\}.$$

The unaffected fundamental price $\overline{S} = (\overline{S}_t)_{t\geq 0}$ of the risky asset evolves according to

$$\mathrm{d}\overline{S}_t = \mu \overline{S}_t\,\mathrm{d}t + \sigma \overline{S}_t\,\mathrm{d}W_t\,, \quad \overline{S}_0 \in (0,\infty), \text{ with } \sigma > 0,\ \mu \in \mathbb{R}, \tag{4.1}$$

as a geometric Brownian motion, in the absence of perturbations by large investor trading. By trading, however, the large investor has market impact on the actual price

$$S_t := f(Y_t)\overline{S}_t\,, \tag{4.2}$$

of the risky asset through some impact process Y, by an increasing positive smooth function $f > 0$ with $f(0) = 1$. The process Y can be interpreted as a volume effect process, representing the transient volume displacement by large trades in a limit order book (LOB) whose shape corresponds to the price impact function f as in Remark 2.1.1. For $\hat{\sigma} > 0$ the effect from perturbations $\hat{\sigma}\,\mathrm{d}B_t - \mathrm{d}A_t$ on the process

$$\mathrm{d}Y_t = -\beta Y_t\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t - \mathrm{d}A_t\,, \quad Y_{0-} = y, \tag{4.3}$$

is transient over time, in that Y is mean reverting towards zero with mean reversion rate $\beta > 0$. Existence and uniqueness of a strong solution to (4.3) are guaranteed for instance by [PTW07, Thm. 4.1]. Sometimes we shall write $Y^{y,A}$ to stress the dependence of Y on its initial state y and the strategy A. The dynamics of Y are of Ornstein-Uhlenbeck type, driven by $\hat{\sigma}\,\mathrm{d}B - \mathrm{d}A$. The mean-reversion property of the OU process has the financial interpretation that in the absence of activity from the large trader, the impact lessens since Y reverts back to the neutral state zero and hence the price recovers to the fundamental price $\overline{S}$, thus modeling the transient component of the impact (in absolute terms).

For $\gamma \geq 0$, the γ-discounted proceeds up to time t from a liquidation strategy A are

$$L_t(y; A) := \int_0^t e^{-\gamma u} f(Y_u)\overline{S}_u\,\mathrm{d}A_u^c + \sum_{\substack{0\leq u\leq t\\ \Delta A_u \neq 0}} e^{-\gamma u}\overline{S}_u \int_0^{\Delta A_u} f(Y_{u-} - x)\,\mathrm{d}x, \tag{4.4}$$

for $t \geq 0$, where $A_t = A_t^c + \sum_{u\leq t} \Delta A_u$ is the (pathwise) decomposition of A into its continuous and pure-jump part, and $Y = Y^{y,A}$ solves (4.3). Jump terms in (4.4) can be justified from a LOB perspective (cf. Remark 2.1.1) or by stability results, see equation (5.15) and Theorem 5.2.7 for details. In particular, if $A^n \to A$ converges in the Skorokhod M_1 topology in probability for, e.g., continuous strategies A^n and possibly non-continuous A, then the above definition ensures that $L(y;A^n) \to L(y;A)$ in M_1 in probability.

As L is an increasing process, the limit $L_\infty := \lim_{t\to\infty} L_t$ exists. The large trader's objective is to maximize expected (discounted) proceeds over an infinite time horizon,

$$\max_{A\in\mathcal{A}(\Theta_{0-})} \mathbb{E}[L_\infty(y;A)] \quad \text{with} \quad v(y,\theta) := \sup_{A\in\mathcal{A}(\theta)} \mathbb{E}[L_\infty(y;A)], \tag{4.5}$$

where $v(y,\theta)$ denotes the value function for $y \in \mathbb{R}$ and $\theta \in [0,\infty)$.

Remark 4.1.1. The value function v is increasing in y and θ. Indeed, monotonicity in θ follows from $\mathcal{A}(\theta_1) \subset \mathcal{A}(\theta_2)$ for $\theta_1 \le \theta_2$. For monotonicity in y, note that for $y_1 \le y_2$ and any strategy $A \in \mathcal{A}(\theta)$ one has $Y_t^{y_1,A} \le Y_t^{y_2,A}$ for all t, implying $L_t(y_1;A) \le L_t(y_2;A)$.

For the rest of this chapter, the function f and scalars $\beta,\mu,\gamma,\sigma,\rho,\hat\sigma$ satisfy

Assumption 4.1.2.

C1. We have $\delta := \gamma - \mu > 0$, that means the drift coefficient $-\delta\overline{S}$ for the γ-discounted fundamental price $e^{-\gamma t}\overline{S}_t$ is negative.

C2. The impact function $f \in C^3(\mathbb{R})$ satisfies $f, f' > 0$ and $(f'/f)' < (\Phi'/\Phi)'$, where

$$\Phi(x) := \Phi_\delta(x) := H_{-\delta/\beta}\Big((\sigma\rho\hat\sigma - \beta x)/\big(\sqrt{\beta}\hat\sigma\big)\Big), \qquad x \in \mathbb{R}, \tag{4.6}$$

with Hermite function H_ν (cf. [Leb72, Sect. 10.2]) and $\sigma,\hat\sigma,\beta > 0$ and $\rho \in [-1,1]$.

C3. The impact function f furthermore satisfies $(f'/f)' < (\Phi''/\Phi')'$.

C4. The function $\lambda(y) := f'(y)/f(y)$, $y \in \mathbb{R}$, is bounded, i.e. there exists $\lambda_{\max} \in (0,\infty)$ such that $0 < \lambda(y) \le \lambda_{\max}$ for all $y \in \mathbb{R}$.

C5. The function $k(y) := \frac{\hat\sigma^2}{2}\frac{f''(y)}{f(y)} - (\beta+\delta) + (\sigma\rho\hat\sigma - \beta y)\frac{f'(y)}{f(y)}$, $y \in \mathbb{R}$, is strictly decreasing.

C6. There exist $y_0 \in \mathbb{R}$ such that $(f'/f)(y_0) = (\Phi'/\Phi)(y_0)$ and $y_\infty \in \mathbb{R}$ such that $(f'/f)(y_\infty) = (\Phi''/\Phi')(y_\infty)$.

Assumption 4.1.2 is satisfied by e.g. $f(y) = \exp(\lambda y)$ with $\lambda \in (0,\infty)$, cf. Lemma 4.4.1 below. See Remark 2.1.1 for the shape of the related multiplicative LOB. Note that Φ is (up to a constant factor) the unique positive and increasing solution of the ODE $\frac{\hat\sigma^2}{2}\Phi''(y) + (\sigma\rho\hat\sigma - \beta y)\Phi'(y) - \delta\Phi(y) = 0$.

The overall negative drift in Assumption C1 ensures that the optimization problem on an infinite time horizon has a finite value. Assumptions C2 and C3 imply uniqueness of the (boundary) points y_0 and y_∞ from Assumption C6 which are needed in Lemma 4.4.3. While C3, uniqueness of y_∞, is not crucial there, it will be needed in (4.46) for the verification. The bound on λ in Assumption C4 is used to show some growth condition on the value function in Lemma 4.5.5, that is required to apply the martingale optimality principle (Proposition 4.5.1). Assumption C5 is needed for the verification Lemma 4.5.7.

Let us now comment on the model and its financial interpretation. The price impact function f can be interpreted through a (static) multiplicative limit order book with *volume effect* process Y, as already described in Remark 2.1.1. By $\beta > 0$ we have mean reversion of Y towards the neutral level zero. This transient nature of Y relates to the liquidity property that [Kyl85] calls resilience. Note that in our model the resilience is stochastic in the sense that the volume effect process Y in (4.3) is, whereas the resilience rate β is constant (differently e.g. to [GH17]).

Remark 4.1.3 (The meaning of y_0)**.** The level y_0 can be interpreted as the optimal level of Y where a small investor should sell an infinitesimal amount of assets. The small investor trades at (discounted) prices $e^{-\gamma t}\overline{S}_t f(Y_t) = e^{-\delta t}\overline{S}_0\mathcal{E}(\sigma W)_t f(Y_t)$, but incurs no impact, i.e. cannot control Y. He wants to maximize expected (infinitesimal) gains $\mathbb{E}[e^{-\delta\tau}\overline{S}_0\mathcal{E}(\sigma W)_\tau f(Y_\tau)]$ when selling at time τ. Assume for simplicity that W and B are independent and we can ignore the factor $\mathcal{E}(\sigma W)$. In the general case, a change of measure as in (4.12) below with a priori integrability assumptions on τ would be necessary. Taking w.l.o.g. $\overline{S}_0 = 1$, the small investor now maximizes $\mathbb{E}[e^{-\delta\tau} f(Y_\tau)]$. Due to the Markovian structure of our problem and monotonicity of f, it is natural to assume that the optimal stopping time would be of the form $\tau = \tau_z := \inf\{t > 0 \mid Y_t \geq z\}$ for some (free boundary) level z, see e.g. [PS06] for the connections between optimal stopping and free boundary problems. So for initial $Y_0 = y$, the small investor's problem reduces to $\max_z \mathbb{E}[e^{-\delta\tau_z} f(Y_{\tau_z})] = \max_z f(z \vee y)\mathbb{E}[e^{-\delta\tau_z}]$. The Laplace transform of the hitting time τ_z is well-known, see e.g. [RW87, V.50] and the discussion in Chapter 7. We have $\mathbb{E}[e^{-\delta\tau_z}] = \Phi_\delta(y)/\Phi_\delta(z)$ for $y < z$. Hence we need to maximize $J(z) := f(y \vee z)\Phi(y \wedge z)/\Phi(z) = f(z)\frac{\Phi(y)}{\Phi(z)}\mathbb{1}_{y<z} + f(y)\mathbb{1}_{y\geq z}$ for given initial impact y. We have $J'(z) = \Phi(y)\big(f'/\Phi - f\Phi'/\Phi^2\big)(z)$ for $y < z$ and $J'(z) = 0$ for $y > z$. In particular, Assumption C6 implies $J'(y_0) = 0$. Starting at some $y < y_0$, we indeed find by Assumption C2 that y_0 maximizes J. Hence the small investor's optimal time to sell will be $\tau_{y_0} = \inf\{t \mid Y_t \geq y_0\}$. This is in fact optimal among all stopping times; we will revisit the issue in Remark 4.6.2.

Remark 4.1.4 (Non-deterministic liquidity)**.** Stochasticity may account for variations of transient impact that cannot be entirely explained by the single agent's own trading activity, and thus not solely described by deterministic functional modeling.

(a) Most of the literature on transient impact considers impact that is a deterministic function of the actions of a single large trader. We consider here an application problem for an individual large trader, but we do not want to assume that she is the only large trader in the market, or that she is as an aggregate of all large traders (a possibility mentioned in [Fre98]). The additional stochastic noise term $\hat{\sigma}\,\mathrm{d}B_t$ in (4.3) can be understood as the aggregate influence on the impact by other large 'noise' traders (acting non-strategically). Questions on strategic behavior between multiple traders (like in [SZ17]) are interesting but beyond the present thesis.

(b) Note that the volatility and as well the drift of the (marginal) price process $S = f(Y_t)\overline{S}_t$ from (4.2), at which (additional infinitesimally) small quantities of the risky assets would be traded, are stochastic via the additional stochastic component of Y. Furthermore, we emphasize that the form of relative price impact function $\Delta \mapsto f(Y_{t-} + \Delta)/f(Y_{t-})$ can vary with Y in general. In the sense of Remark 2.1.1, this means the general shape of the corresponding LOB can exhibit stochastic variations from the large trader's perspective.

(c) Recently, [LN19] suggested to model a signal, which predicts the short-term evolution of prices, as an Ornstein-Uhlenbeck process that modulates the drift of the price dynamics. One can interpret stochasticity of Y as such a signal as follows. For $\lambda = f'/f$

being constant, the log-price can be written as $\log S = (\log \overline{S} + \lambda Y^{\mathrm{sig}}) + \lambda Y^{\mathrm{trans},\Theta}$, where Y^{sig} is a mean-reverting signal with $\mathrm{d}Y^{\mathrm{sig}}_t = -\beta Y^{\mathrm{sig}}_t \,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t$ and $Y^{\mathrm{trans},\Theta}$ is the transient impact from trading with $\mathrm{d}Y^{\mathrm{trans},\Theta}_t = -\beta Y^{\mathrm{trans},\Theta}_t \,\mathrm{d}t + \mathrm{d}\Theta_t$. From this perspective, the optimal liquidation strategy will be adaptive to the signal and depend on the correlation between the signal and $\log \overline{S}$, see Theorem 4.2.1 and Remark 4.2.3.

Remark 4.1.5 (Level of interpretation for the model and relation to additive impact)**.** Noting that a bid-ask spread is not modeled explicitly and price impact f (i.e. the LOB shape) is static, we consider the model as being at a mesoscopic level for low-frequency problems, rather than for market microstructure effects in high frequency. At this level and as pointed out in [AKS16, Rmk. 2.2], it is sensible to think of price impact and liquidity costs as being aggregated over various types of orders. The LOB from Remark 2.1.1 should be interpreted accordingly. Note that in this chapter we deal with monotone strategies and thus only one (bid) side of the LOB is relevant. Considering infinite time horizon can be viewed as approximation for a longer horizon with more analytic tractability. Concerning the question of comparison with additive models for transient impact, positivity of asset prices is desirable from a theoretical point of view, relevant for applications with longer time horizons (as they may occur e.g. for large institutional trades, cf. e.g. [CL95], or for hedging problems with longer maturities), and appears to fit better to common models with multiplicative price evolutions like (4.1). See Remark 3.2.7 for a more detailed discussion and further references.

4.2 The optimal strategy and how it will be derived

This section states the main theorem which describes the solution to the singular stochastic control problem, and outlines afterwards the general course of arguments for proving it in the subsequent sections. To explain ideas, let us first motivate how the variational inequality (4.9), being the dynamical programming equation for the optimization problem at hand, is readily suggested by an application of the martingale optimality principle. To this end, consider for an admissible strategy A the process

$$G_t(y;A) := L_t(y;A) + e^{-\gamma t}\overline{S}_t V(Y_t,\Theta_t), \tag{4.7}$$

where $G_{0-}(y;A) = \overline{S}_0 V(Y_{0-},\Theta_{0-})$ and $V \in C^{2,1}(\mathbb{R}\times[0,\infty);[0,\infty))$ is some function. Suppose V can be chosen such that G is a supermartingale. Then one should have

$$\begin{aligned}\overline{S}_0 V(y,\Theta_{0-}) &= \mathbb{E}[G_{0-}(y;A)]\\ &\geq \lim_{T\to\infty}\mathbb{E}[L_T(y;A)] + \lim_{T\to\infty} e^{-\gamma T}\mathbb{E}[\overline{S}_T V(Y_T,\Theta_T)]\\ &= \mathbb{E}[L_\infty(y;A)]\end{aligned}$$

heuristically, provided that the second summand on the right-hand side converges to 0. Hence, for V being such that G is a supermartingale for every admissible strategy A and a martingale for at least one strategy A^*, one can conclude that V is essentially the value function for (4.5) (modulo the factor $\overline{S}_0$). To describe V, one may apply Itô's formula to get

$$
\begin{aligned}
\mathrm{d}G_t = e^{-\gamma t}\overline{S}_t\Big(&\hat{\sigma}V_y(Y_{t-},\Theta_{t-})\,\mathrm{d}B_t + \sigma V(Y_{t-},\Theta_{t-})\,\mathrm{d}W_t \\
&+ \big((\mu-\gamma)V + (\sigma\rho\hat{\sigma} - \beta Y_{t-})V_y + \tfrac{\hat{\sigma}^2}{2}V_{yy}\big)(Y_{t-},\Theta_{t-})\,\mathrm{d}t \\
&+ \big(f - V_y - V_\theta\big)(Y_{t-},\Theta_{t-})\,\mathrm{d}A_t^c \\
&+ \int_0^{\Delta A_t}\big(f - V_y - V_\theta\big)(Y_{t-}-x,\Theta_{t-}-x)\,\mathrm{d}x\Big).
\end{aligned}
\tag{4.8}
$$

Define, with $\delta = \gamma - \mu$, a differential operator on $C^{2,0}$ functions φ by

$$
\mathcal{L}\varphi(y,\theta) := \frac{\hat{\sigma}^2}{2}\varphi_{yy}(y,\theta) + (\sigma\rho\hat{\sigma} - \beta y)\varphi_y(y,\theta) - \delta\varphi(y,\theta).
$$

By equation (4.8), solving the Hamilton-Jacobi-Bellman (HJB) variational inequality

$$
0 = \max\{f - V_y - V_\theta\ ,\ \mathcal{L}V\} \quad \text{with } V(y,0) = 0,\ y \in \mathbb{R}, \tag{4.9}
$$

would suffice for G to be a local (super-)martingale. This suggests the existence of a *sell region* $\mathcal{S}$ (action region) where the $\mathrm{d}A$-integrand $f - V_y - V_\theta$ is zero and it is optimal to trade (i.e. sell), and a *wait region* $\mathcal{W}$ (inaction region) in which the $\mathrm{d}t$-integrand $\mathcal{L}V$ is zero and it is optimal not to trade. Assume that the two regions

$$
\begin{aligned}
\mathcal{S} &= \{(y,\theta) \in \mathbb{R}\times(0,\infty) : \mathrm{y}(\theta) < y\} \quad \text{and} \\
\mathcal{W} &= \{(y,\theta) \in \mathbb{R}\times(0,\infty) : y < \mathrm{y}(\theta)\}
\end{aligned}
$$

are separated by a free boundary $\{(y,\theta) : y = \mathrm{y}(\theta)\}$. An optimal strategy, i.e. a strategy for which G is a martingale, would be described as follows: if $(Y_{0-},\Theta_{0-}) \in \mathcal{S}$, then perform a block sale of size ΔA_0 such that $(Y_0,\Theta_0) = (Y_{0-} - \Delta A_0, \Theta_{0-} - \Delta A_0) \in \partial\mathcal{S}$. Thereafter, if $\Theta_0 > 0$, sell just enough as to keep the process (Y,Θ) within $\overline{\mathcal{W}}$. In this way, the process (Y,Θ) should be described by a diffusion process that is reflected at the boundary $\partial\mathcal{W} \cap \partial\mathcal{S}$ in direction $(-1,-1)$, i.e. there is waiting in the interior and selling at the boundary until all shares are sold, when (Y,Θ) hits $\{(y,0) : \mathrm{y}(0) \le y\}$. For such reflected diffusions, existence and uniqueness follow from classical results, see Remark 4.3.1, and Theorem 4.3.2 provides important characteristics which are key to the subsequent construction of the optimal control. The solution of the optimal liquidation problem is indeed described by the local time process of a diffusion reflected at a boundary which is explicitly given by an ODE. This main result is stated as Theorem 4.2.1 below.

In the following sections, we will find the value function for our stochastic control problem by constructing a classical solution of the variational inequality (4.9). Provided that the key variational inequalities for the (candidate) solution are satisfied, optimality can be verified by typical martingale arguments, see Proposition 4.5.1. Based on results on reflected diffusions from Theorem 4.3.2, we reformulate in Section 4.3 the optimization problem as a (nonstandard) calculus of variations problem. Its solution, derived in Section 4.4, provides a candidate for the free boundary, separating the regions of action and inaction, together with the value function on that boundary. Moreover, we show a (one-sided) local optimality property of the derived boundary (cf. Theorem 4.4.6). This will be crucial in Section 4.5 (cf. proof of Lemma 4.5.7) to verify (4.9) for the candidate value function, constructed there, in order to finally conclude on p. 76 the proof for

Theorem 4.2.1. *Let Assumption 4.1.2 be satisfied. Then the ordinary differential equation*

$$\mathrm{y}'(\theta) = \left(\frac{((\Phi')^2 - \Phi\Phi'')(f'\Phi' - f\Phi'')/\Phi}{(\Phi\Phi'' - (\Phi')^2)f'' + (\Phi'\Phi'' - \Phi\Phi''')f' + (\Phi'\Phi''' - (\Phi'')^2)f} \right)\big(\mathrm{y}(\theta)\big) \tag{4.10}$$

with initial condition $\mathrm{y}(0) = y_0$ admits a unique solution $\mathrm{y} : [0,\infty) \to \mathbb{R}$, that is strictly decreasing and maps $[0,\infty)$ bijectively to $(y_\infty, y_0]$, for y_0 and y_∞ from Assumption C6.

The boundary function y characterizes the solution of problem (4.5) as the strategy $A^* = (\Delta + K)\mathbb{1}_{[\![0,\tau]\!]}$, where $\Delta := \Theta_{0-}\mathbb{1}_{\{Y_{0-} \geq y_0 + \Theta_{0-}\}} + \tilde{\Delta}\mathbb{1}_{\{Y_{0-} < y_0 + \Theta_{0-}, \tilde{\Delta} \geq 0\}}$ with $\tilde{\Delta} \leq \Theta_{0-}$ satisfying $Y_{0-} - \tilde{\Delta} = \mathrm{y}(\Theta_{0-} - \tilde{\Delta})$, and where (Y, K) is the unique continuous adapted process on $[\![0,\tau]\!]$ with non-decreasing K which solves the y-reflected SDE

$$\begin{aligned} Y_t &\leq \mathrm{y}(\Theta_{0-} - \Delta - K_t)\,, \\ \mathrm{d}Y_t &= -\beta Y_t\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t - \mathrm{d}K_t\,, \\ \mathrm{d}K_t &= \mathbb{1}_{\{Y_t = \mathrm{y}(\Theta_{0-} - \Delta - K_t)\}}\,\mathrm{d}K_t\,, \end{aligned}$$

starting in $(Y_{0-} - \Delta, 0)$, for time to liquidation $\tau := \inf\{t \geq 0 : K_t = \Theta_{0-} - \Delta\}$. Moreover, τ has finite moments.

Since y is strictly monotone as we will show in Lemma 4.4.3, the ODE (4.10) is easily solved by the inverse $\mathrm{y} = \theta^{-1}$ of

$$\theta(y) := \int_{y_0}^{y} \left(\frac{(\Phi\Phi'' - (\Phi')^2)f'' + (\Phi'\Phi'' - \Phi\Phi''')f' + (\Phi'\Phi''' - (\Phi'')^2)f}{((\Phi')^2 - \Phi\Phi'')(f'\Phi' - f\Phi'')/\Phi} \right)(x)\,\mathrm{d}x,$$

for $y \in (y_\infty, y_0]$.

Remark 4.2.2. The optimal control A^* acts as follows: 1) If $Y_{0-} \geq y_0 + \Theta_{0-}$, sell everything immediately at time 0 and stop trading; 2) Otherwise, if (Θ_{0-}, Y_{0-}) is such that $\mathrm{y}(\Theta_{0-}) < Y_{0-} < y_0 + \Theta_{0-}$, perform an initial block trade of size $A_0^* := \Delta > 0$ so that $Y_0 = Y_{0-} - \Delta$ is on the boundary $Y_0 = \mathrm{y}(\Theta_0)$. Now being in the wait region $\overline{\mathcal{W}}$, sell as much as to keep with the least effort the state process (Y, Θ) in $\overline{\mathcal{W}}$ until all assets are liquidated at time τ (cf. Figure 4.1: waiting e.g. at times $t \in [87, 100]$ since then impact Y_t is less than $\mathrm{y}(\Theta_t)$).

The inverse local time $\tau_\ell := \inf\{t > 0 : K_t > \ell\}$ is simply how long it takes to liquidate ℓ assets (after an initial block sale). For $\tau > 0$ (case 2 in Remark 4.2.2) its Laplace transform is

$$\mathbb{E}\big[e^{-\alpha\tau_\ell}\big] = \frac{\Phi_\alpha(Y_0)}{\Phi_\alpha(\mathrm{y}(\Theta_0))} \exp\!\left(\int_0^\ell \big(\mathrm{y}'(\Theta_0 - x) + 1\big) \frac{\Phi'_\alpha(\mathrm{y}(\Theta_0 - x))}{\Phi_\alpha(\mathrm{y}(\Theta_0 - x))}\,\mathrm{d}x \right) \tag{4.11}$$

for $\alpha > 0$ and $0 \leq \ell \leq \Theta_0 = \Theta_{0-} - \Delta$, as it will be shown in the proof of Theorem 4.2.1. Using analyticity of Φ_α w.r.t. the parameter α, one easily gets that τ_ℓ has finite moments. Moreover, the Laplace transform (4.11) gives access to the distribution of the time to liquidation τ by efficient numerical inversion, as in e.g. [AW95].

Remark 4.2.3 (Volatility of the fundamental price)**.** If correlation ρ is not zero, the optimal strategy and the shape of the free boundary do depend on the volatility σ of the

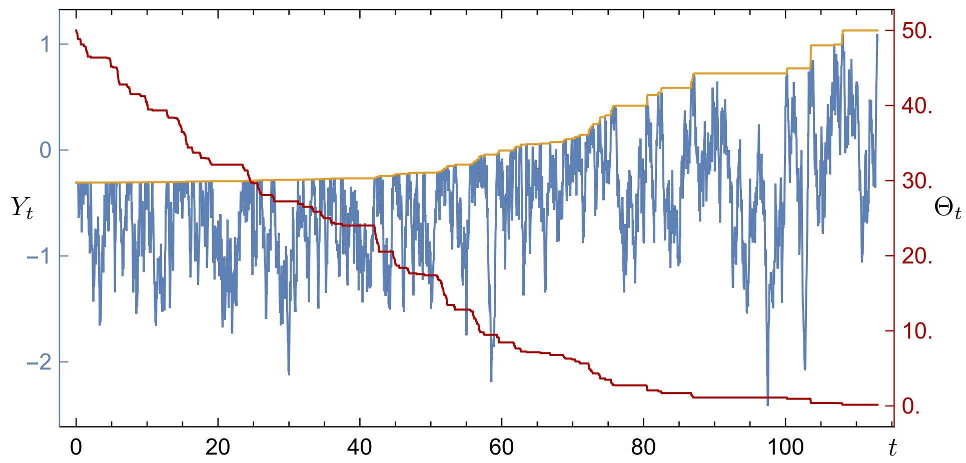


Figure 4.1: Sample path of impact Y_t (blue), asset position Θ_t (red, decreasing) and reflecting boundary $\mathrm{y}(\Theta_t)$ (orange, increasing) for optimally liquidating $\Theta_0 = 50$ assets (after an initial block trade Δ), with $\delta = 0.1$, $\beta = 1$, $\rho = 0$, $\hat{\sigma} = 1$ and $f(\cdot) = \exp(\cdot)$.

fundamental price process. This is a notable difference to many additive impact models, where the optimal liquidation strategy does not depend on the martingale part of the fundamental price process, cf. e.g. [LS13, Sect. 2.2]. To stress the dependence on ρ, we write Φ^ρ for Φ in (4.6), denote by F^ρ the right-hand side of (4.10) and by y_0^ρ the root of $f'/f - (\Phi^\rho)'/\Phi^\rho$. So the solution y^ρ of the ODE $(\mathrm{y}^\rho)'(\theta) = F^\rho\big(\mathrm{y}^\rho(\theta)\big)$ with $\mathrm{y}^\rho(0) = y_0^\rho$ is the optimal boundary function from Theorem 4.2.1. In the special case of constant λ, i.e. $f(y) = e^{\lambda y}$, we have $F^\rho(y) = F^0(y - \sigma\rho\hat{\sigma}/\beta)$ since $\Phi^\rho(y) = \Phi^0(y - \sigma\rho\hat{\sigma}/\beta)$, and thus $\mathrm{y}^\rho(\theta) = \mathrm{y}^0(\theta) + \sigma\rho\hat{\sigma}/\beta$. For general f, investigating y_0 and y_∞ from Assumption C6 still reveals a similar displacement of the boundary. Thus, when impact and fundamental price are positively correlated ($\rho > 0$), it is optimal to trade slower if fundamental price volatility is larger, since the wait region increases.

4.3 Reformulation as a calculus of variations problem

In this section we will recast the free boundary problem of the variational inequality (4.9) as a (nonstandard, at first) calculus of variations problem. To sketch the idea, suppose that the large trader has to liquidate $\Theta_0 \geq 0$ shares and that (Y_0, Θ_0) is already on the free boundary between sell and wait regions (after an initial jump or waiting). Let $\mathrm{y} : [0, \Theta_0] \to \mathbb{R}$ be a C^1 function with $\mathrm{y}(\Theta_0) = Y_0$ and $\mathrm{y}' < 0$ (we expect the optimal boundary to be such). To find the optimal boundary curve y, we will optimize expected proceeds over the set of y-reflected strategies $A := A^{\mathrm{refl}}(\mathrm{y}, \Theta_0)$ from

Definition. Let (Y, A) be the (unique) pair of continuous adapted processes with non-decreasing A such that $Y_t \leq \mathrm{y}(\Theta_0 - A_t)$ and

$$\begin{aligned} \mathrm{d}Y_t &= -\beta Y_t \,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t - \mathrm{d}A_t\,, && Y_0 = \mathrm{y}(\Theta_0)\,,\\ \mathrm{d}A_t &= \mathbb{1}_{\{Y_t = \mathrm{y}(\Theta_0 - A_t)\}}\,\mathrm{d}A_t\,, && A_0 = 0\,, \end{aligned}$$

on $[\![0,\tau]\!]$ for $\tau := \inf\{t \geq 0 : A_t = \Theta_0\}$. We call $A^{\text{refl}}(\mathrm{y}, \Theta_0) := A$ a y-*reflected strategy.*

Remark 4.3.1. Existence and uniqueness of a strong solution (Y, A) follows from (a careful extension of) classical results, cf. [DI93], by considering the pair (Y, A) as a (degenerate) diffusion in $\mathbb{R}^2$ with oblique direction of reflection $(-1, +1)$ at a smooth boundary. Considered as a one-dimensional diffusion, the process Y is reflected at a boundary that moves with its local time A. In this sense, we call the reflection *elastic*. Chapter 7 is devoted to stochastic differential equations with such kind of elastic reflection.

Viewing Y as a diffusion with reflection at y, we can rewrite expected proceeds from A as a deterministic functional of y, see (4.19) below, whose maximizer should describe the optimal strategy. For this step we rely crucially on a representation for the Laplace transform of the inverse local time of reflected diffusions from Theorem 4.3.2. Since the integrand of (4.19) depends on the whole path y, a reparametrization is necessary to obtain a tractable calculus of variations problem (4.21) – (4.22).

Let τ_{Θ_0} be the stopping time when $A = \Theta_0$. For the continuous y-reflected strategy A with proceeds $L := L(\mathrm{y}(\Theta_0); A)$, we have by [DM82, Thm. VI.57] for any $T \in [0, \infty)$,

$$\begin{aligned}\mathbb{E}[L_T] &= \mathbb{E}\Big[\int_0^{\tau_{\Theta_0}\wedge T} f(Y_t)e^{-\delta t}\mathcal{E}(\sigma W)_t \,\mathrm{d}A_t\Big] \\ &= \mathbb{E}\Big[\mathcal{E}(\sigma W)_T \int_0^{\tau_{\Theta_0}\wedge T} f(Y_t)e^{-\delta t}\,\mathrm{d}A_t\Big].\end{aligned}$$

For fixed T, let $\mathbb{Q}$ be the measure given by $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = \mathcal{E}(\sigma W)_T$ on $\mathcal{F}_T$. Then

$$\mathbb{E}[L_T] = \mathbb{E}^{\mathbb{Q}}\Big[\int_0^{\tau_{\Theta_0}\wedge T} f(Y_t)e^{-\delta t}\,\mathrm{d}A_t\Big]. \tag{4.12}$$

Girsanov's theorem gives that the process $\widetilde{B}_t := B_t - [B, \sigma W]_t = B_t - \sigma\rho t$ is a Brownian motion under $\mathbb{Q}$. Therefore, we have under $\mathbb{Q}$

$$\mathrm{d}Y_t = (\sigma\rho\hat{\sigma} - \beta Y_t)\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}\widetilde{B}_t - \mathrm{d}A_t\,,$$

i.e. the impact process Y is a (reflected) Ornstein-Uhlenbeck process with shifted (non-zero) mean reversion level, and A is its local time on the boundary. We cannot directly pass to the limit $T \to \infty$ in (4.12) because the measure change $\mathbb{Q}$ depends on T. However, note that the right-hand side of (4.12) depends only on the law of the reflected diffusion (Y, A) under the measure $\mathbb{Q}$. That is why we consider the reflected diffusion (X, A^X) with the following dynamics under $\mathbb{P}$: for $g(a) := \mathrm{y}(\Theta_0 - a)$ let

$$\mathrm{d}X_t = (\sigma\rho\hat{\sigma} - \beta X_t)\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t - \mathrm{d}A_t^X, \qquad X_0 = g(0)\,, \tag{4.13}$$

$$\mathrm{d}A_t^X = \mathbb{1}_{\{X_t = g(A_t^X)\}}\,\mathrm{d}A_t^X, \qquad A_0^X = 0\,, \tag{4.14}$$

$$\tau_\ell^X := \inf\{t > 0 : A_t^X > \ell \text{ or } A_t^X = \Theta_0\}\,, \tag{4.15}$$

such that in addition $X_t \leq g(A_t^X)$, on $[\![0, \tau_{\Theta_0}^X]\!]$. Existence and uniqueness of a strong solution (X, A^X) until $\tau_{\Theta_0}^X$ follows as in Remark 4.3.1.

Now, by (4.12) we have $\mathbb{E}[L_T] = \mathbb{E}[\int_0^{\tau^X_{\Theta_0}\wedge T} f(X_t)e^{-\delta t}\,\mathrm{d}A^X_t]$, which gives for $T\to\infty$ by monotone convergence on both sides

$$\begin{aligned}\mathbb{E}[L_\infty] &= \mathbb{E}\Big[\int_0^{\tau^X_{\Theta_0}} f(X_t)e^{-\delta t}\,\mathrm{d}A^X_t\Big] = \mathbb{E}\Big[\int_0^{\tau^X_{\Theta_0}} f\big(g(A^X_t)\big)e^{-\delta t}\,\mathrm{d}A^X_t\Big]\\ &= \mathbb{E}\Big[\int_0^{\Theta_0} f\big(g(\ell)\big)e^{-\delta\tau^X_\ell}\,\mathrm{d}\ell\Big] = \int_0^{\Theta_0} f\big(g(\ell)\big)\mathbb{E}\big[e^{-\delta\tau^X_\ell}\big]\,\mathrm{d}\ell\,, \end{aligned}\tag{4.16}$$

using (4.14). To express the latter as a functional of the free boundary only, we need

Theorem 4.3.2. *The Laplace transform of τ^X_ℓ from* (4.13)–(4.15) *for $\Theta_0=\theta$ is*

$$\mathbb{E}\big[e^{-\delta\tau^X_\ell}\big] = \exp\Big(\int_{\theta-\ell}^{\theta}\big(\mathrm{y}'(x)-1\big)\frac{\Phi'_\delta(\mathrm{y}(x))}{\Phi_\delta(\mathrm{y}(x))}\,\mathrm{d}x\Big)\quad \textit{for } \ell<\theta. \tag{4.17}$$

Proof. We will identify the Laplace transform by calculating the terms in (4.16) at first for f being replaced with arbitrary test functions φ, and then using ideas from calculus of variations. To identify $q(y,\theta) := \mathbb{E}[\int_0^T e^{-\delta t}\varphi(X_t)\,\mathrm{d}A^X_t]$ for continuous functions $\varphi:\mathbb{R}\to[0,\infty)$ with $X_0=y\le \mathrm{y}(\theta)$, $\Theta_0=\theta$ and $T:=\tau^X_\theta$, it suffices to construct q such that

$$M_t := \int_0^t e^{-\delta u}\varphi(X_u)\,\mathrm{d}A^X_u + e^{-\delta t}q\big(X_t,\theta-A^X_t\big)$$

is a martingale on $[\![0,T]\!]$ with $e^{-\delta t}q(X_t,\theta-A^X_t)\to 0$ in L^1 as $t\to T$. Consider the state space $\mathcal{I}:=\{(y,\theta): y<\mathrm{y}(\theta)\}$. To check the martingale property, assuming that we have $q\in C^{2,1}(\mathcal{I})\cap C^{1,1}(\overline{\mathcal{I}})$, Itô's formula yields (similarly to (4.8)) that $q_y+q_\theta=\varphi$ on $\partial\mathcal{I}$ and $\mathcal{L}q(y,\theta)=0$ in $\mathcal{I}$. Moreover, for q increasing in y we have $q(y,\theta)=\Phi(y)C(\theta)$ with $\Phi=\Phi_\delta$ from (4.6) and some function $C\in C^1$. Let $H(\theta):=q(\mathrm{y}(\theta),\theta)$. The condition $q_y+q_\theta=\varphi$ leads to

$$H'(\theta) = \Phi'(\mathrm{y}(\theta))C(\theta)\mathrm{y}'(\theta) + \big(\varphi(\mathrm{y}(\theta)) - \Phi'(\mathrm{y}(\theta))C(\theta)\big) = A(\theta)H(\theta)+B(\theta)$$

where $A(\theta):=\big(\mathrm{y}'(\theta)-1\big)\Phi'(\mathrm{y}(\theta))/\Phi(\mathrm{y}(\theta))$ and $B(\theta):=\varphi(\mathrm{y}(\theta))$. Solving this ODE for H gives (since $H(0)=0$)

$$H(\theta) = \int_0^\theta \varphi\big(\mathrm{y}(z)\big)\exp\Big(\int_z^\theta\big(\mathrm{y}'(x)-1\big)\frac{\Phi'_\delta(\mathrm{y}(x))}{\Phi_\delta(\mathrm{y}(x))}\,\mathrm{d}x\Big)\,\mathrm{d}z,$$

which yields the candidate $q(y,\theta)=\Phi(y)H(\theta)/\Phi(\mathrm{y}(\theta))$. It is straightforward to check $q\in C^{2,1}(\mathcal{I})\cap C^{1,1}(\overline{\mathcal{I}})$ and $q_y+q_\theta=\varphi$ on $\partial\mathcal{I}$, giving that M is a martingale, using boundedness of $q_y(X,\theta-A^X)$ on $[\![0,T]\!]$. By monotonicity of q in y, hence $q(y,\theta)\le H(\theta)$, we obtain $e^{-\delta t}q(X_t,\theta-A^X_t)\to 0$ in L^1 as $t\to T$ via dominated convergence, so as in (4.16) we find

$$\int_0^\theta \varphi(\mathrm{y}(z))\underbrace{\Big(\mathbb{E}[e^{-\delta\tau^X_{\theta-z}}] - \exp\Big(\int_z^\theta\big(\mathrm{y}'(x)-1\big)\frac{\Phi'_\delta(\mathrm{y}(x))}{\Phi_\delta(\mathrm{y}(x))}\,\mathrm{d}x\Big)\Big)}_{=:\Delta(z)}\,\mathrm{d}z = 0. \tag{4.18}$$

Note that $z \mapsto \mathbb{E}[\exp(-\delta\tau^X_{\theta-z})]$ is left-continuous. Hence, if $\Delta(z_1) > 0$ for some $z_1 \in (0,\theta]$, there exists $z_0 < z_1$ such that $\Delta > 0$ on (z_0, z_1). Since $\mathtt{y}$ is bijective (recall that $\mathtt{y}' < 0$), we can find a continuous function φ with $\varphi \circ \mathtt{y} > 0$ inside (z_0, z_1) and $\varphi \circ \mathtt{y} = 0$ outside (z_0, z_1), which yields $\int_0^\theta \varphi(\mathtt{y}(z))\Delta(z)\,\mathrm{d}z > 0$, contradicting (4.18). Similarly, $\Delta(z_1) < 0$ also leads to a contradiction. Therefore $\Delta = 0$ on $(0,\theta]$. □

Remark 4.3.3. Let us note that Theorem 4.3.2 generalizes to general (regular) diffusions reflected at increasing boundaries by taking Φ_δ to be the increasing non-negative δ-eigenfunction of the generator of the diffusion. Indeed, the proof would not change. In Chapter 7, we extend Theorem 4.3.2 to non-decreasing reflection boundaries and investigate an approximation scheme for such reflected diffusions, that provides a more intuitive understanding of (4.17) via the total lengths of excursions away from the boundary.

Using Theorem 4.3.2 and (4.16) we derive the following representation for the proceeds from a $\mathtt{y}$-reflected strategy in terms of the boundary:

$$\mathbb{E}[L_\infty] = \int_0^{\Theta_0} f\big(g(\ell)\big) \exp\Big(-\int_0^\ell \big(g'(a)+1\big)\frac{\Phi_\delta'(g(a))}{\Phi_\delta(g(a))}\,\mathrm{d}a\Big)\,\mathrm{d}\ell\,. \tag{4.19}$$

Since the $\mathrm{d}\ell$-integrand in (4.19) depends on the whole path of g, classical calculus of variations methods are not directly available. Since by definition $g(a) = \mathtt{y}(\Theta_0 - a)$ we get with $\mathtt{r}(\ell) := \int_0^\ell (1-\mathtt{y}'(x))\frac{\Phi'(\mathtt{y}(x))}{\Phi(\mathtt{y}(x))}\,\mathrm{d}x$ that

$$\mathbb{E}[L_\infty] = e^{-\mathtt{r}(\Theta_0)}\int_0^{\Theta_0} f\big(\mathtt{y}(\ell)\big)e^{\mathtt{r}(\ell)}\,\mathrm{d}\ell. \tag{4.20}$$

Since $\Phi', \Phi > 0$ and $\mathtt{y}' < 0$, the function $\mathtt{r}$ in strictly increasing and thus has an inverse $\mathtt{r}^{-1}$. Fixing $R := \mathtt{r}(\Theta_0)$ and setting $w(r) := \mathtt{y}(\mathtt{r}^{-1}(r))$, we find

$$\mathtt{r}^{-1}(r) = \int_0^r \left(w'(z) + \frac{\Phi(w(z))}{\Phi'(w(z))}\right)\mathrm{d}z.$$

Hence, by the reparametrization $\mathtt{y}(\theta) = w(\mathtt{r}(\theta))$, finding a maximizing function $\mathtt{y}$ for (4.20) reduces to the problem of finding a function w which maximizes

$$J(w) := \int_0^R f\big(w(r)\big)e^{-(R-r)}\left(w'(r)+\frac{\Phi(w(r))}{\Phi'(w(r))}\right)\mathrm{d}r \quad (=\mathbb{E}[L_\infty]) \tag{4.21}$$

$$\text{subject to the condition } K(w) := \int_0^R \left(w'(r)+\frac{\Phi(w(r))}{\Phi'(w(r))}\right)\mathrm{d}r = \Theta_0\,. \tag{4.22}$$

4.4 Solving the calculus of variations problem

In this section, we solve (locally) the calculus of variations problem of maximizing (4.21) subject to (4.22) by employing necessary and sufficient conditions on the first and second variation of the functionals involved. We obtain the candidate free boundary function $\mathtt{y}(\theta)$, see equations (4.28) and (4.29), and show its local optimality in Lemma 4.4.4. We then relate our results on the calculus of variations problem to the initial optimal execution problem in Theorem 4.4.6. This will be crucial later for Section 4.5 to verify the desired inequality in the sell region, presented in Lemma 4.5.7.

A maximizer w of the isoperimetric problem (4.21) – (4.22) also maximizes $J+mK$ for some constant $m := m(R)$ that is the Lagrange multiplier, cf. [GF00, Theorem 2.12.1]. Considering perturbations $w(r)+h(r)$ of w with $h(0)=h(R)=0$, a necessary condition for an extremum w of a functional $J+mK$ is that its first variation $\boldsymbol{\delta}(J+mK)$ vanishes at w, see [GF00, Thm. 1.3.2]. Integration by parts yields the Euler-Lagrange equation

$$0 = F_w - \frac{\mathrm{d}}{\mathrm{d}r}F_{w'} + \Big(G_w - \frac{\mathrm{d}}{\mathrm{d}r}G_{w'}\Big)m\,, \tag{4.23}$$

with $G(r,w,w') := w' + \Phi(w)/\Phi'(w)$ and $F(r,w,w') := f(w)e^{-(R-r)}G(r,w,w')$, the integrands of K and J, respectively.

Since we assume to start on the (yet unknown) boundary, one side is fixed, i.e. $w(R) = \mathrm{y}(\Theta_0)$. But the other end $w(0)$ is free. Thus, integration by parts of $\boldsymbol{\delta}(J+mK)$ with perturbations $w(r)+h(r)$ of w where $h(0)\neq 0$ imposes as an additional condition for $\boldsymbol{\delta}(J+mK)$ to vanish that

$$0 = \big(F_{w'} + mG_{w'}\big)\big|_{r=0}\,.$$

This *natural boundary condition* (cf. [GF00, Sect. 1.6]) yields the Lagrange multiplier $m(R) = -f(y_0)e^{-R}$ for $y_0 := \mathrm{y}(0) = w(0)$. After multiplication with $e^R\Phi'(w)^2$, equation (4.23) simplifies to

$$e^r\Phi(w)\big(f'(w)\Phi'(w) - f(w)\Phi''(w)\big) = f(y_0)\big(\Phi'(w)^2 - \Phi(w)\Phi''(w)\big)\,. \tag{4.24}$$

Inserting $r=0$ gives a condition for y_0, namely

$$f'(y_0)\Phi(y_0) = f(y_0)\Phi'(y_0).$$

Assumption C6 guarantees existence and C2 uniqueness of y_0. On the other hand, differentiating both sides of (4.24) with respect to r gives the ODE for w

$$\begin{aligned}0 = \big(&e^r(f'\Phi' - f\Phi'')\Phi' + e^r(f''\Phi' - f\Phi''')\Phi - f(y_0)(\Phi'\Phi'' - \Phi\Phi''')\big)w' \\ &+ e^r(f'\Phi' - f\Phi'')\Phi,\end{aligned} \tag{4.25}$$

where $f = f(w(r))$, $f' = f'(w(r))$, $\Phi = \Phi(w(r))$, etc.

Both sides in the above equality (4.24) are negative on the boundary $w(r)$, due to

Lemma 4.4.1. *The positive, increasing eigenfunctions $\Phi = \Phi_\delta$ corresponding to the eigenvalue $\delta > 0$ of the generator of an Ornstein-Uhlenbeck process satisfy*

$$\big(\Phi^{(n)}(x)\big)^2 < \Phi^{(n-1)}(x)\Phi^{(n+1)}(x)$$

for all $x\in\mathbb{R}$ and $n\in\mathbb{N}$. In particular, $(\Phi')^2 < \Phi\Phi''$. Moreover, for $n\in\mathbb{N}$

$$\lim_{x\to-\infty}\Phi^{(n)}(x)/\Phi^{(n-1)}(x) = 0 \quad \textit{and} \quad \lim_{x\to+\infty}\Phi^{(n)}(x)/\Phi^{(n-1)}(x) = +\infty.$$

Proof. Since $H'_\nu(x) = 2\nu H_{\nu-1}(x)$ for complex ν (see [Leb72, eq. (10.4.4)]), equation (4.6) implies

$$\Phi_\delta^{(n)}\Phi_\delta^{(n+2)} - \big(\Phi_\delta^{(n+1)}\big)^2 = \big(\Phi_{\delta+n\beta}\Phi''_{\delta+n\beta} - (\Phi'_{\delta+n\beta})^2\big)\frac{2^{2n}}{\hat\sigma^{2n}\beta^n}\prod_{k=0}^{n}(\delta+k\beta)^2\,,$$

so it suffices to prove $(\Phi')^2 < \Phi''\Phi$ for every $\delta, \beta, \sigma, \hat{\sigma} > 0$ and $\rho \in [-1,1]$ in (4.6). This is equivalent to showing $(H'_\nu)^2 < H''_\nu H_\nu$ for every $\nu < 0$. Since $\Gamma(-\nu) > 0$ and $H_\nu(x) = \Gamma(-\nu)^{-1}\int_0^\infty e^{-t^2-2xt}t^{-\nu-1}\,\mathrm{d}t$ for $\nu < 0$ (cf. [Leb72, eq. (10.5.2)]), the function $\varphi_x(t) := e^{-t^2-2xt}t^{-\nu-1}$ is the density of an absolutely continuous finite measure μ on $[0,\infty)$. For the probability measure $\tilde{\mathbb{P}}[A] := \mu([0,\infty))^{-1}\mu(A)$ consider two independent random variables $X, Y \sim \tilde{\mathbb{P}}$. By [Kle08, Thm. 6.28], we can exchange differentiation and integration (in the integral representation of H_ν above) to see that $H''_\nu(x)H_\nu(x) - H'_\nu(x)^2 = 4\,\tilde{\mathbb{E}}[X^2 - XY]$. Symmetry gives $2\,\tilde{\mathbb{E}}[X^2-XY] = \tilde{\mathbb{E}}[(X-Y)^2] \geq 0$. Since X and Y are independent with absolutely continuous distribution, Fubini's theorem yields $\tilde{\mathbb{P}}[X = Y] = 0$, so $\tilde{\mathbb{E}}[(X-Y)^2] > 0$.

The asymptotic behavior of $\Phi^{(n)}/\Phi^{(n-1)}$ follows from [Leb72, eq. (10.6.4)] in the case $x \to -\infty$ and from [Leb72, eq. (10.6.7)] in the case $x \to +\infty$. □

Now (4.24) gives a representation of r given y_0 and w as

$$r = \log\frac{f(y_0)}{\Phi(w)} + \log\frac{\Phi'(w)^2 - \Phi(w)\Phi''(w)}{f'(w)\Phi'(w) - f(w)\Phi''(w)}, \tag{4.26}$$

which we can use to simplify the ODE (4.25) (assuming $w' \neq 0$ everywhere) to

$$\frac{1}{w'} = -\frac{\Phi'}{\Phi} + \frac{f\Phi''' - f''\Phi'}{f'\Phi' - f\Phi''} + \frac{\Phi'\Phi'' - \Phi\Phi'''}{(\Phi')^2 - \Phi\Phi''},$$

reading the right hand side as a function of $w(r)$. With $\mathbb{y}(\theta) = w(\mathbb{r}(\theta))$ and $r := \mathbb{r}(\theta)$, we get $\mathbb{y}'(\theta) = w'(r)\mathbb{r}'(\theta) = w'(r)(1 - \mathbb{y}'(\theta))\Phi'(\mathbb{y}(\theta))/\Phi(\mathbb{y}(\theta))$, which simplifies to

$$\begin{aligned}\mathbb{y}'(\theta) &= \frac{\Phi'(\mathbb{y})}{\Phi'(\mathbb{y}) + \Phi(\mathbb{y})/w'(r)}\\ &= \frac{1}{\Phi}\frac{((\Phi')^2 - \Phi\Phi'')(f'\Phi' - f\Phi'')}{(\Phi\Phi'' - (\Phi')^2)f'' + (\Phi'\Phi'' - \Phi\Phi''')f' + (\Phi'\Phi''' - (\Phi'')^2)f}\\ &= \frac{M_2(\mathbb{y}(\theta))}{M_1'(\mathbb{y}(\theta))},\end{aligned} \tag{4.27}$$

$$\text{where}\quad M_1 := \frac{f\Phi' - f'\Phi}{(\Phi')^2 - \Phi\Phi''} \quad\text{and}\quad M_2 := \frac{f'\Phi' - f\Phi''}{(\Phi')^2 - \Phi\Phi''}. \tag{4.28}$$

By (4.24) and Lemma 4.4.1 we have $M_2(\mathbb{y}(\theta)) > 0$ for any θ. We get $M_1'(\mathbb{y}(\theta)) < 0$ by

Lemma 4.4.2. *Under Assumption C2, $M_1'(y) < 0$ for all $y \in \mathbb{R}$.*

Proof. Let $G := \Phi'/\Phi$ and $H := \Phi''/\Phi'$. We have $G, G', H, H' > 0$ and $G < H$ by Lemma 4.4.1. With $\lambda(y) = f'(y)/f(y) > 0$, thus $f''/f = \lambda' + \lambda^2$, we get

$$(G')^2\Phi M_1'/f = \lambda' G' + (\lambda^2 - \lambda H)G' + (G^2 - \lambda G)H'.$$

So $M_1'(y) < 0$ if and only if $\lambda'(y)G'(y) < q(\lambda(y))$ where the right-hand side is $q(\lambda) := (H - \lambda)\lambda G' + (\lambda - G)GH'$. The function q is quadratic in λ and takes its minimum in

$$\lambda^* := \frac{HG' + GH'}{2G'} \qquad \text{with value} \qquad q(\lambda^*) = \frac{(HG' + GH')^2}{4G'} - G^2H'.$$

Note also, that $G' = (H-G)G$. We find that

$$\begin{aligned}4G'\left(\lambda' G' - q(\lambda)\right) &\le 4G'\left(\lambda' G' - q(\lambda^*)\right) < 4G'\left((G')^2 - q(\lambda^*)\right)\\ &= 4(G')^3 - (GH' + G'H)^2 + 4G'G^2H'\\ &= G^2\Big(4G(H-G)^3 - \big(H' + (H-G)H\big)^2 + 4(H-G)GH'\Big)\\ &= -G^2\big(H' + H^2 + 2G^2 - 3GH\big)^2 \le 0,\end{aligned}$$

using that $\lambda'(y) < G'(y)$, $y \in \mathbb{R}$, by Assumption C2. So $M_1'(y) < 0$ for all $y \in \mathbb{R}$. □

Lemma 4.4.3. *Let f satisfy Assumptions C2, C3 and C6. Then there exists a unique solution $\theta \mapsto \mathrm{y}(\theta)$, $\theta \in [0,\infty)$, of the ODE*

$$\mathrm{y}' = M_2(\mathrm{y})/M_1'(\mathrm{y}), \quad \mathrm{y}(0) = y_0, \tag{4.29}$$

and y is strictly decreasing to $\lim_{\theta\to\infty} \mathrm{y}(\theta) = y_\infty$ (with y_0 and y_∞ from Assumption C6).

Proof. Since M_2/M_1' is locally Lipschitz by $f \in C^3(\mathbb{R})$, there exists a unique maximal solution $\mathrm{y} : [0,\theta_{\max}) \to \mathbb{R}$ of (4.29). We have $M_2(\mathrm{y}(\theta)) > 0$ and $M_1' < 0$ by Lemma 4.4.2, thus $\mathrm{y}' < 0$. Assume $\theta_{\max} < \infty$, which implies $\lim_{\theta\to\theta_{\max}} \mathrm{y}(\theta) = -\infty$. However, note that $\{(\theta,\mathrm{y}(\theta)) : 0 \le \theta < \theta_{\max}\}$ and $[0,\infty) \times \{y_\infty\}$ are trajectories of the two-dimensional autonomous dynamical system induced by the field $(\theta, y) \mapsto (1, M_2(y)/M_1'(y))$. Since trajectories of autonomous dynamical systems cannot cross, and $y_\infty < y_0$ by Lemma 4.4.1, we must have $y_\infty < \mathrm{y}(\theta)$ for all $\theta \in [0,\theta_{\max})$, which contradicts $\theta_{\max} < \infty$.

Moreover, $\mathrm{y}^{-1}(y) = \int_{y_0}^{y} (M_1'/M_2)(x)\,\mathrm{d}x$ is finite for every $y \in (y_\infty, y_0]$. Since $\theta_{\max} = \infty$, it follows that $\mathrm{y}(\theta) \to y_\infty$ as $\theta \to \infty$. □

By considering the first variation $\boldsymbol{\delta}(J+mK)$, we found a candidate boundary function y in terms of a possible extremum $w : [0,R] \to \mathbb{R}$ of $J+mK$. Calculating the second variation $\boldsymbol{\delta}^2(J+mK)$ at w, we find that w is indeed a local maximizer.

Lemma 4.4.4. *The functional $\hat{J} := J + mK : C^1([0,R]) \to \mathbb{R}$ defined by (4.21) – (4.22) with $m = -f(y_0)e^{-R}$ has a strict local maximizer $w(r) = \mathrm{y}(\mathrm{r}^{-1}(r))$, with y solving (4.29), in the following sense. There exists $\varepsilon > 0$ such that for all perturbations $0 \not\equiv h \in C^1([0,R])$ with endpoints $h(0) = h(R) = 0$ and $\|h\|_{W^{1,\infty}} = \|h\|_\infty \vee \|h'\|_\infty < \varepsilon$ it holds*

$$\hat{J}(w+h) < \hat{J}(w).$$

Proof. For a C^1-perturbation $h : [0,R] \to \mathbb{R}$ of w with $h(0) = h(R) = 0$ we have by [GF00, Sect. 5.25, (10) and (11)]

$$\boldsymbol{\delta}^2(J+mK)[w;h] = \int_0^R \left(Ph'(r)^2 + Qh(r)^2\right)\mathrm{d}r$$

with $P = P(r, w(r), w'(r))$ and $Q = Q(r, w(r), w'(r))$ given by

$$\begin{aligned}P &= \tfrac{1}{2}\big(F_{w'w'} + mG_{w'w'}\big) = 0,\\ Q &= \frac{1}{2}\Big(F_{ww} + mG_{ww} - \frac{\mathrm{d}}{\mathrm{d}r}\Big(F_{ww'} + mG_{ww'}\Big)\Big)\\ &= \frac{1}{2}e^{-(R-r)}\left(\frac{\Phi}{\Phi'}f'' + 2\Big(\frac{\Phi}{\Phi'}\Big)'f' + \Big(\frac{\Phi}{\Phi'}\Big)''f - f'\right) + \frac{1}{2}\Big(\frac{\Phi}{\Phi'}\Big)''m,\end{aligned}$$

with f, Φ and their derivatives being evaluated at $w(r)$ when no argument is mentioned. Differentiating (4.23) with respect to r yields

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}r} e^{-(R-r)}\left(\frac{\Phi}{\Phi'} f' + \left(\frac{\Phi}{\Phi'}\right)' f - f\right) + m\,\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{\Phi}{\Phi'}\right)' \\
&= e^{-(R-r)}\left(\frac{\Phi}{\Phi'} f' + \left(\frac{\Phi}{\Phi'}\right)' f - f\right) \\
&\quad + e^{-(R-r)}\left(\frac{\Phi}{\Phi'} f'' + 2\left(\frac{\Phi}{\Phi'}\right)' f' + \left(\frac{\Phi}{\Phi'}\right)'' f - f'\right) w' + \left(\frac{\Phi}{\Phi'}\right)'' m\, w' \\
&= e^{-(R-r)}\left(\frac{\Phi}{\Phi'} f' + \left(\frac{\Phi}{\Phi'}\right)' f - f\right) + 2Qw' \\
&= e^{-(R-r)} \frac{\Phi}{(\Phi')^2}\left(f'\Phi' - f\Phi''\right) + 2Qw' \,.
\end{aligned} \tag{4.30}$$

By equation (4.24) and Lemma 4.4.1, the first summand in (4.30) is negative along $w(r)$. Since $w(r) = \mathbb{y}(\mathbb{r}^{-1}(r))$ and $\mathbb{r}^{-1}$ is strictly increasing, we have $w' < 0$ by Lemma 4.4.3. So $Q(r, w(r), w'(r)) < -\kappa < 0$ on $[0, R]$ by (4.30) for some constant $\kappa = \kappa_R$, giving that the second variation is negative definite at w, i.e. for $h \not\equiv 0$,

$$\boldsymbol{\delta}^2(J+mK)[w;h] = \int_0^R Q(r, w(r), w'(r)) h(r)^2 \,\mathrm{d}r < -\kappa \int_0^R h(r)^2 \,\mathrm{d}r < 0 \,. \tag{4.31}$$

To shorten notation, let $\hat{F} := F + mG$, so $\hat{J} := J + mK = \int_0^R \hat{F}\,\mathrm{d}r$. Unless the arguments are explicitly written, take $\hat{F} = \hat{F}(r, w(r), w'(r))$. Taylor's theorem gives $\hat{J}(w+h) - \hat{J}(w) = \boldsymbol{\delta}\hat{J}[w;h] + \boldsymbol{\delta}^2\hat{J}[w;h] + \mathcal{E}(h)$ with first variation $\boldsymbol{\delta}\hat{J}[w;h] = 0$ by (4.23), second variation $\boldsymbol{\delta}^2\hat{J}[w;h] = \int_0^R Qh^2\,\mathrm{d}r < 0$ by (4.31) and remainder

$$\mathcal{E}(h) = \int_0^R \left(\sum_{|\boldsymbol{\alpha}|=3} \partial^{\boldsymbol{\alpha}} \hat{F}\big(r, \boldsymbol{w} + \xi_r \boldsymbol{h}\big) \frac{\boldsymbol{h}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}\right) \mathrm{d}r$$

for some $\xi_r \in [0,1]$, with $\boldsymbol{w} = (w(r), w'(r))^\top$, $\boldsymbol{h} = (h(r), h'(r))^\top$ and multi-index $\boldsymbol{\alpha} \in \mathbb{N}_0^2$, considering $\hat{F}(r, \cdot)$ as an function on $\mathbb{R}^2$. Since $\hat{F}$ is affine in w' we get

$$\mathcal{E}(h) = \int_0^R \Big(\frac{1}{6}\hat{F}_{www}(r, \boldsymbol{w} + \xi_r \boldsymbol{h})h + \frac{1}{2}\hat{F}_{www'}(r, \boldsymbol{w} + \xi_r \boldsymbol{h})h'\Big) h^2 \,\mathrm{d}r =: \int_0^R A h^2 \,\mathrm{d}r$$

Note that by compactness of $[0, R]$ we have uniform convergence

$$\sup_{r \in [0,R]} \sup_{\xi \in [0,1]} \big|A\big(h(r), h'(r), w(r), w'(r), \xi, r\big)\big| \to 0$$

as $\|h\|_{W^{1,\infty}} \to 0$. Now choose $\varepsilon > 0$ small enough such that

$$\big|A\big(h(r), h'(r), w(r), w'(r), \xi, r\big)\big| < \kappa/2$$

for all $r \in [0, R]$, $\xi \in [0, 1]$ and h with $\|h\|_{W^{1,\infty}} < \varepsilon$. Hence, with $h \not\equiv 0$

$$\hat{J}(w+h) - \hat{J}(w) = \int_0^R (Q + A)h^2 \,\mathrm{d}r < -\frac{\kappa}{2} \int_0^R h^2 \,\mathrm{d}r < 0 \,. \qquad \square$$

Note that the definition $w(r) := \mathbb{y}(\mathbb{r}^{-1}(r))$ does not depend on the interval boundary R. Hence the optimizer w over $[0, R]$ from Lemma 4.4.4 is optimal for all $R > 0$. We can calculate the value $J(w)$ of our optimizer explicitly.

Lemma 4.4.5. *For the optimal w from Lemma 4.4.4 we have*

$$J(w) = (\Phi M_1)(\mathbb{y}(\Theta_0)) = (\Phi M_1)(w(R)).$$

Proof. By direct calculation we have $fM_1'/(\Phi M_2^2) = ((f\Phi' - f'\Phi)/(f'\Phi' - f\Phi''))'$. Moreover, (4.24) gives $e^r = f(y_0)/(\Phi M_2)(w(r))$. With $r = \mathbb{r}(\ell)$ and using (4.29), we get from (4.20) that

$$\begin{aligned} J(w) &= e^{-\mathbb{r}(\Theta_0)} \int_0^{\Theta_0} f(\mathbb{y}(\ell)) e^{\mathbb{r}(\ell)} \,\mathrm{d}\ell \\ &= (\Phi M_2)(\mathbb{y}(\Theta_0)) \int_0^{\Theta_0} \Big(\frac{f}{\Phi M_2}\Big)(\mathbb{y}(\ell)) \,\mathrm{d}\ell \\ &= (\Phi M_2)(\mathbb{y}(\Theta_0)) \int_{y_0}^{\mathbb{y}(\Theta_0)} \Big(\frac{fM_1'}{\Phi M_2^2}\Big)(x) \,\mathrm{d}x \\ &= (\Phi M_2)(\mathbb{y}(\Theta_0)) \Big[\frac{f\Phi' - f'\Phi}{f'\Phi' - f\Phi''}\Big]_{y_0}^{\mathbb{y}(\Theta_0)} \\ &= (\Phi M_1)(\mathbb{y}(\Theta_0)). \end{aligned}$$

□

Now we can translate the results obtained so far back to the state space of impact and asset position. The following theorem will be crucial for our analysis in the verification arguments in Section 4.5.

Theorem 4.4.6. *The function $\mathbb{y} : [0, \infty) \to \mathbb{R}$ defined by equation (4.29) is a (one-sided) local maximizer of $\mathbb{E}[L_\infty(A^{refl}(\mathbb{y}, \Theta_0))]$ in the sense that, for every $\theta > 0$ there exists $\varepsilon > 0$ such that for any decreasing $\tilde{\mathbb{y}} \in C^1([0, \infty))$ with $\mathbb{y}(\cdot) \le \tilde{\mathbb{y}}(\cdot) \le y_0$, $\mathbb{y} = \tilde{\mathbb{y}}$ on $[\theta, \infty)$ and $0 < \|\mathbb{y} - \tilde{\mathbb{y}}\|_{W^{1,\infty}} < \varepsilon$, it holds*

$$\mathbb{E}\big[L_\infty\big(A^{refl}(\mathbb{y}, \theta)\big)\big] > \mathbb{E}\big[L_\infty\big(A^{refl}(\tilde{\mathbb{y}}, \theta)\big)\big].$$

Proof. For sake of clarity, we write $J = J_R$ and $K = K_R$ to emphasize the dependence of the functionals J, K on R. Call $w(r)$ the parametrization of $\mathbb{y}$ and $\tilde{w}(r)$ the parametrization of $\tilde{\mathbb{y}}$.

Fix $\theta > 0$ and choose $R, \hat{R}, \hat{\theta}$ such that $\mathbb{y}(\theta) = w(R)$, $\tilde{\mathbb{y}}(\theta) = \tilde{w}(\hat{R})$ and $w(\hat{R}) = \mathbb{y}(\hat{\theta})$. So $R := \mathbb{r}_{\mathbb{y}}(\theta)$, $\hat{R} := \mathbb{r}_{\tilde{\mathbb{y}}}(\theta) = \int_0^\theta \frac{\Phi'}{\Phi}(\tilde{\mathbb{y}}(x))\,\mathrm{d}x + \int_{\tilde{\mathbb{y}}(\theta)}^{\tilde{\mathbb{y}}(0)} \frac{\Phi'}{\Phi}(u)\,\mathrm{d}u$ and $\hat{\theta} := \mathbb{r}_{\mathbb{y}}^{-1}(\hat{R})$. By $\mathbb{y} \not\equiv \tilde{\mathbb{y}}$, $\mathbb{y}(\cdot) \le \tilde{\mathbb{y}}(\cdot)$ with equality outside $(0, \theta)$ and monotonicity of Φ'/Φ, we have $\hat{R} > R$ and thus $\hat{\theta} > \theta$.

Now, $K_{\hat{R}}(w) = \hat{\theta}$ and $K_{\hat{R}}(\tilde{w}) = \theta$. Moreover, $J_r(w) = (\Phi M_1)(w(r))$ by Lemma 4.4.5. So if $\|w - \tilde{w}\|_{W^{1,\infty}}$ is small enough, by Lemma 4.4.4 we get

$$\begin{aligned} J_R(w) &= \big(\Phi M_1\big)\big(w(R)\big) - \big(\Phi M_1\big)\big(w(\hat{R})\big) + J_{\hat{R}}(w) \\ &= \big(\Phi M_1\big)\big(w(R)\big) - \big(\Phi M_1\big)\big(w(\hat{R})\big) + e^{-\hat{R}} f(y_0)\hat{\theta} + \big(J_{\hat{R}} - e^{-\hat{R}} f(y_0) K_{\hat{R}}\big)(w) \\ &> \big(\Phi M_1\big)\big(w(R)\big) - \big(\Phi M_1\big)\big(w(\hat{R})\big) + e^{-\hat{R}} f(y_0)\hat{\theta} + \big(J_{\hat{R}} - e^{-\hat{R}} f(y_0) K_{\hat{R}}\big)(\tilde{w}) \\ &= \big(\Phi M_1\big)\big(\mathbb{y}(\hat{\theta} - \eta)\big) - \big(\Phi M_1\big)\big(\mathbb{y}(\hat{\theta})\big) + e^{-\hat{R}} f(y_0)\eta + J_{\hat{R}}(\tilde{w}) \\ &=: \Psi(\eta) + J_{\hat{R}}(\tilde{w})\,. \end{aligned}$$

where $\eta := \hat{\theta} - \theta > 0$. By (4.26) we get $e^{-\hat{R}} f(y_0) = (\Phi M_2)(\mathrm{y}(\hat{\theta}))$. With (4.27) follows

$$\begin{aligned}\Psi'(\eta) &= -\left((\Phi M_1)' \frac{M_2}{M_1'}\right)\big(\mathrm{y}(\hat{\theta} - \eta)\big) + \big(\Phi M_2\big)\big(\mathrm{y}(\hat{\theta})\big) \\ &= -\left(\frac{\Phi' M_1 M_2}{M_1'} + \Phi M_2\right)\big(\mathrm{y}(\hat{\theta} - \eta)\big) + \big(\Phi M_2\big)\big(\mathrm{y}(\hat{\theta})\big).\end{aligned}$$

Hence $\Psi'(0) = -(\Phi' M_1 M_2 / M_1')(\mathrm{y}(\hat{\theta}))$. Since $M_1 > 0$ on $(-\infty, y_0)$, $M_2 > 0$ on $(y_\infty, y_0]$, $M_1' < 0$ by Lemma 4.4.2 and $\Phi' > 0$, it follows $\Psi'(0) > 0$. So $\Psi(\eta) > 0$ for $\eta > 0$ small enough. Therefore we have by (4.21)

$$\mathbb{E}\big[L_\infty\big(A^{\text{refl}}(\mathrm{y}, \theta)\big)\big] = J_R(w) > J_{\hat{R}}(\tilde{w}) = \mathbb{E}\big[L_\infty\big(A^{\text{refl}}(\tilde{\mathrm{y}}, \theta)\big)\big].$$

The bounds on η and $\|w - \tilde{w}\|_{W^{1,\infty}}$ are satisfied for small enough $\varepsilon > 0$, because $(\mathrm{y}, \ell) \mapsto \mathrm{r}_\mathrm{y}(\ell)$ and $(\mathrm{y}, \ell) \mapsto \mathrm{r}_\mathrm{y}^{-1}(\ell)$ are continuous in $W^{1,\infty} \times \mathbb{R}$, so $\|w - \tilde{w}\|_{W^{1,\infty}} \to 0$, $\hat{R} \to R$ and $\hat{\theta} \to \theta$ as $\varepsilon \to 0$. □

4.5 Constructing the value function and verification

In this section, we construct a candidate for the value function and verify the variational inequality (4.9) in Lemmas 4.5.6 and 4.5.7, relying on results from the previous sections. This will be sufficient to conclude the proof of our main result, Theorem 4.2.1.

Having defined a candidate boundary via the ODE (4.29) to separate the sell and wait regions $\mathcal{S}$ and $\mathcal{W}$, we will now construct a solution V of the variational inequality (4.9) that will give the value function of the optimal liquidation problem. As a direct consequence of Lemma 4.4.5, we get its value along the boundary

$$V_{\text{bdry}}(\theta) := V\big(\mathrm{y}(\theta), \theta\big) = \Phi\big(\mathrm{y}(\theta)\big) M_1\big(y(\theta)\big)\,. \tag{4.32}$$

Inside the wait region $\mathcal{W}$, which we assume is to the left of the boundary, we require $V = V^{\mathcal{W}}$ to satisfy $\frac{\hat{\sigma}^2}{2} V_{yy} + (\sigma\rho\hat{\sigma} - \beta y) V_y = \delta V$. Note that $V^{\mathcal{W}}$ solves the same ODE in y as Φ. Since V should be also monotonically increasing, the only possibility is that $V^{\mathcal{W}}(y, \theta) = C(\theta)\Phi(y)$ for some increasing function $C : [0, \infty) \to [0, \infty)$. Using the boundary condition $V^{\mathcal{W}}(\mathrm{y}(\theta), \theta) = V_{\text{bdry}}(\theta)$, in light of equation (4.32) we then have

$$V^{\mathcal{W}}(y, \theta) := \Phi(y) C(\theta) \tag{4.33}$$

for $y \le \mathrm{y}(\theta)$ and $\theta \ge 0$, where $C(\theta) := M_1(\mathrm{y}(\theta))$. On the other hand, in the sell region we require for $V = V^{\mathcal{S}}$ to satisfy $f = V_y^{\mathcal{S}} + V_\theta^{\mathcal{S}}$. We divide $\mathcal{S}$ in two parts:

$$\begin{aligned}\mathcal{S}_1 &:= \{(y, \theta) \in \mathbb{R} \times (0, \infty) : \mathrm{y}(\theta) < y < y_0 + \theta\}\,, \\ \mathcal{S}_2 &:= \{(y, \theta) \in \mathbb{R} \times (0, \infty) : y_0 + \theta < y\}\,.\end{aligned}$$

Let $\Delta := \Delta(y, \theta) \ge 0$ denote the $\|\cdot\|_\infty$-distance of a point $(y, \theta) \in \overline{\mathcal{S}}$ to the boundary $\partial\mathcal{S}$ in direction $(-1, -1)$. This means in $\overline{\mathcal{S}}_1$ (but not in $\mathcal{S}_2$) that

$$\mathrm{y}(\theta - \Delta) = y - \Delta\,. \tag{4.34}$$

Inside $\overline{\mathcal{S}}_1$, we need to have

$$V^{\mathcal{S}_1}(y,\theta) := V^{\mathcal{W}}(y-\Delta,\theta-\Delta) + \int_{y-\Delta}^{y} f(x)\,\mathrm{d}x\,, \tag{4.35}$$

since $V^{\mathcal{S}_1}_y + V^{\mathcal{S}_1}_\theta = f$ in $\overline{\mathcal{S}}$ and $V^{\mathcal{S}_1}(\mathrm{y}(\theta),\theta) = V^{\mathcal{W}}(\mathrm{y}(\theta),\theta)$. Similarly, in $\overline{\mathcal{S}}_2$,

$$V^{\mathcal{S}_2}(y,\theta) := \int_{y-\theta}^{y} f(x)\,\mathrm{d}x. \tag{4.36}$$

To wrap up, the candidate value function is defined by:

$$V = V^{\mathcal{W}} \text{ on } \overline{\mathcal{W}}, \quad V = V^{\mathcal{S}_1} \text{ on } \overline{\mathcal{S}}_1, \quad V = V^{\mathcal{S}_2} \text{ on } \overline{\mathcal{S}}_2. \tag{4.37}$$

The rest of this section is devoted to verifying that V is a classical solution of the HJB variational inequality (4.9) and thus concluding the proof of Theorem 4.2.1 by an application of the martingale optimality principle. We first formalize the heuristic verification from Section 4.2.

4.5.1 Martingale optimality principle

Recall that v is the value function of the optimal liquidation problem (cf. (4.5)). Analogous to Proposition 2.2.8 for deterministic Y, we have can apply martingale optimality in our setup of stochastic Y.

Proposition 4.5.1 (Martingale optimality principle).
Consider a $C^{2,1}$ function $V : \mathbb{R} \times [0,\infty) \to [0,\infty)$ with the following properties:

1. *For every $\Theta_{0-} \geq 0$, there exist constants C_1, C_2 so that*

   $$V(y,\theta) \leq C_1 \exp(C_2 y) \vee 1 \quad \text{for all } (y,\theta) \in \mathbb{R} \times [0,\Theta_{0-}];$$

2. *For every $\Theta_{0-} \geq 0$ and $A \in \mathcal{A}(\Theta_{0-})$, the process G from (4.7) is a supermartingale, where $Y = Y^{y,A}$ is defined in (4.3), and additionally $G_0(y;A) \leq G_{0-}(y;A)$.*

Then we have $\overline{S}_0 \cdot V(y,\theta) \geq v(y,\theta)$.
Moreover, if there exists $A^* \in \mathcal{A}(\Theta_{0-})$ such that $G(y;A^*)$ is a martingale starting in $G_0(y;A^*) = G_{0-}(y;A^*)$, then we have $\overline{S}_0 V(y,\theta) = v(y,\theta)$ and $v(y,\theta) = \mathbb{E}[L_\infty(y;A^*)]$ for $\Theta_{0-} = \theta \geq 0$. In this case, any strategy A for which $G(y;A)$ is not a martingale would be suboptimal.

Proof. By the supermartingale property we have for every $T \geq 0$

$$\begin{aligned} \overline{S}_0 V(Y_{0-},\Theta_{0-}) &\geq \mathbb{E}[G_0(y;A)] \geq \mathbb{E}[L_T(y;A) + e^{-\gamma T}\overline{S}_T V(Y_T,\Theta_T)] \\ &= \mathbb{E}[L_T(y;A)] + e^{-\gamma T}\mathbb{E}[\overline{S}_T V(Y_T,\Theta_T)] \\ &= \mathbb{E}[L_T(y;A)] + e^{-\delta T}\overline{S}_0 \mathbb{E}[\mathcal{E}(\sigma W)_T V(Y_T,\Theta_T)]. \end{aligned} \tag{4.38}$$

By monotone convergence, the first summand in (4.38) tends to $\mathbb{E}[L_\infty(y;A)]$ for $T \to \infty$. To see that the second summand converges to 0, consider the Ornstein-Uhlenbeck process $\mathrm{d}X_t = -\beta X_t\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t,\ X_0 = y$. An application of Itô's formula gives

$$e^{\beta t}(Y_t - X_t) = \int_{[0,t]} e^{\beta u}\,\mathrm{d}\Theta_u \quad \forall t \geq 0. \tag{4.39}$$

Since Θ is non-increasing, we conclude $Y_t \le X_t$ for all $t \ge 0$. Let $p, q > 1$ be conjugate, i.e. $1 = 1/q + 1/p$. Using Hölder's inequality and the bound on V,

$$\begin{aligned}
\mathbb{E}\big[\mathcal{E}(\sigma W)_T V(Y_T, \Theta_T)\big] &\le \mathbb{E}\big[\mathcal{E}(\sigma W)_T^p\big]^{1/p} \mathbb{E}\big[V(Y_T, \Theta_T)^q\big]^{1/q} \\
&= \mathbb{E}\big[\exp\big(p\sigma W_T - \tfrac{1}{2}p\sigma^2 T\big)\big]^{1/p} \mathbb{E}[V(Y_T, \Theta_T)^q]^{1/q} \\
&= \mathbb{E}[\mathcal{E}(p\sigma W)_T]^{1/p} \exp\Big(\tfrac{1}{p}\big(\tfrac{1}{2}p^2\sigma^2 T - \tfrac{1}{2}p\sigma^2 T\big)\Big) \mathbb{E}\big[V(Y_T, \Theta_T)^q\big]^{1/q} \\
&= \exp\Big(\frac{p-1}{2}\sigma^2 T\Big) \mathbb{E}[V(Y_T, \Theta_T)^q]^{1/q} \\
&\le \exp\Big(\frac{p-1}{2}\sigma^2 T\Big) \mathbb{E}[C_1^q \exp(qC_2 Y_T) \vee 1]^{1/q} \\
&\le \exp\Big(\frac{p-1}{2}\sigma^2 T\Big) \mathbb{E}[C_1^q \exp(qC_2 X_T) \vee 1]^{1/q}.
\end{aligned}$$

Using the fact that X is a Gaussian process with mean $\mathbb{E}[X_T] = ye^{-\beta T}$ and variance $\operatorname{Var}(X_T) = \frac{\hat{\sigma}^2}{2\beta}(1 - e^{-2\beta T})$, we get for $K := \mathbb{E}[C_1^q \exp(qC_2 X_T) \vee 1]$ that

$$\begin{aligned}
K &\le 1 + C_1^q \exp\Big(qC_2 \mathbb{E}[X_T] + \frac{1}{2}q^2 C_2^2 \operatorname{Var}(X_T)\Big) \\
&\le 1 + C_1^q \exp\Big(qC_2 y + \frac{\hat{\sigma}^2}{4\beta}q^2 C_2^2\Big).
\end{aligned}$$

This bound on K is independent of T. Now choosing $p > 1$ such that $\frac{p-1}{2}\sigma^2 < \delta$ ensures that $\exp(-\delta T)\exp(\frac{p-1}{2}\sigma^2 T)$ is exponentially decreasing in T, and thus the second summand in (4.38) converges to 0 for $T \to \infty$. This implies that $\overline{S}_0 V(y, \theta) \ge \mathbb{E}[L_\infty(y; A)]$ for all $A \in \mathcal{A}(\theta)$ and yields the first part of the claim. The second part follows similarly by noting that, if $A^* \in \mathcal{A}(\theta)$ is such that $G(y; A^*)$ is a martingale and $G_0(y; A) = G_{0-}(y; A)$, then we have equalities instead of inequalities in the estimates leading to (4.38). By taking $T \to \infty$ we conclude that $\overline{S}_0 V(y, \theta) = \mathbb{E}[L_\infty(y; A^*)]$. Since $\overline{S}_0 V(y, \theta) \ge v(y, \theta)$ by the first part of the claim, we deduce the optimality of A^*. □

To justify later why the stochastic integrals in (4.8) are true martingales, we need the following technical

Lemma 4.5.2. *Let $\Theta_{0-} \ge 0$ be given and $F \in C^{2,1}(\mathbb{R} \times [0, \infty); \mathbb{R})$ be such that there exist constants $C_1, C_2 \ge 0$ with $|F(y, \theta)| \le C_1 \exp(C_2 y) \vee 1$ for all $(y, \theta) \in \mathbb{R} \times [0, \Theta_{0-}]$. For an admissible strategy $A \in \mathcal{A}(\Theta_{0-})$ let $Y^A =: Y$ denote the impact process defined by* (4.3) *for $y \in \mathbb{R}$. Then the stochastic integral processes*

$$\int_0^{\cdot} \overline{S}_u F(Y_u, \Theta_u)\,\mathrm{d}B_u \quad \textit{and} \quad \int_0^{\cdot} \overline{S}_u F(Y_u, \Theta_u)\,\mathrm{d}W_u \quad \textit{are true martingales.}$$

Proof. By the exponential growth of F, it suffices to check $\mathbb{E}[\int_0^t \overline{S}_u^2 \exp(2C_2 Y_u)\,\mathrm{d}u] < \infty$ for every $t \ge 0$. Consider an Ornstein-Uhlenbeck process X given by the dynamics $\mathrm{d}X_t = -\beta X_t\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t$, with $X_0 = y$. As in the proof of Proposition 4.5.1 (see (4.39)), we have $Y_t \leq X_t$ for all $t \geq 0$. In particular,

$$\mathbb{E}\Big[\int_0^t \overline{S}_u^2 \exp(2C_2 Y_u)\,\mathrm{d}u\Big] \leq \mathbb{E}\Big[\int_0^t \overline{S}_u^2 \exp(2C_2 X_u)\,\mathrm{d}u\Big]$$
$$= \int_0^t \mathbb{E}[\overline{S}_u^2 \exp(2C_2 X_u)]\,\mathrm{d}u \leq \int_0^t \sqrt{\mathbb{E}[\overline{S}_u^4]\mathbb{E}[\exp(4C_2 X_u)]}\,\mathrm{d}u < \infty,$$

using the Cauchy-Schwarz inequality and the fact that X is a Gaussian process. □

4.5.2 Verification and proof of Theorem 4.2.1

Now we verify that V is a classical solution of the variation inequality (4.9) with the boundary condition $V(y,0) = 0$ for all $y \in \mathbb{R}$. That $V(y,0) = 0$ is clear because $M_1(y_0) = 0$. The rest will be split into several lemmas.

Lemma 4.5.3 (Smooth pasting). *Let* $(y_b, \theta_b) \in \overline{\mathcal{W}} \cap \overline{\mathcal{S}}$. *Then*

$$\Phi(y_b)C'(\theta_b) + \Phi'(y_b)C(\theta_b) = f(y_b)\,, \tag{4.40}$$
$$\Phi'(y_b)C'(\theta_b) + \Phi''(y_b)C(\theta_b) = f'(y_b)\,. \tag{4.41}$$

Proof. This follows easily from $C(\theta_b) = M_1(y_b)$ and $C'(\theta_b) = M_2(y_b)$, see the definition of C and (4.29), together with the definitions of M_1 and M_2, see (4.28). Note that when $(y_b, \theta_b) = (y_0, 0)$ we take the right derivative of C at 0 and the equalities still hold true. □

Remark 4.5.4. It might be interesting to point out that (4.40) and (4.41) are sufficient to derive the boundary between the sell and the wait regions. Indeed, solving (4.40) – (4.41) with respect to $C(\theta_b)$ and $C'(\theta_b)$, it is easy to see that $C(\theta_b) = M_1(y_b)$ and $C'(\theta_b) = M_2(y_b)$. On the other hand, by the chain rule one gets $\theta'(y_b)C'(\theta_b) = M_1'(y_b)$ and thus we derive for the boundary parametrization $\theta(\cdot) = \mathrm{y}^{-1}(\cdot)$ in the appropriate range

$$\theta'(y_b) = \frac{M_1'}{M_2}(y_b),$$

which gives the ODE for the boundary in (4.29). To get the initial condition y_0, note that the boundary condition $V(\cdot,0) \equiv 0$ gives $C(0) = 0$, i.e. $M_1(y_0) = 0$, exactly as in Lemma 4.4.3. Thus, one could derive the candidate boundary function $\mathrm{y}(\cdot)$ after assuming sufficient smoothness of the function V along the boundary. This is similar to the classical approach in the singular stochastic control literature, cf. [KS86, Section 6]. The reason why we chose the seemingly longer derivation via calculus of variation techniques is the local (one-sided) optimality that we derived in Theorem 4.4.6 and that will be crucial in our verification of the inequalities of the candidate value function in the sell region, see Lemma 4.5.7. In the special case of $\lambda(\cdot)$ being constant, a more direct approach to verify the variational inequality is however available, see Section 4.5.3.

The smooth-pasting property translates to smoothness of V. Moreover, exponential bound on V and V_y will be needed to rely on the verification results from Section 4.5.1.

Lemma 4.5.5. *The function V is $C^{2,1}(\mathbb{R} \times [0,\infty))$. Moreover, for every Θ_{0-} there exist constants C_1, C_2, that depend on Θ_{0-}, such that both $V(y,\theta)$ and $V_y(y,\theta)$ are non-negative and bounded from above by $C_1 \exp(C_2 y) \vee 1$ for all $(y,\theta) \in \mathbb{R} \times [0, \Theta_{0-}]$.*

Proof. Inside $\mathcal{W}$, the function V is already $C^{2,1}$ by construction and the fact that $C(\theta) = M_1(\mathrm{y}(\theta))$ is continuously differentiable since $\mathrm{y}(\cdot)$ and $M_1(\cdot)$ are so.

For $(y,\theta) \in \mathcal{S}_1$, set $(y_b, \theta_b) := (y - \Delta(y,\theta), \theta - \Delta(y,\theta))$ and $\Delta := \Delta(y,\theta)$ (recall (4.34)). We have by (4.35) for the first and (4.40) for the second equality

$$\begin{aligned} V_y^{\mathcal{S}_1} &= \Phi'(y_b)C(\theta_b)\,(1-\Delta_y) + \Phi(y_b)C'(\theta_b)\,(-\Delta_y) + f(y) - f(y_b)\,(1-\Delta_y) \\ &= \Phi'(y-\Delta)C(\theta-\Delta) + f(y) - f(y-\Delta). \end{aligned} \tag{4.42}$$

Since f, Δ, C and Φ' are continuously differentiable, V_y will also be so. Hence by (4.41),

$$\begin{aligned} V_{yy}^{\mathcal{S}_1} &= \Phi''(y_b)C(\theta_b)\,(1-\Delta_y) + \Phi'(y_b)C'(\theta_b)\,(-\Delta_y) + f'(y) - f'(y_b)\,(1-\Delta_y) \\ &= V_{yy}^{\mathcal{W}}(y_b,\theta_b) + f'(y) - f'(y_b), \end{aligned} \tag{4.43}$$

which is continuous. On the other hand, by (4.35) and (4.41) we have

$$\begin{aligned} V_\theta^{\mathcal{S}_1}(y,\theta) &= \Phi'(y_b)C(\theta_b)(-\Delta_\theta) + \Phi(y_b)C'(\theta_b)(1-\Delta_\theta) - f(y_b)(-\Delta_\theta) \\ &= \Phi(y_b)C'(\theta_b), \end{aligned} \tag{4.44}$$

which is continuous. For $(y,\theta) \in \overline{\mathcal{W}} \cap \overline{\mathcal{S}}$ on the boundary, the left derivative w.r.t. y is

$$\lim_{x\searrow 0} \frac{1}{x}\big(V(y,\theta) - V(y-x,\theta)\big) = \Phi(y)C(\theta),$$

while the right derivative is again given by (4.42) and is equal to the left derivative since $\Delta(y,\theta) = 0$ in this case. Hence, V is continuously differentiable w.r.t. y on the boundary with derivative $V_y(y,\theta) = \Phi'(y)C(\theta)$. Similarly, the left derivative of V_y on the boundary is $\Phi''(y)C(\theta)$ and is equal to the right derivative which is given by (4.43) with $y = y_b$. The left derivative of V w.r.t. θ on the boundary is equal to the right derivative (given by (4.44)). Therefore, V is $C^{2,1}$ inside $\overline{\mathcal{W}} \cup \mathcal{S}_1$.

For $(y,\theta) \in \mathcal{S}_2$, we have that $V_y^{\mathcal{S}_2} = f(y) - f(y-\theta)$, $V_{yy}^{\mathcal{S}_2} = f'(y) - f'(y-\theta)$ and $V_\theta^{\mathcal{S}_2} = f(y-\theta)$ by (4.36), which are all continuous. On the boundary between $\mathcal{S}_1$ and $\mathcal{S}_2$, the left derivative of V w.r.t. y is given by (4.42) while the right derivative is $f(y) - f(y_0)$. Since $\theta - \Delta = 0$ in this case and $C(0) = 0$, they are equal and hence V is continuously differentiable w.r.t. y there; similarly for V_{yy}. The left derivative of V w.r.t. θ there is given by (4.44) with $(y_b,\theta_b) = (y_0, 0)$. The right derivative w.r.t. θ is $f(y-\theta) = f(y_0)$. They are equal by (4.41) and $C(0) = 0$. Therefore, V is $C^{2,1}$ on $\overline{\mathcal{S}}_1 \cup \mathcal{S}_2$. It remains to check smoothness on $\{(y,0) : y \in \mathbb{R}\}$. The derivatives w.r.t. y there are 0. V is continuously differentiable w.r.t. θ in this case because $\mathrm{y}(\cdot)$, C, and Δ are continuously differentiable w.r.t. θ also at $\theta = 0$ (we consider the right derivatives in this case).

To conclude the proof, the bound of V and V_y can be argued as follows. In the wait region, which is contained in $(-\infty, y_0] \times [0,\infty)$, we have $V(y,\theta) = C(\theta)\Phi(y)$ and $V_y(y,\theta) = C(\theta)\Phi'(y)$. Since Φ, Φ' are strictly increasing in y (see (4.6) and [Leb72, Chapter 10] for properties of the Hermite functions), V and V_y will be bounded by a constant there. Now, in the sell region we have $f - V_y - V_\theta = 0$. However, $V_\theta > 0$ because in $\mathcal{S}_1$ (4.44) holds and $C'(\theta_b) = M_2(\mathrm{y}(\theta_b)) > 0$, while in $\mathcal{S}_2$ we have that $V_\theta(y,\theta) = f(y-\theta) > 0$. Similarly, $V_y > 0$ in the sell region. Therefore, $0 < V_y(y,\theta) < f(y) \le \exp(\lambda_\infty y) \vee 1$ by Assumption C4. Hence, integrating in y gives $V(y,\theta) \le V(0,\theta) + \exp(\lambda_\infty y)/\lambda_\infty$ for $y \ge 0$, which implies $V(y,\theta) \le C_1 \exp(C_2 y) \vee 1$ for appropriate constants C_1, C_2. □

Next we prove that V solves the variational inequality (4.9).

Lemma 4.5.6. *The function* $V^{\mathcal{W}} : \overline{\mathcal{W}} \to [0,\infty)$ *from* (4.33) *satisfies*

$$\mathcal{L}V^{\mathcal{W}}(y,\theta) = 0 \quad \textit{and} \quad f(y) < V^{\mathcal{W}}_y(y,\theta) + V^{\mathcal{W}}_\theta(y,\theta) \textit{ for } y < \mathrm{y}(\theta).$$

Proof. From (4.27), we get representations $V^{\mathcal{W}}_\theta = \Phi(y)M_1'(\mathrm{y}(\theta))\mathrm{y}'(\theta) = \Phi(y)M_2(\mathrm{y}(\theta))$ and $V^{\mathcal{W}}_y = \Phi'(y)M_1(\mathrm{y}(\theta))$. Recall that at $y = \mathrm{y}(\theta)$ we have by (4.40) the equality $V^{\mathcal{W}}_y + V^{\mathcal{W}}_\theta = f(\mathrm{y}(\theta))$. Now consider $y < \mathrm{y}(\theta)$. By Lemma 4.4.2, we then have $M_1(y) > M_1(\mathrm{y}(\theta))$ giving

$$\Big(\frac{f}{\Phi}\Big)'(y) > \Big(\frac{\Phi'}{\Phi}\Big)'(y)M_1\big(\mathrm{y}(\theta)\big) = \frac{\mathrm{d}}{\mathrm{d}y}\bigg(M_1\big(\mathrm{y}(\theta)\big)\frac{\Phi'(y)}{\Phi(y)} + M_2\big(\mathrm{y}(\theta)\big)\bigg).$$

Therefore, $y \mapsto (f - V^{\mathcal{W}}_y(y,\theta) + V^{\mathcal{W}}_\theta(y,\theta))/\Phi(y)$ is increasing in y. Since at $y = \mathrm{y}(\theta)$ it equals to 0, we get the claimed inequality. □

It remains to verify the inequality in the sell region. The proof is more subtle and that is where Theorem 4.4.6 plays a crucial role. Recall Assumption 4.1.2 and note that y_∞ from Lemma 4.4.3 is unique by condition C3.

Lemma 4.5.7. *The functions* $V^{\mathcal{S}_1}$ *and* $V^{\mathcal{S}_2}$ *satisfy on* $\overline{\mathcal{S}}_1$ *and* $\mathcal{S}_2$ *respectively*

$$\mathcal{L}V^{\mathcal{S}_1} \le 0, \quad \mathcal{L}V^{\mathcal{S}_2} < 0.$$

Moreover, the inequality inside $\overline{\mathcal{S}}_1$ *is strict except on the boundary between the wait region and the sell region* ($\overline{\mathcal{W}} \cap \overline{\mathcal{S}}_1$) *where we have equality.*

Proof. First consider region $\overline{\mathcal{S}}_1$. Recall from Lemma 4.5.5 (see (4.42) – (4.43)) that in this case

$$\begin{aligned}
V^{\mathcal{S}_1}_y(y,\theta) &= V^{\mathcal{W}}_y(y-\Delta,\theta-\Delta) + f(y) - f(y-\Delta),\\
V^{\mathcal{S}_1}_{yy}(y,\theta) &= V^{\mathcal{W}}_{yy}(y_b,\theta_b) + f'(y) - f'(y_b),
\end{aligned}$$

where $y = y_b + \Delta(y,\theta)$ and $\theta = \theta_b + \Delta(y,\theta)$. Fix $(y_b,\theta_b) \in \overline{\mathcal{W}} \cap \overline{\mathcal{S}}_1$ and consider the perturbation $\Delta \mapsto (y,\theta) = (y_b+\Delta, \theta_b+\Delta)$. Set

$$\begin{aligned}
h(\Delta) &:= \mathcal{L}V^{\mathcal{S}_1}(y_b+\Delta,\theta_b+\Delta)\\
&= \tfrac{\hat\sigma^2}{2}V^{\mathcal{W}}_{yy}(y_b,\theta_b) - \tfrac{\hat\sigma^2}{2}f'(y_b) + \sigma\rho\hat\sigma V^{\mathcal{W}}_y(y_b,\theta_b) - \sigma\rho\hat\sigma f(y_b) - \delta V^{\mathcal{W}}(y_b,\theta_b)\\
&\quad + \tfrac{\hat\sigma^2}{2}f'(y) - \beta y V^{\mathcal{W}}_y(y_b,\theta_b) + \beta y f(y_b) + (\sigma\rho\hat\sigma - \beta y)f(y) - \delta\int_{y_b}^{y} f(x)\,\mathrm{d}x\,.
\end{aligned}$$

Note that $h(0) = 0$ by Lemma 4.5.6 and to show $h(\Delta) < 0$ for $\Delta > 0$, it suffices to prove $h'(\Delta) < 0$ for all $\Delta > 0$. We have for all $\Delta \ge 0$ at $y = y_b + \Delta$ that

$$h'(\Delta) = \beta\big(f(y_b) - V^{\mathcal{W}}_y(y_b,\theta_b)\big) + f(y)\underbrace{\bigg(\frac{\hat\sigma^2}{2}\frac{f''(y)}{f(y)} - (\beta+\delta) + (\sigma\rho\hat\sigma - \beta y)\frac{f'(y)}{f(y)}\bigg)}_{=k(y)}, \quad (4.45)$$

where at $\Delta = 0$ we consider the right derivative $h'(0+)$. Now we show that $k(y) < 0$ for all $y \geq y_\infty$. To this end, recall that Φ solves the ODE $\delta\Phi(x) = \frac{\hat{\sigma}^2}{2}\Phi''(x) + (\sigma\rho\hat{\sigma} - \beta x)\Phi'(x)$. Differentiating w.r.t. x and dividing by $\Phi'(x)$ yields

$$0 = \frac{\hat{\sigma}^2}{2}\left(\frac{\Phi''(x)}{\Phi'(x)}\right)' + \frac{\hat{\sigma}^2}{2}\frac{\Phi''(x)^2}{\Phi'(x)^2} - (\beta + \delta) + (\sigma\rho\hat{\sigma} - \beta x)\frac{\Phi''(x)}{\Phi'(x)}$$

So at the left end y_∞ of our boundary, we have

$$\begin{aligned} k(y_\infty) &= \frac{\hat{\sigma}^2}{2}\left(\frac{f'}{f}\right)'(y_\infty) + \frac{\hat{\sigma}^2}{2}\frac{\Phi''(y_\infty)^2}{\Phi'(y_\infty)^2} - (\beta + \delta) + (\sigma\rho\hat{\sigma} - \beta y_\infty)\frac{\Phi''(y_\infty)}{\Phi'(y_\infty)} \\ &= \frac{\hat{\sigma}^2}{2}\left(\frac{f'}{f}\right)'(y_\infty) - \frac{\hat{\sigma}^2}{2}\left(\frac{\Phi''}{\Phi'}\right)'(y_\infty) < 0 \end{aligned} \tag{4.46}$$

by Assumption C3. With Assumption C5 we get $k(y) < 0$ for every $y \geq y_\infty$.

In particular, $k(y_b+\Delta) < 0$ for all $\Delta \geq 0$. Since f is positive and increasing, the product $\Delta \mapsto (fk)(y_b + \Delta)$ is decreasing. Therefore, proving $h'(0+) \leq 0$ is sufficient to show the inequality in $\mathcal{S}_1$. To stress the dependence of h on the point $(y_b, \theta_b) = (\mathrm{y}(\theta_b), \theta_b)$, we also write $h(\Delta) = h_{\theta_b}(\Delta)$. Note that $h_\theta(\Delta)$ is continuous in θ and Δ on $[0,\infty) \times [0,\infty)$.

Assume $h'_{\theta_b}(0+) > 0$ at some boundary point (y_b, θ_b) with $\theta_b > 0$. By continuity of h' on θ and Δ there exists some $\varepsilon > 0$ such that $\mathcal{L}V^{\mathcal{S}_1} > 0$ on $U := \overline{\mathcal{S}}_1 \cap B_\varepsilon(y_b, \theta_b)$. This will lead to a contradiction to the fact that the candidate boundary is a (one-sided) strict local maximizer of our stochastic optimization problem with strategies described by the local times of reflected diffusions, see Theorem 4.4.6.

Indeed, fix $\Theta_0 > \theta_b + \varepsilon$ and consider a perturbation $\tilde{\mathrm{y}}(\cdot) \in C^1$ of the boundary $\mathrm{y}(\cdot)$ which satisfies the conditions of Theorem 4.4.6 and $\mathrm{y}(\theta) < \tilde{\mathrm{y}}(\theta) \leq y_0$ in $(\tilde{\mathrm{y}}(\theta), \theta) \in U$ and such that $\tilde{\mathrm{y}}$ and y coincide outside of U. For the corresponding reflection strategies $\tilde{A} := A^{\mathrm{refl}}(\tilde{\mathrm{y}}, \Theta_0)$ and $A := A^{\mathrm{refl}}(\mathrm{y}, \Theta_0)$ denote by $\tilde{\Theta}_t := \Theta_0 - \tilde{A}_t$ and $\Theta_t := \Theta_0 - A_t$ their asset position processes. The liquidation times of $\tilde{A}$ and A are $\tilde{\tau} := \inf\{t \geq 0 : \tilde{A}_t = \Theta_0\}$ and $\tau := \inf\{t \geq 0 : A_t = \Theta_0\}$, respectively. By Theorem 4.3.2 (see also the discussion after (4.11)), we have $T := \tilde{\tau} \vee \tau < \infty$ a.s. Fix initial impact $Y_{0-}^{\tilde{A}} = Y_{0-}^{A} = \mathrm{y}(\Theta_0)$. To compare the strategies A and $\tilde{A}$, consider the processes $G(\mathrm{y}(\Theta_0); A)$ and $G(\mathrm{y}(\Theta_0); \tilde{A})$ from (4.7) for our candidate value function (which is $C^{2,1}$ by Lemma 4.5.5). Since $V(\cdot, 0) = 0$, we have $L_T(\tilde{A}) = G_T(\tilde{A})$ and $L_T(A) = G_T(A)$. However, since $(Y^{\tilde{A}}, \tilde{\Theta})$ spends a positive amount of time in the region $\{\mathcal{L}V > 0\}$ until time T and always remains in the region $\{\mathcal{L}V \geq 0\}$, the perturbed strategy $\tilde{A}$ generates larger proceeds (in expectation) than A.

Indeed, by (4.8) applied for $G(\tilde{A})$ and $G(A)$, using monotone convergence (twice) and arguments as in the proof of Proposition 4.5.1 for the first equality (by (4.19) expected proceeds are bounded), and Lemma 4.5.2 for the stochastic integrals in the second line (noting the growth condition from Lemma 4.5.5), we get

$$\begin{aligned} \mathbb{E}[L_\infty(\tilde{A}) - L_\infty(A)] &= \lim_{n\to\infty} \mathbb{E}[G_{n\wedge T}(\tilde{A}) - G_{n\wedge T}(A)] \\ &= \lim_{n\to\infty} \mathbb{E}\Big[\int_0^{n\wedge T} \dots \,\mathrm{d}W_t + \int_0^{n\wedge T} \dots \,\mathrm{d}B_t + \int_0^{n\wedge T} \mathcal{L}V(Y_t^{\tilde{A}}, \tilde{\Theta}_t)\,\mathrm{d}t\Big] \\ &= \mathbb{E}\Big[\int_0^T \mathcal{L}V(Y_t^{\tilde{A}}, \tilde{\Theta}_t)\,\mathrm{d}t\Big] > 0\,. \end{aligned}$$

This contradicts Theorem 4.4.6, so $h'(0+) \le 0$ and hence the inequality in $\mathcal{S}_1$ must hold.

It remains to consider the case $(y,\theta) \in \overline{\mathcal{S}}_2$, where $V_y^{\mathcal{S}_2} = f(y) - f(y-\theta)$ and $V_{yy}^{\mathcal{S}_2} = f'(y) - f'(y-\theta)$. Fix $y - \theta =: a \ge y_0$ and consider $\mathcal{L}V^{\mathcal{S}_2}$ as a function of θ. We have

$$\begin{aligned}\mathcal{L}V^{\mathcal{S}_2}(y,\theta) = {} & \frac{\hat{\sigma}^2}{2}\big(f'(a+\theta) - f'(a)\big) + \big(\sigma\rho\hat{\sigma} - \beta(a+\theta)\big)\big(f(a+\theta) - f(a)\big) \\ & - \delta \int_a^{a+\theta} f(x)\,\mathrm{d}x.\end{aligned}$$

Differentiating the right-hand side w.r.t. θ we get $f(a+\theta)k(a+\theta)$, which is again decreasing in θ because $a \ge y_0$. Since at $\theta = 0$ we have $\mathcal{L}V^{\mathcal{S}_2}(y,\theta) = 0$ we deduce the desired inequality. □

Note that in the particular case of constant $\lambda = f'/f$, a more direct approach is available, see Section 4.5.3 below. Now we have all the ingredients in place to complete the

Proof of Theorem 4.2.1. The function V constructed in (4.37) is a classical solution of the variational inequality (4.9) because of Lemmas 4.5.5, 4.5.6 and 4.5.7. Thus, for each admissible strategy A the process $G(y;A)$ from (4.7) is a supermartingale with $G_0(y;A) \le G_{0-}(y;A)$: the growth condition on V_y and V from Lemma 4.5.5 guarantees that the stochastic integral processes in (4.8) are true martingales by an application of Lemma 4.5.2, while the variational inequality gives the supermartingale property on $[0-,\infty)$. Moreover, for the described strategy A^*, whose existence and uniqueness on $[\![0,\tau]\!]$ follows from classical results, cf. Remark 4.3.1, the process $G(y;A^*)$ is a true martingale with $G_0(y;A^*) = G_{0-}(y;A^*)$ by our construction of V and the validity of the variational inequality in the respective regions. Therefore A^* is an optimal strategy by Proposition 4.5.1. Any other strategy will be suboptimal because the respective inequalities are strict in the sell and wait region, i.e., for any other strategy the process G will be a strict supermartingale.

The Laplace transform formula (4.11) was derived in Theorem 4.3.2 for a $\mathbb{y}$-reflected strategy when the state process starts on the boundary. If the state process starts in $Y_0 = x$ in the wait region, the behavior of the process until time $H^{x\to z}$ when it hits the boundary for the first time (at $z := \mathbb{y}(\Theta_0)$) is independent from future excursions from the boundary, and hence the multiplicative factor in (4.11), see e.g. [RW87, Prop. V.50.3]: for $x < z \in \mathbb{R}$ and $\alpha > 0$, $\mathbb{E}[\exp(-\alpha H^{x\to z})] = \Phi_\alpha(x)/\Phi_\alpha(z)$. □

4.5.3 Alternative verification for exponential impact function

The most difficult part in the proof of Lemma 4.5.7 is showing $h'(0+) \le 0$ for given (y_b,θ_b) on the boundary, i.e. $y_b \in (y_\infty, y_0]$, $\theta_b = \theta\!\!\!/(y_b)$, with $h'(\Delta)$ from (4.45). Written in terms of y_b only, we need to show $g(y_b) \le 0$ for $y_b \in (y_\infty, y_0]$ where

$$g(y) := \beta\left(f(y) - \Phi'\frac{f\Phi' - f'\Phi}{(\Phi')^2 - \Phi\Phi''}\right) + f(y)k(y),$$

using $V_y^{\mathcal{W}}(y_b,\theta_b) = \Phi'(y_b)M_1(y_b)$ and $M_1 = (f\Phi' - f'\Phi)/((\Phi')^2 - \Phi\Phi'')$ by (4.28). Direct calculations give $g(y_0) = \beta f(y_0) + f(y_0)k(y_0) = \frac{\hat{\sigma}^2}{2}\big((f'/f)' - (\Phi'/\Phi)'\big)(y_0) < 0$ and, using (4.46), also $g(y_\infty) < 0$. This implies $g(y_\infty)/f(y_\infty) < 0$ and $g(y_0)/f(y_0) < 0$, so it would suffice to prove monotonicity of g/f on $[y_\infty, y_0]$ to deduce $g < 0$ on $[y_\infty, y_0]$.

Now consider constant $\lambda = f'/f$. In this case $k'(y) = -\beta\lambda$ and we get

$$\left(\frac{g}{f}\right)' = \frac{\beta}{(\Phi\Phi'' - (\Phi')^2)^2} \frac{\Phi^3(\Phi'')^2}{\Phi'} \left(\frac{(\Phi')^2}{\Phi\Phi'}\right)' \left(\frac{\Phi'}{\Phi} - \lambda\right).$$

Since $\Phi'(y)/\Phi(y) \le \lambda$ for $y \le y_0$, we find $(g/f)' \le 0$ on $(-\infty, y_0]$ thanks to the following result.

Lemma 4.5.8. *The function $y \mapsto \frac{\Phi'(y)^2}{\Phi(y)\Phi''(y)}$ is increasing.*

The proof is due to Torben Koch [Koc19].

Proof. Fix arbitrary numbers $x < y$ and denote by $\tau := \inf\{t \ge 0 \mid X_t = y\}$ the first hitting time of level y by an Ornstein-Uhlenbeck process X with $\mathrm{d}X_t = (\sigma\rho\hat\sigma - \beta X_t)\,\mathrm{d}t + \hat\sigma\,\mathrm{d}B_t$, starting in $X_0 = x$. For $\kappa > 0$, we have the Laplace transform $\mathbb{E}[e^{-\kappa\tau}] = \Phi_\kappa(x)/\Phi_\kappa(y)$, cf. [RW87, V.50] or (7.8), which means $\mathbb{E}[e^{-(\delta+n\beta)\tau}] = \Phi^{(n)}(x)/\Phi^{(n)}(y)$ for integer $n \ge 0$. Now, Hölder's inequality yields

$$\frac{\Phi'(x)}{\Phi'(y)} = \mathbb{E}\big[e^{-\frac{\delta}{2}\tau} e^{-\frac{\delta+2\beta}{2}\tau}\big] \le \sqrt{\mathbb{E}\big[e^{-\delta\tau}\big]}\sqrt{\mathbb{E}\big[e^{-(\delta+2\beta)\tau}\big]} = \sqrt{\frac{\Phi(x)}{\Phi(y)}}\sqrt{\frac{\Phi''(x)}{\Phi''(y)}},$$

which implies $\frac{\Phi'(x)^2}{\Phi(x)\Phi''(x)} \le \frac{\Phi'(y)^2}{\Phi(y)\Phi''(y)}$ by positivity of Φ, Φ', Φ''. □

4.6 Relation to optimal stopping

Now that we have solved the optimal liquidation problem (4.5), we can relate it to an optimal stopping problem in a similar way as it is done in the recent [FK19, Sect. 4.1] for a model that closely resembles our setup, but with linear $f(y) = y - c$ (in our notation). For the relation between singular control and optimal stopping problems, cf. e.g. [KS84]. One theme in the literature is to utilize this relation to solve singular control problems by solving the associated optimal stopping problems, see e.g. [DAFF17]. Such approach usually hinges on convexity properties of the objective functional.

Throughout this section, denote by X^y the Ornstein-Uhlenbeck process with dynamics $\mathrm{d}X^y_t = (\sigma\rho\hat\sigma - \beta X^y_t)\,\mathrm{d}t + \hat\sigma\,\mathrm{d}B_t$, with $X^y_0 = y$. Recall that the infinitesimal generator of X^y is $\phi \mapsto \mathcal{L}\phi + \delta\phi$ for the operator $\mathcal{L}\phi(y) = \frac{\hat\sigma^2}{2}\phi''(y) + (\sigma\rho\hat\sigma - \beta y)\phi'(y) - \delta\phi(y)$. We also write $\mathcal{L}\phi(y,\theta) := \mathcal{L}(\phi(\cdot,\theta))(y)$ for functions ϕ of two variables.

Corollary 4.6.1. *The function $u(y,\theta) := V_y(y,\theta) + V_\theta(y,\theta)$ for $y \in \mathbb{R}$, $\theta \in [0,\infty)$ is the value function of an optimal stopping problem,*

$$u(y,\theta) = \sup_\tau \mathbb{E}\left[e^{-\delta\tau} f(X^y_\tau) - \int_0^\tau e^{-\delta t}\beta C(\theta)\Phi'(X^y_t)\,\mathrm{d}t\right] \tag{4.47}$$

where the supremum is taken over all $(\mathcal{F}_t)_{t\ge0}$-stopping times τ and $C(\theta) = M_1(\mathrm{y}(\theta))$ as in (4.33). Moreover, the optimal stopping time for (4.47) is given by

$$\tau^*(y,\theta) = \inf\{t \ge 0 \mid X^y_t \ge \mathrm{y}(\theta)\}.$$

Proof. Fix $\theta \geq 0$ and note that $u(\cdot,\theta) \in C^1(\mathbb{R})$ by Lemma 4.5.5. Moreover, we find $u_{yy} \in L^\infty_{\mathrm{loc}}$ since $u_y(\cdot,\theta) \in C(\mathbb{R}) \cap C^1(\mathbb{R}\setminus\{\mathrm{y}(\theta)\})$, that is, $u(\cdot,\theta) \in W^{2,\infty}_{\mathrm{loc}}$. With a version of the Itô formula for such u, see e.g. [Pro05, Ch. IV, Thm. 71], standard arguments (cf. [PS06, Ch. IV]) give that it suffices to prove that $w = u(\cdot,\theta)$ solves

$$\begin{aligned} \mathcal{L}w(y) - \beta C(\theta)\Phi'(y) = 0 \quad &\text{and} \quad f(y) - w(y) < 0 \quad \text{for } y < \mathrm{y}(\theta),\\ \mathcal{L}w(y) - \beta C(\theta)\Phi'(y) < 0 \quad &\text{and} \quad f(y) - w(y) = 0 \quad \text{for } y > \mathrm{y}(\theta). \end{aligned}$$

By construction of V in $\mathcal{S}$ and Lemma 4.5.6, we immediately get $f(y) = u(y,\theta)$ for $y > \mathrm{y}(\theta)$ and $f(y) < u(y,\theta)$ for $y < \mathrm{y}(\theta)$. Inside $\mathcal{W}$, we have $V(y,\theta) = C(\theta)\Phi(y)$, so that $\mathcal{L}V_y = \beta V_y$ and $\mathcal{L}V_\theta = 0$. Hence, $\mathcal{L}u(y,\theta) = \beta C(\theta)\Phi'(y)$ for all $y < \mathrm{y}(\theta)$. Now for $y > \mathrm{y}(\theta)$ let

$$g(y) := \mathcal{L}f(y) - \beta C(\theta)\Phi'(y) = (fk)(y) + \beta\big(f(y) - C(\theta)\Phi'(y)\big).$$

Recall the auxiliary function $h(\Delta)$ from the proof of Lemma 4.5.7, where we showed $h'(0+) \leq 0$. By (4.45) we have $g(\mathrm{y}(\theta)) = h'(0+) \leq 0$. Since Φ'/f is increasing on (y_∞, ∞) by Assumptions C3 and C6 and k is decreasing by Assumption C5, we get that g/f is decreasing for $y > \mathrm{y}(\theta)$ and therefore $g(y) < 0$. □

Remark 4.6.2. Using optional projection, we can rewrite the objective in (4.47) in terms of the fundamental price $\overline{S}$ and impact Y up to localization: For fixed $T \in [0,\infty)$ we have by [DM82, Thm. VI.57] that

$$\begin{aligned} &\mathbb{E}\Big[e^{-\gamma(\tau\wedge T)}\overline{S}_{\tau\wedge T} f(Y_{\tau\wedge T}) - \int_0^{\tau\wedge T} e^{-\gamma t}\overline{S}_t \beta C(\theta)\Phi(Y_t)\,\mathrm{d}t\Big]\\ &\qquad = \overline{S}_0\mathbb{E}\Big[e^{-\delta(\tau\wedge T)} f(X_{\tau\wedge T}) - \int_0^{\tau\wedge T} e^{-\delta t}\beta C(\theta)\Phi(X_t)\,\mathrm{d}t\Big], \end{aligned}$$

for the *uncontrolled* process $Y = Y^0$ with $\mathrm{d}Y_t = -\beta Y_t\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t$ and $Y_0 = y = X_0$.

Noting that $C(0) = 0$, we again see the small investor's stopping problem from Remark 4.1.3. For $\theta > 0$ we have $C(\theta) > 0$, so the integral term in (4.47) can be understood as an additional penalty for waiting compared to the (marginal) gains a small investor would receive.

4.7 Sensitivity analysis for the impact stochasticity

We will discuss the sensitivity of the free boundary y on the noise parameter $\hat{\sigma}$ of the impact process Y. To stress the dependence on $\hat{\sigma}$, denote $y_0^{\hat{\sigma}} := y_0$, $y_\infty^{\hat{\sigma}} := y_\infty$ and $\mathrm{y}^{\hat{\sigma}} := \mathrm{y}$, so that $\mathrm{y}^{\hat{\sigma}}$ solves the ODE $(\mathrm{y}^{\hat{\sigma}})'(\theta) = G^{\hat{\sigma}}(\mathrm{y}^{\hat{\sigma}}(\theta))$ with $\mathrm{y}^{\hat{\sigma}}(0) = y_0^{\hat{\sigma}}$ and $G^{\hat{\sigma}}$ given by the right-hand side of (4.10). By equation (4.27) and Lemma 4.4.2 we have $G^{\hat{\sigma}} \in C^1(\mathbb{R})$. We also have $y_\infty^{\hat{\sigma}} < \mathrm{y}^{\hat{\sigma}}(\theta) \leq y_0^{\hat{\sigma}}$ with decreasing $\mathrm{y}^{\hat{\sigma}}$ for all $\hat{\sigma} > 0$. We will now investigate the limiting behavior for $\hat{\sigma} \searrow 0$.

Formally setting $\hat{\sigma} = 0$ in (4.3) yields the deterministic impact model (2.2) from Chapter 2 with linear resilience $h(y) = \beta y$. Note that Assumption 4.1.2 implies Assumption 2.2.2 in this case. According to Theorem 2.2.4, the optimal monotone strategy for deterministic impact dynamics is characterized by a free boundary function $\mathbb{y}^0$ that solves the ODE $(\mathbb{y}^0)'(\theta) = G^0(\mathbb{y}^0(\theta))$ with $\mathbb{y}^0(0) = y_0^0$ where $\lambda(y_0^0) = -\delta/(\beta y_0^0)$ and

$$G^0(y) := \frac{\delta\big(\beta y\lambda(y) + \beta + \delta\big)}{\big(\beta y\lambda(y) + \delta\big)\big(\beta y\lambda(y) + \beta + \delta\big) + \beta y\big(\beta\lambda(y) + \beta y\lambda'(y)\big)}. \tag{4.48}$$

Denote by y_∞^0 the unique solution to $\lambda(y_\infty^0) = -(\beta + \delta)/(\beta y_\infty^0)$. Then we have $y_\infty^0 < \mathbb{y}^0(\theta) \le y_0^0 < 0$ for all $\theta \ge 0$ and $\mathbb{y}^0$ is decreasing.

We will first show that $y_0^{\hat\sigma} \to y_0^0$, $y_\infty^{\hat\sigma} \to y_\infty^0$ and $G^{\hat\sigma}(y) \to G^0(y)$ pointwise for all $y < 0$, as $\hat\sigma \searrow 0$. Note that $G^{\hat\sigma}$ may be rewritten as

$$G^{\hat\sigma}(y) = \Big(A_0 \frac{(A_0 - A_1)(\lambda - A_1)}{(\lambda' + \lambda^2)(A_1 - A_0) + \lambda\cdot(A_0A_1 - A_1A_2) + (A_0A_1A_2 - A_0A_1^2)}\Big)(y)\,, \tag{4.49}$$

with $A_n := \Phi^{(n+1)}/\Phi^{(n)}$. To obtain the limit of $A_n(y)$ as $\hat\sigma \searrow 0$, for $n \ge 0$, we utilize [Leb72, eq. (10.6.6)], which yields in particular for real $z > 0$ and arbitrary ν the asymptotic behavior

$$H_\nu(z) = (2z)^\nu\big(1 + \mathcal{O}(|z|^{-2})\big). \tag{4.50}$$

Moreover, we know the derivatives of Hermite functions by [Leb72, eq. (10.4.4)] as

$$H_\nu'(z) = 2\nu H_{\nu-1}(z). \tag{4.51}$$

Lemma 4.7.1. *Let $y < 0$ and $n \in \mathbb{N}_0$. Then we have $A_n(y) \to -\frac{\delta+n\beta}{\beta y}$ for $\hat\sigma \searrow 0$.*

Proof. By (4.6) and (4.51) we have the n^{th} derivative

$$\Phi^{(n)}(y) = 2^n \frac{\prod_{k=0}^{n-1}(\delta + k\beta)}{\beta^{n/2}\hat\sigma^n} H_{-(\delta+n\beta)/\beta}(z).$$

with $z := (\sigma\rho\hat\sigma - \beta y)/(\sqrt{\beta}\hat\sigma)$. Since $z \to +\infty$ as $\hat\sigma \searrow 0$ for $y < 0$, we find by (4.50) the limit

$$\begin{aligned} A_n(y) &= 2\frac{\delta + n\beta}{\sqrt{\beta}\hat\sigma} \cdot \frac{H_{-(\delta+(n+1)\beta)/\beta}(z)}{H_{-(\delta+n\beta)/\beta}(z)} \\ &= 2\frac{\delta + n\beta}{\sqrt{\beta}\hat\sigma} \cdot \frac{(2z)^{-(\delta+(n+1)\beta)/\beta}}{(2z)^{-(\delta+n\beta)/\beta}} \cdot \frac{1 + \mathcal{O}(|z|^{-2})}{1 + \mathcal{O}(|z|^{-2})} \\ &= \frac{\delta + n\beta}{\sqrt{\beta}\hat\sigma} \cdot \frac{\sqrt{\beta}\hat\sigma}{\sigma\rho\hat\sigma - \beta y} \cdot \frac{1 + \mathcal{O}(|z|^{-2})}{1 + \mathcal{O}(|z|^{-2})} \to -\frac{\delta + n\beta}{\beta y} \end{aligned}$$

for all $y < 0$, as $\hat\sigma \searrow 0$. □

Now, since $y := \mathbb{y}^{\hat\sigma}(\theta) \le y_0^{\hat\sigma}$, for $\theta, \hat\sigma \ge 0$, and $y_0^0 < 0$, we can assume $y < 0$ (and therefore apply Lemma 4.7.1) for all $\hat\sigma \ge 0$ small enough by the following

Lemma 4.7.2. *We have $y_0^{\hat\sigma} \to y_0^0$ and $y_\infty^{\hat\sigma} \to y_\infty^0$ as $\hat\sigma \searrow 0$.*

Proof. First, we will show that necessarily $y_0^{\hat\sigma} < 0$ for all $\hat\sigma > 0$ small enough. By (4.51),

$$\frac{\Phi'(y)}{\Phi(y)} = 2\frac{\delta}{\sqrt{\beta}\hat\sigma}\frac{H_{-\delta/\beta-1}(z)}{H_{-\delta/\beta}(z)} \qquad \text{for all } y \in \mathbb{R}, \text{ with } z := \tfrac{\sigma\beta\hat\sigma-\beta y}{\sqrt{\beta}\hat\sigma}.$$

If $y = 0$, then $z \equiv \sigma\sqrt{\beta}$ and therefore $\Phi'(0)/\Phi(0) \to +\infty$ as $\hat\sigma \searrow 0$. Since we know by Lemma 4.4.1 that Φ'/Φ is strictly increasing for all $\hat\sigma > 0$, it follows $\Phi'(y)/\Phi(y) \to \infty$ as $\hat\sigma \searrow 0$ for all $y \geq 0$. Hence the solution $y = y_0^{\hat\sigma}$ of $\lambda(y) = \Phi'(y)/\Phi(y)$ needs to be negative for all $\hat\sigma > 0$ small enough.

Now we can apply Lemma 4.7.1 to find $A_0 \to -\delta/(\beta y)$. Therefore, the function g with $g(\hat\sigma, y) := \lambda(y) - (\Phi^{\hat\sigma})'(y)/\Phi^{\hat\sigma}(y)$, if $\hat\sigma > 0$, and $g(x,y) := \lambda(y) + \delta/(\beta y)$, if $x \leq 0$, is continuous on $(-\varepsilon, \varepsilon) \times (-\infty, -\varepsilon)$ for $\varepsilon > 0$ small enough. Since $y_0^{\hat\sigma}$ is unique for all $\hat\sigma \geq 0$, it follows by [Kum80, Cor. 1.1] that the implicit function $\hat\sigma \mapsto y_0^{\hat\sigma}$ exists and is continuous in a neighborhood of zero.

For $y_\infty^{\hat\sigma}$, the proof is analogous, using $\lambda(y_\infty^{\hat\sigma}) = A_1(y_\infty^{\hat\sigma})$ and $A_1(y) \to -(\delta+\beta)/(\beta y)$. □

As a further step before showing convergence of $\mathrm{y}^{\hat\sigma}$ to y^0, we investigate $G^{\hat\sigma}$.

Lemma 4.7.3. *There exists $\varepsilon > 0$ such that the function $(\hat\sigma, y) \mapsto G^{\hat\sigma}(y)$ is $C^{0,1}$-continuous on $[0,\varepsilon] \times [y_\infty^0 - \varepsilon, y_0^0 + \varepsilon]$ with uniform bound on the derivatives. In particular, the family $(G^{\hat\sigma})_{\hat\sigma \geq 0}$ is equicontinuous on $[y_\infty^0 - \varepsilon, y_0^0 + \varepsilon]$.*

Proof. For continuity it suffices to show continuity as $\hat\sigma \searrow 0$. The denominator in (4.10) is strictly negative on $[y_\infty^0, y_0^0]$, so $G^0(y)$ is continuous on a neighborhood $[y_\infty^0 - \varepsilon, y_0^0 + \varepsilon] \subset (-\infty, 0)$. Now, direct calculations using (4.49) and Lemma 4.7.1 show $G^{\hat\sigma}(y) \to G^0(y)$ as $\hat\sigma \searrow 0$ for every $y \in [y_\infty^0 - \varepsilon, y_0^0 + \varepsilon]$:

$$\begin{aligned}
G^{\hat\sigma}(y) &= \Big(\frac{A_0(A_0 - A_1)(\lambda - A_1)}{(\lambda' + \lambda^2)(A_1 - A_0) + \lambda \cdot (A_0A_1 - A_1A_2) + A_0A_1A_2 - A_0A_1^2}\Big)(y) \\
&\to \frac{\frac{-\delta}{\beta y}\big(\frac{-\delta}{\beta y} + \frac{\delta+\beta}{\beta y}\big)\big(\lambda(y) + \frac{\delta+\beta}{\beta y}\big)}{(\lambda' + \lambda^2)(y)\big(-\frac{\delta+\beta}{\beta y} + \frac{\delta}{\beta y}\big) + \lambda(y)\big(\frac{\delta(\delta+\beta)}{\beta^2 y^2} - \frac{(\delta+\beta)(\delta+2\beta)}{\beta^2 y^2}\big) - \frac{\delta(\delta+\beta)(\delta+2\beta)}{\beta^3 y^3} + \frac{\delta(\delta+\beta)^2}{\beta^3 y^3}} \\
&= \frac{-\delta\beta\big(\beta y\lambda(y) + \beta + \delta\big)}{-\beta^3 y^2(\lambda' + \lambda^2)(y) - 2\beta^2 y(\beta+\delta)\lambda(y) - \beta\delta(\beta+\delta)} \\
&= G^0(y)
\end{aligned}$$

by (4.48).

For equicontinuity, note that $G^{\hat\sigma}$ is continuously differentiable on $[y_\infty^0 - \varepsilon, y_0^0 + \varepsilon]$ for all $\hat\sigma \geq 0$. As shown above, the denominator of (4.49) converges to a non-zero value for $y \in [y_\infty^0 - \varepsilon, y_0^0 + \varepsilon]$. Hence the derivative of (4.49), which is a function of λ and A_n, converges for every $y \in [y_\infty^0 - \varepsilon, y_0^0 + \varepsilon]$ to a finite value as $\hat\sigma \searrow 0$ that, moreover, is continuous in y: Let $\psi(\hat\sigma, y) := \frac{\mathrm{d}}{\mathrm{d}y}G^{\hat\sigma}(y)$, for $\hat\sigma > 0$, and $\psi(0,y) := \lim_{\hat\sigma\searrow 0}\psi(\hat\sigma, y)$. So ψ is continuous on $K := [0,\varepsilon] \times [y_\infty^0 - \varepsilon, y_0^0 + \varepsilon]$. Now we have a common upper bound $\sup_{(\hat\sigma,y)\in K}|\psi(\hat\sigma,y)| \wedge \sup_{y\in[y_\infty^0-\varepsilon, y_0^0+\varepsilon]}\big|\frac{\mathrm{d}}{\mathrm{d}y}G^0(y)\big| < \infty$ for $\big|\frac{\mathrm{d}}{\mathrm{d}y}G^{\hat\sigma}(y)\big|$, $\hat\sigma \geq 0$, which implies equicontinuity of $(G^{\hat\sigma})_{\hat\sigma\geq 0}$. □

Arzelà-Ascoli now gives local uniform convergence of $G^{\hat\sigma}$ and thereby local uniform convergence of ODE solutions $\mathrm{y}^{\hat\sigma}$ by

Theorem 4.7.4. *For every $\theta \geq 0$ we have $\mathrm{y}^{\hat{\sigma}} \to \mathrm{y}^0$ uniformly on $[0, \theta]$.*

Proof. By Lemma 4.7.2 there exists $\varepsilon > 0$ such that $\mathrm{y}^{\hat{\sigma}}(x) \in K := [y^0_\infty - \varepsilon, y^0_0 + \varepsilon]$ for all $x \geq 0$ and $\hat{\sigma} \in [0, \varepsilon]$. Lemma 4.7.3 gives that $L := \sup_{y \in K} \sup_{\hat{\sigma} \in [0,\varepsilon]} |(G^{\hat{\sigma}})'(y)| < \infty$. Moreover, equicontinuity and uniform boundedness of $(G^{\hat{\sigma}})_{\hat{\sigma} \geq 0}$, together with pointwise convergence, imply uniform convergence by [Wal98, Thm. II.7.IV], i.e. we have $\eta(\hat{\sigma}) := \sup_{y \in K} \big| G^{\hat{\sigma}}(y) - G^0(y) \big| \to 0$ as $\hat{\sigma} \searrow 0$. Now, for $\delta > 0$ it follows

$$
\begin{aligned}
C_\delta := \sup_{x \in [0,\delta]} |\mathrm{y}^{\hat{\sigma}}(x) - \mathrm{y}^0(x)| &\leq |y^{\hat{\sigma}}_0 - y^0_0| + \int_0^\delta \big| G^{\hat{\sigma}}\big(\mathrm{y}^{\hat{\sigma}}(x)\big) - G^0\big(\mathrm{y}^0(x)\big) \big| \,\mathrm{d}x \\
&\leq |y^{\hat{\sigma}}_0 - y^0_0| + \int_0^\delta \big| G^{\hat{\sigma}}\big(\mathrm{y}^{\hat{\sigma}}(x)\big) - G^{\hat{\sigma}}\big(\mathrm{y}^0(x)\big) \big| \,\mathrm{d}x + \int_0^\delta \big| G^{\hat{\sigma}}\big(\mathrm{y}^0(x)\big) - G^0\big(\mathrm{y}^0(x)\big) \big| \,\mathrm{d}x \\
&\leq |y^{\hat{\sigma}}_0 - y^0_0| + \delta C_\delta \underbrace{\sup_{y \in K} \big| (G^{\hat{\sigma}})'(y) \big|}_{\leq L} + \delta \underbrace{\sup_{y \in K} \big| G^{\hat{\sigma}}(y) - G^0(y) \big|}_{= \eta(\hat{\sigma})},
\end{aligned}
$$

which implies for $\delta \in (0, 1/L)$ that $C_\delta \leq \big(|y^{\hat{\sigma}}_0 - y^0_0| + \delta\eta(\hat{\sigma})\big)/(1 - \delta L) \to 0$ as $\hat{\sigma} \searrow 0$, since $y^{\hat{\sigma}}_0 \to y^0_0$ by Lemma 4.7.2. Using $\mathrm{y}^{\hat{\sigma}}(\delta) \to \mathrm{y}^0(\delta)$ instead of $y^{\hat{\sigma}}_0 \to y^0_0$ in the above estimation, we get

$$
C_{2\delta} = \sup_{x \in [0,2\delta]} |\mathrm{y}^{\hat{\sigma}}(x) - \mathrm{y}^0(x)| \leq (C_\delta + \delta\eta(\hat{\sigma}))/(1 - \delta L) \to 0 \quad \text{as } \hat{\sigma} \searrow 0.
$$

With $n := \lceil \theta/\delta \rceil$ iterative steps we cover all of $[0, \theta]$. □

5 Skorokhod M_1/J_1 stability for gains from large investors' strategies

This chapter is devoted to proving continuity of a controlled SDE solution in Skorokhod's M_1 and J_1 topologies and also uniformly, in probability, as a non-linear functional of the control strategy. The functional comes from a generalization of the price impact models introduced in Chapters 2 and 4. We show that M_1-continuity is the key to ensure that proceeds and wealth processes from (self-financing) càdlàg trading strategies are determined as the continuous extensions for those from continuous strategies. Returning to our overall theme of optimal liquidation problems, we demonstrate by example how continuity properties are useful to identify asymptotically realizable proceeds. This chapter presents a selection of the results and examples in [BBF19] and an extension of the main result of [BBF19, Thm. 3.7] to a more general setup with possibly stochastic liquidity. Section 5.1 sets the model and defines the proceeds functional for finite variation strategies. In Section 5.2 we extend this definition to a more general set of strategies and prove our main result of the chapter, Theorem 5.2.7. Section 5.3 gives one particular application related to the stochastic liquidity model of Chapter 4 where our stability result is a necessary prerequisite to extend the set of admissible strategies, since finite variation controls are sub-optimal in this new setup. Technical lemmas are deferred to Section 5.4.

5.1 Additive or multiplicative price impact

Extending on Chapters 2 and 4, without activity of large traders, the unaffected (discounted) price process of the risky asset would evolve according to the stochastic differential equation

$$\mathrm{d}\overline{S}_t = \overline{S}_{t-}(\xi_t \,\mathrm{d}\langle M\rangle_t + \mathrm{d}M_t)\,, \qquad \overline{S}_0 > 0, \tag{5.1}$$

where M is a locally square-integrable martingale that is quasi-left continuous (i.e. for any finite predictable stopping time τ, $\Delta M_\tau := M_\tau - M_{\tau-} = 0$ a.s.) with $\Delta M > -1$ and ξ is a predictable and bounded process. In particular, the predictable quadratic variation process $\langle M\rangle$ is continuous [JS03, Thm. I.4.2], and the unaffected (fundamental) price process $\overline{S} > 0$ can have jumps. We moreover assume that $\langle M\rangle = \int_0^\cdot \alpha_s \,\mathrm{d}s$ with a (locally) Lipschitz and $L^0(\mathbb{P})$ bounded density α, and that the martingale part of $\overline{S}$ is square integrable on compact time intervals. The assumptions on M are satisfied e.g. for $M = \int \sigma \,\mathrm{d}W$, where W is a Brownian motion and σ is a predictable stochastic volatility process that is bounded, or for Lévy processes M satisfying some integrability and lower bound on jumps.

We will assume throughout this chapter that strategies Θ are predictable càdlàg processes. The large investor's *market impact process* Y has dynamics

$$\mathrm{d}Y_t = -h(Y_t)\,\mathrm{d}t + \mathrm{d}N_t + \mathrm{d}\Theta_t \tag{5.2}$$

for some initial condition $Y_{0-} \in \mathbb{R}$, where N is a locally square-integrable continuous martingale whose quadratic variation process $\langle N \rangle$ is absolutely continuous w.r.t. the Lebesgue measure, $\langle N \rangle_t = \int_0^t \beta_t \, \mathrm{d}t$ with a (locally) Lipschitz continuous and $L^0(\mathbb{P})$ bounded β, such that the continuous covariation process $\langle N, M \rangle$ is of the form $\langle N, M \rangle_t = \int_0^t \gamma_t \, \mathrm{d}t$, again with γ locally Lipschitz and $L^0(\mathbb{P})$ bounded. With $N \equiv 0$, we are in the setup of Chapter 2 and with $N_t = \hat{\sigma} B_t$ for a Brownian motion B correlated to M we can capture the model of Chapter 4. The article [BBF19] which this chapter is based on essentially considers the case $N \equiv 0$ where $\beta_t = \gamma_t = 0$. But there the mean reversion of Y is w.r.t. $\mathrm{d}\langle M \rangle_t$ instead of $\mathrm{d}t$.

We assume that $h : \mathbb{R} \to \mathbb{R}$ is Lipschitz with $h(0) = 0$ and $h(y) \operatorname{sgn}(y) \geq 0$ for all $y \in \mathbb{R}$. The Lipschitz assumption on h guarantees existence and uniqueness of Y in a pathwise sense, see [PTW07, proof of Thm. 4.1] and Proposition 5.4.1 below. The sign assumption on h gives *transience* of the impact which recovers towards 0 (if $h(y) \neq 0$ for $y \neq 0$) when the large trader is inactive. The function h gives the speed of resilience at any level of Y_t and we will refer to it as *the resilience function*. For example, when $h(y) = \beta y$ for some constant $\beta > 0$, the market recovers at exponential rate (as in [OW13, AFS10, Løk14] and Chapter 4). Note that we also allow for $h \equiv 0$ in which case the impact is permanent as in [BB04]. Clearly, the process Y depends on Θ, and sometimes we will indicate this dependence as a superscript $Y = Y^{\Theta}$. Moreover, we will sometimes also indicate the dependence on the martingale N as $Y = Y^{N,\Theta}$. Note in particular, that we have $Y^{N,\Theta} = Y^{0,N+\Theta}$.

If the large investor trades according to a continuous strategy Θ, the observed price S at which infinitesimal quantities $\mathrm{d}\Theta$ are traded (see (5.4)) is given via (5.2) by

$$S_t := g(\overline{S}_t, Y_t)\,, \tag{5.3}$$

where the *price impact function* $(x, y) \mapsto g(x, y)$ is $C^{2,1}$ and non-negative with g_{xx} being locally Lipschitz in y, meaning that on every compact interval $I \subset \mathbb{R}$ there exists $K > 0$ such that $|g_{xx}(x, y) - g_{xx}(x, z)| \leq K|y - z|$ for all $x, y, z \in I$. Moreover, we assume $g(x, y)$ to be non-decreasing in both x and y. In particular, selling (buying) by the large trader causes the price S to decrease (increase). This price impact is transient due to (5.2).

Example 5.1.1. [BB04] consider a family of semimartingales $(S^{\theta})_{\theta \in \mathbb{R}}$ being parametrized by the large trader's risky asset position θ. In our setup, this corresponds to general price impact function g and $h \equiv 0$, meaning that impact is permanent. A known example in the literature on transient price impact is the additive case, $S = \overline{S} + f(Y)$, where [OW13] take $f(y) = \lambda y$ to be linear, motivated from a block-shaped limit order book. For generalizations to non-linear increasing $f : \mathbb{R} \to [0, \infty)$, see [AFS10, PSS11]. Note that we require $0 \leq g \in C^{2,1}$ for Theorem 5.2.7, see Remark 5.2.9. A (somewhat technical) modification of the model by [OW13], that fits with our setup and ensures positive asset prices, could be to take $g(\overline{S}, Y) = \varphi(\overline{S} + f(Y))$ with a non-negative increasing $\varphi \in C^2$ satisfying $\varphi(x) = x$ on $[\varepsilon, \infty)$ and $\varphi(\cdot) = 0$ on $(-\infty, -\varepsilon]$ for some $\varepsilon > 0$. An example that naturally ensures positive asset prices is multiplicative impact $S = f(Y)\overline{S}$ as in as in Chapters 2 to 4, for f being strictly positive, non-decreasing, and with $f \in C^1$ (to satisfy the conditions on g). Also here, the function f can be interpreted as resulting from a limit order book, see Remark 2.1.1.

While impact and resilience are given by general non-parametric functions, note that these are static. Considering such a model as a low (rather than high) frequency model, we do consider approximations by continuous and finite variation strategies to be relevant. To start, let Θ be a continuous process of finite variation (f.v., being adapted). Then, the cumulative proceeds (negative expenses), denoted by $L(\Theta)$, that are the variations in the bank account to finance buying and selling of the risky asset according to the strategy, can be defined (pathwise) in an unambiguous way. Indeed, proceeds over period $[0, T]$ from a strategy Θ that is continuous should be (justified also by Lemma 5.2.1)

$$L_T(\Theta) := -\int_0^T S_u \, \mathrm{d}\Theta_u = -\int_0^T g(\overline{S}_u, Y_u) \, \mathrm{d}\Theta_u. \tag{5.4}$$

Our main task in this chapter is to extend by stability arguments the model from continuous to more general trading strategies, in particular such involving block trades and even more general ones with càdlàg paths, assuming transient price impact but no further frictions, like e.g. bid-ask spread. To this end, we will adopt the following point of view: approximately similar trading behavior should yield similar proceeds. The next section will make precise what we mean by "similar" by considering different topologies on the càdlàg path space. It turns out that the natural extension of the functional L from the space of continuous f.v. paths to the space of càdlàg f.v. paths which makes the functional L continuous in all of the considered topologies is as follows: for discontinuous trading we take the proceeds from a block market buy or sell order of size $|\Delta\Theta_\tau|$, executed immediately at a predictable stopping time $\tau < \infty$, to be given by

$$-\int_0^{\Delta\Theta_\tau} g(\overline{S}_{\tau-}, Y_{\tau-} + x) \, \mathrm{d}x, \tag{5.5}$$

and so the proceeds up to T from a f.v. strategy Θ with continuous part Θ^c are

$$L_T(\Theta) := -\int_0^T g(\overline{S}_u, Y_u) \, \mathrm{d}\Theta^c_u - \sum_{\substack{\Delta\Theta_t \neq 0 \\ 0 \le t \le T}} \int_0^{\Delta\Theta_t} g(\overline{S}_{t-}, Y_{t-} + x) \, \mathrm{d}x. \tag{5.6}$$

Note that a block sell order means that $\Delta\Theta_t < 0$, so the average price per share for this trade satisfies $S_t \le -\frac{1}{\Delta\Theta_t} \int_0^{\Delta\Theta_t} g(\overline{S}_t, Y_{t-} + x) \, \mathrm{d}x \le S_{t-}$. Similarly, the average price per share for a block buy order, $\Delta\Theta_t > 0$, is between S_{t-} and S_t. The expression in (5.5) could be justified from a limit order book perspective for some cases of g, as noted in Example 5.1.1. But we will derive it in the next section using stability considerations.

The form (5.6) appears as a usual choice for the objective functional in the singular stochastic control literature, see e.g. [Zhu92]. Already [Tak97, p. 609] justifies jump terms as in (5.6) with an M_1 approximation argument, although there he misqualifies M_1 convergence as pointwise convergence at continuity points of the limit, which is only true for monotone processes.

Remark 5.1.2. The aim to define a model for trading under price impact for general strategies is justified by applications in finance, which encompass trade execution, utility optimization and hedging. While also e.g. [BB04, BR17, ÇJP04] define proceeds for semimartingale strategies, their definitions are not ensuring continuity in the M_1 topology, in contrast to Theorem 5.2.7. Another difference to [BB04, BR17] is that our presentation is not going to rely on non-linear stochastic integration theory due to Kunita or, respectively, Carmona and Nualart.

5.2 Continuity of the proceeds in various topologies

In this section we will discuss questions about continuity of the proceeds process $\Theta \mapsto L.(\Theta)$ with respect to various topologies: the ucp topology and the Skorokhod J_1 and (in particular) M_1 topologies. Each one captures different stability features, the suitability of which may vary with application context.

Let us observe that for a continuous bounded variation trading strategy Θ the proceeds from trading should be given by (5.4). To this end, let us make just the assumption that

$$\text{a block order of a size } \Delta \text{ at some (predictable) time } t \text{ is executed at some average price per share which is between } S_{t-} = g(\overline{S}_t, Y_{t-}) \text{ and } S_t = g(\overline{S}_t, Y_t), \text{ where } \Delta Y_t = \Delta. \tag{5.7}$$

The assumption is natural, stating that a block trade is executed at an average price per share that is somewhere between the asset prices observed immediately before and after the execution. Assumption (5.7) means that proceeds by a simple strategy as in (5.9) are

$$L_t(\Theta^n) = -\sum_{k:\ t_k \le t} \xi_k (\Theta_{t_k} - \Theta_{t_{k-1}}) \tag{5.8}$$

for some random variable ξ_k between $g(\overline{S}_{t_k}, Y^{\Theta^n}_{t_k-})$ and $g(\overline{S}_{t_k}, Y^{\Theta^n}_{t_k})$. Note that at this point we have not specified the proceeds (negative expenses) from block trades, but we only assume that they satisfy some natural bounds. Yet, this is indeed already sufficient to derive the functional (5.4) for continuous strategies as a limit of simple ones.

Lemma 5.2.1. *For $T > 0$, approximate a continuous f.v. process $(\Theta_t)_{t\in[0,T]}$ by a sequence $(\Theta^n_t)_{t\in[0,T]}$ of simple trading strategies given as follows: For a sequence of partitions $\{0 = t_0 < t_1 < \cdots < t_{m_n} = T\}$, $n \in \mathbb{N}$, with $\sup_{1\le k\le m_n}|t_k - t_{k-1}| \to 0$ for $n \to \infty$, let*

$$\Theta^n_t := \Theta_0 + \sum_{k=1}^{m_n} (\Theta_{t_k} - \Theta_{t_{k-1}}) \mathbb{1}_{[t_k,T]}(t)\,, \quad t \in [0,T]. \tag{5.9}$$

Assume (5.7) holds. Then $\sup_{0\le t\le T}|L_t(\Theta^n) + \int_0^t S_u \,\mathrm{d}\Theta_u| \to 0$ a.s. for $n \to \infty$.

Proof. Note that $\sup_{u\in[0,T]}|\Theta^n_u - \Theta_u| \to 0$ as $n \to \infty$. The solution map $\Theta \mapsto Y^\Theta$ is continuous with respect to the uniform norm, see Proposition 5.4.1. Therefore,

$$\sup_{u\in[0,T]} |Y^{\Theta^n}_u - Y^\Theta_u| \to 0 \quad \text{a.s. for } n \to \infty. \tag{5.10}$$

Note that for $Y := Y^\Theta$, $\Delta\Theta_{t_k} := \Theta_{t_k} - \Theta_{t_{k-1}} = \Delta\Theta^n_{t_k}$ and ξ_k between $g(\overline{S}_{t_k}, Y^{\Theta^n}_{t_k-})$ and $g(\overline{S}_{t_k}, Y^{\Theta^n}_{t_k})$ we have

$$\begin{aligned} |\xi_k - g(\overline{S}_{t_k}, Y_{t_k})| &\le \mathrm{Lip}_g(\overline{S}_{t_k}, \omega) \max\{|Y_{t_k} - Y^{\Theta^n}_{t_k}|, |Y_{t_k} - Y^{\Theta^n}_{t_k-}|\} \\ &\le \mathrm{Lip}_g(\overline{S}_{t_k}, \omega)\big(|Y_{t_k} - Y^{\Theta^n}_{t_k}| + |\Delta\Theta_{t_k}|\big), \end{aligned}$$

where $\mathrm{Lip}_g(x,\omega)$ denotes the Lipschitz constant of $y \mapsto g(x,y)$ on a compact set, depending on the (bounded) realizations for $\omega \in \Omega$ of Y^{Θ} and Y^{Θ^n}, $n \in \mathbb{N}$, on the interval $[0,T]$; such a compact set exits since Θ is continuous and $\sup_{u\in[0,T]}\big|Y^{\Theta}_u - Y^{\Theta^n}_u\big|$ can be bounded by a factor times the uniform distance between Θ and Θ^n on $[0,T]$, cf. [PTW07, proof of Thm. 4.1]. Hence,

$$L_t(\Theta^n) = -\sum_{k:\ t_k\le t} g(\overline{S}_{t_k}, Y^{\Theta}_{t_k})\big(\Theta_{t_k} - \Theta_{t_{k-1}}\big) + \mathcal{E}^n_t\,, \tag{5.11}$$

$$\text{where } |\mathcal{E}^n_t| \le \Big(\sup_{u\in[0,T]} \mathrm{Lip}_g(\overline{S}_u,\omega)\Big)\sum_{k=1}^{n}\big(\big|Y_{t_k} - Y^{\Theta^n}_{t_k}\big| + |\Delta\Theta_{t_k}|\big)|\Delta\Theta_{t_k}| \tag{5.12}$$

$$\le C(\omega)\Big(\sup_{1\le k\le n}\big|Y_{t_k} - Y^{\Theta^n}_{t_k}\big|\Big)|\Theta(\omega)|_{\mathrm{TV}} + C(\omega)\sum_{k=1}^{n}|\Delta\Theta_{t_k}|^2 \tag{5.13}$$

$$\to 0 \quad \text{a.s. for } n\to\infty \text{ (uniformly in } t), \tag{5.14}$$

thanks to (5.10) and the fact that Θ has continuous paths of finite variation. The claim follows since by dominated convergence the Riemann-sum process in (5.11) converges a.s. to the Stieltjes-integral process $-\int_0^{\cdot} S_u \,\mathrm{d}\Theta_u$ uniformly on $[0,T]$. □

Example 5.2.2 (Continuity issues for an alternative "ad-hoc" definition of proceeds)**.** Consider the problem of optimally liquidating $\Theta_{0-} = 1$ risky asset in time $[0,T]$ while maximizing expected proceeds. In view of assumption (5.7), an alternative but possibly "ad-hoc" definition for proceeds $\tilde{L}_T$ of simple strategies could be to consider just some price for each block trade, similarly to [BB04, Section 3] or [HH11, Example 2.4]. For multiplicative impact $g(\overline{S},Y) = \overline{S}f(Y)$, taking e.g. the price directly after the impact would yield for simple strategies Θ^n that trade at times $\{0 = t^n_0 < t^n_1 < \cdots < t^n_n = T\}$ the proceeds $\tilde{L}_T(\Theta^n) = -\sum_{k=0}^n \overline{S}_{t^n_k} f(Y^{\Theta^n}_{t^n_k})\Delta\Theta^n_{t^n_k}$. The family $(\Theta^n)_n$ of strategies which liquidate an initial position of size 1 until time $1/n$ in n equidistant blocks of uniform size is given by $\Theta^n_t := \sum_{k=1}^n \frac{n-k+1}{n}\mathbb{1}_{[\frac{k-1}{n^2},\frac{k}{n^2})}(t)$. With unaffected price $\overline{S}_t = e^{-\delta t}\widetilde{M}_t$ for a continuous martingale $\widetilde{M}$, and permanent impact ($h\equiv 0$), i.e. $Y_t = \Theta_t - 1$, this yields $\mathbb{E}[\tilde{L}_T(\Theta^n)] \to \int_0^1 f(-y)\,\mathrm{d}y$ for $n\to\infty$. Given $\delta \ge 0$, for any non-increasing simple strategy $\Theta = \sum_{k=1}^n \Theta_{\tau_k}\mathbb{1}_{[\![\tau_{k-1},\tau_k[\![}$ with $\Theta_{0-} = 1$ holds that $\mathbb{E}[\tilde{L}(\Theta)] \le \int_0^1 f(-y)\,\mathrm{d}y$ with strict inequality for $\delta > 0$. So the control sequence (Θ^n) is only asymptotically optimal among all simple monotone liquidation strategies.

Remark 5.2.3. Note that Example 5.2.2 is a toy example, since for permanent impact the optimal strategy (considering asymptotically realizable proceeds) is trivial and in case $\delta = 0$ any strategy is optimal, cf. [GZ15, Prop. 3.5(III) and the comment preceding it]. Nevertheless, this example shows that the object of interest are *asymptotically realizable* proceeds, an insight due to [BB04]. For analysis, it thus appears convenient and sensible not to make a formal distinction of (sub-optimal) realizable and asymptotically realizable proceeds, but to consider the latter and interpret strategies accordingly. Investigating asymptotically realizable proceeds can help to answer questions on modeling issues, e.g. whether the large investor could sidestep liquidity costs entirely and in effect act as a small investor, cf. [BB04, ÇJP04]. One could impose, like [ÇST10], additional constraints on strategies to avoid such issues; But in such tweaked models one could not investigate the effects from some given illiquidity friction alone, in isolation from other constraints, because results from an analysis will be consequences of the combination of both frictions.

Using Itô's formula for $G(\overline{S}_t, Y_t^\Theta)$ for the C^2 function $G(x,y) := \int_c^y g(x,z)\,\mathrm{d}z$ with constant c, we can obtain the following alternative representation of the functional in (5.4) for continuous f.v. strategies:

$$\begin{aligned} L(\Theta) = &\int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-}^\Theta)\,\mathrm{d}\overline{S}_u + \int_0^\cdot g(\overline{S}_{u-}, Y_{u-}^\Theta)\,\mathrm{d}N_u + \int_0^\cdot k(\overline{S}_u, Y_u^\Theta, \alpha_u, \beta_u, \gamma_u)\,\mathrm{d}u \\ &- \big(G(\overline{S}_\cdot, Y_\cdot^\Theta) - G(\overline{S}_0, Y_{0-}^\Theta)\big) \\ &+ \sum_{\substack{\Delta\overline{S}_u \neq 0 \\ 0\le u\le \cdot}} \big(G(\overline{S}_u, Y_u^\Theta) - G(\overline{S}_{u-}, Y_u^\Theta) - G_x(\overline{S}_{u-}, Y_u^\Theta)\Delta\overline{S}_u\big)\,, \end{aligned} \tag{5.15}$$

where $k(x,y,a,b,c) := \frac{1}{2}G_{xx}(x,y)x^2 - g(x,y)h(y)a + \frac{1}{2}g_y(x,y)b + g_x(x,y)c$ and using that $\overline{S}$ and Y have no common jumps. The advantage of this representation is that the right-hand side of (5.15) makes sense for any predictable process Θ with càdlàg paths in contrast to the term in (5.4) This form of the proceeds will turn out to be helpful for the stability analysis. We will show that the right-hand side in (5.15) is continuous in the control Θ when the path-space of Θ, the càdlàg path space, is endowed with various topologies. Hence, it can be used to define the proceeds for general trading strategies by continuity. Next section is going to discuss the topologies that will be of interest.

5.2.1 The Skorokhod space and its M_1 and J_1 topologies

We are going to derive a continuity result (Theorem 5.2.7) for the functional L in different topologies on the space $D \equiv D([0,T]) := D([0,T];\mathbb{R})$ of real-valued càdlàg paths on the time interval $[0,T]$. Following the convention by [Sko56], we take each element in $D[0,T]$ to be left-continuous at time T.[1] One could also consider initial and terminal jumps by extending the paths, see Remark 5.2.6. At this point, let us remark that finite horizon T is not essential for the results below, whose analysis carries over to the time interval $[0,\infty)$ because the topology on $D([0,\infty))$ is induced by the topologies of $D([0,T])$ for $T \ge 0$. More precisely, for the topologies we are interested in, $x_n \to x$ as $n \to \infty$ in $D([0,\infty))$ if $x_n \to x$ in $D([0,t])$ for the restrictions of x_n, x on $[0,t]$, for any t being a continuity point of x, see [Whi02, Sect. 12.9].

Convergence in the uniform topology is rather strong, in that approximating a path with a jump is only possible if the approximating sequence has jumps of comparable size at the same time. If one is interested in stability with respect to slight shift of the execution in time, then a familiar choice that also makes D separable, the Skorokhod J_1 topology, might be appropriate; for comprehensive study, see [Bil99, Ch. 3]. However, also here an approximating sequence for a path with jumps needs jumps of comparable size, if only at nearby times. To capture the occurrence of the so-called *unmatched jumps*, i.e. jumps that appear in the limit of continuous processes, another topology on D is more appropriate, the Skorokhod M_1 topology. Recall that $x_n \to x$ in (D, d_{M_1}) if $d_{M_1}(x_n, x) \to 0$ as $n \to \infty$, with

$$d_{M_1}(x_n, x) := \inf\big\{\|u - u_n\| \vee \|r - r_n\| \;\big|\; (u,r) \in \Pi(x), (u_n, r_n) \in \Pi(x_n)\big\}\,, \tag{5.16}$$

[1] This is implicitly assumed also in [Whi02], see the compactness criterion in Thm. 12.12.2 which is borrowed from [Sko56].

where $\|\cdot\|$ denotes the uniform norm on $[0,1]$ and $\Pi(x)$ is the set of all *parametric representations* $(u,r) : [0,1] \to \Gamma(x)$ of the completed graph (with vertical connections at jumps) $\Gamma(x)$ of $x \in D$, see [Whi02, Sect. 3.3]. In essence, two functions $x, y \in D$ are near to each other in M_1 if one could run continuously a particle on each graph $\Gamma(x)$ and $\Gamma(y)$ from the left endpoint toward the right endpoint such that the two particles are nearby in time and space. In particular, it is easy to see that a simple jump path could be approximated in M_1 by a sequence of absolutely continuous paths, in contrast to the uniform and the J_1 topologies. More precisely, we have the following

Proposition 5.2.4. *Let $x \in D([0,T])$ and consider the Wong-Zakai approximation sequence $(x_n) \subset D([0,T])$ defined by $x_n(t) := n\int_{t-1/n}^{t} x(s)\,\mathrm{d}s$, $t \in [0,T]$. Then*

$$x_n \to x \quad \text{for } n \to \infty, \quad \text{in } (D([0,T]), M_1).$$

Proof. To ease notation, we embed a path x in $D([0,\infty))$ and consider the corresponding approximating sequence for the extended path on $[0,\infty)$. The claim follows by restricting to the domain $[0,T]$, as 0 and T are continuity points of x, cf. [Whi02, Sect. 12.9]. The idea is to construct explicitly parametric representations of $\Gamma(x)$ and $\Gamma(x_n)$ that are close enough. For this purpose, we need to add "fictitious" time to be able to parametrize the segments that connect jump points of x. Indeed, let (a_k) be a fixed convergent series of strictly positive numbers and let $t_1, t_2, \dots$ be the jump times of x ordered such that $|\Delta x(t_1)| \geq |\Delta x(t_2)| \geq \dots$ and $t_k < t_{k+1}$ if $|\Delta x(t_k)| = |\Delta x(t_{k+1})|$. Set $\delta(t) := \sum_k a_k \mathbb{1}_{\{t_k \leq t\}}$, the total "fictitious" time added to parametrize the jumps of x up to time t.

Consider the time-changes $\gamma_n(t) := n\int_{t-1/n}^{t}(\delta(u)+u)\,\mathrm{d}u$ and $\gamma_0(t) := \delta(t)+t$, $t \geq 0$, together with their continuous inverses $\gamma_n^{-1}(s) := \inf\{u > 0 \mid \gamma_n(u) > s\}$ for $s \geq 0$, $n \geq 0$. It is easy to check that we have

$$\gamma_n^{-1}(s) - 1/n < \gamma_0^{-1}(s) < \gamma_n^{-1}(s) < \infty \quad \text{for } s \geq 0, \tag{5.17}$$

because $\gamma_n(t) < \gamma_0(t) < \gamma_n(t+1/n)$, cf. [KPP95, Lemma 6.1]. Consider the sequence $u_n(s) := x_n(\gamma_n^{-1}(s))$ for $s \geq 0$ and let

$$u(s) := \begin{cases} x(\gamma_0^{-1}(s)) & \text{if } \eta_1(s) = \eta_2(s), \\ x(\gamma_0^{-1}(s)) \cdot \frac{s-\eta_1(s)}{\eta_2(s)-\eta_1(s)} + x(\gamma_0^{-1}(s)-) \cdot \frac{\eta_2(s)-s}{\eta_2(s)-\eta_1(s)} & \text{if } \eta_1(s) \neq \eta_2(s), \end{cases}$$

where $[\eta_1(s), \eta_2(s)]$ is the "fictitious" time added for a jump at time $t = \gamma_0^{-1}(s)$, i.e. $\eta_1(s) := \sup\{\tilde{s} \mid \gamma_0^{-1}(\tilde{s}) < \gamma_0^{-1}(s)\}$ and $\eta_2(s) := \inf\{\tilde{s} \mid \gamma_0^{-1}(\tilde{s}) > \gamma_0^{-1}(s)\}$, as in [KPP95, p. 368]. Then [KPP95, Lemma 6.2] gives $\lim_{n\to\infty} u_n = u$, uniformly on bounded intervals; our setup corresponds to $f \equiv 1$ there, so our u_n, u correspond to $V^{1/n}, V$ there.

Now the claim follows by observing that (u_n, γ_n^{-1}) is a parametric representation of the completed graph of x_n, i.e. $(u_n, \gamma_n^{-1}) \in \Pi(x_n)$, and $(u, \gamma_0^{-1}) \in \Pi(x)$ which are arbitrarily close when n is big. □

Remark 5.2.5. A direct corollary of Proposition 5.2.4 is that $D([0,T])$ is the closure of the set of absolutely continuous functions in the Skorokhod M_1 topology, in contrast to the uniform or Skorokhod J_1 topologies where a jump in the limit can only be approximated by jumps of comparable sizes.

Remark 5.2.6 (Extended paths)**.** To include trading strategies that could additionally have initial and terminal jumps in our analysis, one may embed the paths of such strategies in the slightly larger space $D([-\varepsilon, T+\varepsilon];\mathbb{R})$ for some $\varepsilon > 0$, e.g. $\varepsilon = 1$, by setting $x(s) = x(0-)$ for $s \in [-\varepsilon, 0)$ and $x(s) = x(T+)$ for $s \in (T, T+\varepsilon]$; we will refer to thereby embedded paths as *extended paths*. This extension is relevant when trying to approximate jumps at terminal time by absolutely continuous strategies in a non-anticipative way as e.g. in Proposition 5.2.4 where it is clear that a bit more time could be required after a jump occurs in order to approximate it. In particular, by considering extended paths the result of Proposition 5.2.4 holds if one allows for initial and terminal jumps of x, but convergence holds in the extended paths space.

5.2.2 Main stability results

Our main result is stability of the functional L defined by the right-hand side of (5.15) for processes Θ with càdlàg paths.

Theorem 5.2.7. *Let a sequence of predictable processes (Θ^n) converge to the predictable process Θ in (D,ρ), in probability, where ρ denotes the uniform topology, the Skorokhod J_1 or M_1 topology, being generated by a suitable metric d. Assume that (Θ^n) is bounded by an L^0 variable, i.e. there exists $K \in L^0(\mathbb{P})$ such that $\sup_{0\le t\le T}|\Theta^n_t| \le K$ for all n. Then the sequence of processes $L(\Theta^n)$ converges to $L(\Theta)$ in (D,ρ) in probability, i.e.*

$$\mathbb{P}\big[d\big(L(\Theta^n), L(\Theta)\big) \ge \varepsilon\big] \to 0 \qquad \text{for } n\to\infty \text{ and } \varepsilon > 0. \tag{5.18}$$

In particular, there is a subsequence $L(\Theta^{n_k})$ that converges a.s. to $L(\Theta)$ in (D,ρ).

Note that e.g. for almost sure convergence $\Theta^n \to \Theta$ in (D,ρ), the $L^0(\mathbb{P})$ boundedness condition is automatically fulfilled.

Proof. By considering subsequences, one could assume that the sequence (Θ^n) converges to Θ in (D,ρ) a.s. The idea for the proof is to show that each summand in the definition of L is continuous. But as D endowed with J_1 or M_1 is not a topological vector space, since addition is not continuous in general, further arguments will be required. Addition is continuous (and hence also multiplication) if for instance the summands have no common jumps, see [JS03, Prop. VI.2.2] for J_1 and [Whi02, Cor. 12.7.1] for M_1. In our case however, there are three terms in L that can have common jumps, namely the stochastic integral process $\int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u$, the sum $\Sigma := \sum_{u\le\cdot}\big(G(\overline{S}_u, Y_u) - G(\overline{S}_{u-}, Y_u) - G_x(\overline{S}_{u-}, Y_u)\Delta\overline{S}_u\big)$ of jumps and the term $-G(\overline{S}, Y)$. At jump times of Θ (i.e. of Y) which are predictable stopping times, $\overline{S}$ does not jump since it is quasi-left continuous. Hence the only common jump times can be jumps times of $\overline{S}$ which are totally inaccessible. If $\Delta\overline{S}_\tau \neq 0$, we have for the jumps that $\Delta(\int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u)_\tau = G_x(\overline{S}_{\tau-}, Y_\tau)\Delta\overline{S}_\tau$ and also that $\Delta(-G(\overline{S}, Y))_\tau = -\big(G(\overline{S}_\tau, Y_\tau) - G(\overline{S}_{\tau-}, Y_\tau)\big)$, because $\Delta Y_\tau = 0$ a.s. Since moreover $\Delta\Sigma_\tau = G(\overline{S}_\tau, Y_\tau) - G(\overline{S}_{\tau-}, Y_\tau) - G_x(\overline{S}_{\tau-}, Y_\tau)\Delta\overline{S}_\tau$, one has cancellation of jumps at jump times of $\overline{S}$. However, these are times of continuity for Y and this will be crucial below to deduce continuity of addition on the support of $\big(\int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u, \Sigma, -G(\overline{S}, Y)\big)$ in $(D,\rho)\times(D,\rho)\times(D,\rho)$.

First consider the case where $(N + \Theta^n)$ is a uniformly bounded sequence. Since N is continuous, we get $N + \Theta^n \to N + \Theta$ in (D, ρ). Then the processes

$$\mathrm{d}Y_t^n = -h(Y_t^n)\,\mathrm{d}t + \mathrm{d}N_t + \mathrm{d}\Theta_t^n, \quad Y_{0-}^n = y\,,$$

are uniformly bounded, so we can assume w.l.o.g. that h, G, G_x and k are ω-wise Lipschitz continuous and bounded (it is so on the range of all Y^n, Y, which is contained in a compact subset of $\mathbb{R}$). By Proposition 5.4.1 we have $Y^n \to Y$ in (D, ρ), almost surely. This implies $(\overline{S}, Y^n) \to (\overline{S}, Y)$ almost surely, by absence of common jumps of $\overline{S}$ and Y, cf. [JS03, Prop. VI.2.2b] for J_1 and[2] [Whi02, Thm. 12.6.1 and 12.7.1] for M_1. By the Lipschitz property of G and (for the M_1 case) monotonicity of $G(\cdot, y)$ and $G(x, \cdot)$, we get

$$G(\overline{S}, Y^n) \to G(\overline{S}, Y) \quad \text{in } (D, \rho), \text{ a.s.} \tag{5.19}$$

Indeed, for the M_1 topology, it is easy to see that $(G(u^1, u^2), r) \in \Pi(G(\overline{S}, Y))$ for any parametric representation $((u^1, u^2), r)$ of $(\overline{S}, Y)$, because at jump times t of $G(\overline{S}, Y)$, $z \mapsto r(z) \equiv t$ is constant on an interval $[z_1, z_2]$, and either u^1 or u^2 is constant on $[z_1, z_2]$.

Note that jump times of Θ and Y coincide, and form a random countable subset of $[0, T]$. Moreover, convergence in (D, ρ) implies local uniform convergence at continuity points of the limit (for ρ being the M_1 topology, cf. [Whi02, Lemma 12.5.1], for the J_1 topology cf. [JS03, Prop. VI.2.1]). Hence, $Y_t^n \to Y_t$ for almost all $t \in [0, T]$, $\mathbb{P}$-a.s. By Lipschitz continuity of k, we get $k(\overline{S}_t, Y_t^n, \alpha_t, \beta_t, \gamma_t) \to k(\overline{S}_t, Y_t, \alpha_t, \beta_t, \gamma_t)$, for almost-all $t \in [0, T]$, $\mathbb{P}$-a.s. By dominated convergence, we conclude that

$$\int_0^{\cdot} k(\overline{S}_u, Y_u^n, \alpha_u, \beta_u, \gamma_u)\,\mathrm{d}u \to \int_0^{\cdot} k(\overline{S}_u, Y_u, \alpha_u, \beta_u, \gamma_u)\,\mathrm{d}u$$

uniformly on $[0, T]$, a.s. Hence these two summands in the definition of L, see (5.15), are (ω-wise) continuous in Θ.

Now we treat the stochastic integral and jump terms in (5.15). By the above arguments we can also deal with the drift in the process $\overline{S}$, since $\langle M \rangle$ is absolutely continuous w.r.t. Lebesgue measure. Thus we may assume w.l.o.g. that $\overline{S}$ is a martingale. In particular, up to a localization argument (see below for details), we can assume that $\overline{S}$ and N are bounded and therefore the stochastic integrals are true martingales, since their integrands $G_x(\overline{S}_{u-}, Y_{u-}^n)$ and $g(\overline{S}_{u-}, Y_{u-}^n)$ are bounded. Having $Y^n \to Y$ a.e. on the space $(\Omega \times [0, T], \mathbb{P} \otimes \mathrm{Leb}([0, T]))$, we can conclude convergence of the stochastic integrals in the uniform topology, in probability. Dominated convergence on $\big([0, T], \mathrm{Leb}([0, T])\big)$ yields

$$\int_0^T (Y_{u-}^n - Y_{u-})^2\,\mathrm{d}\langle \overline{S} \rangle_u \to 0 \quad \text{as } n \to \infty, \ \mathbb{P}\text{-a.s.}$$

Since Y^n, Y are uniformly bounded one gets, again by dominated convergence, that

$$\mathbb{E}\Big[\int_0^T (Y_{u-}^n - Y_{u-})^2\,\mathrm{d}\langle \overline{S} \rangle_u\Big] \to 0 \qquad \text{and} \qquad \mathbb{E}\Big[\int_0^T (Y_{u-}^n - Y_{u-})^2\,\mathrm{d}\langle N \rangle_u\Big] \to 0,$$

[2] Using the strong M_1 topology in $D([0, \infty); \mathbb{R}^2)$.

as $n \to \infty$, i.e. $Y^n_- \to Y_-$ in $L^2(\Omega \times [0,T], \mathrm{d}\mathbb{P} \otimes \mathrm{d}\langle \overline{S} \rangle)$ and in $L^2(\Omega \times [0,T], \mathrm{d}\mathbb{P} \otimes \mathrm{d}\langle N \rangle)$. By localization (to bound $\overline{S}$ and use that $G_x(x,y)$ and $g(x,y)$ are locally Lipschitz in y), Itô's isometry and Doob's martingale inequality, we get

$$\mathbb{P}\Big[\sup_{0 \le t \le T} \Big| \int_0^t G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u - \int_0^t G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u \Big| \ge \varepsilon \Big] \to 0 \qquad \text{and}$$
$$\mathbb{P}\Big[\sup_{0 \le t \le T} \Big| \int_0^t g(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}N_u - \int_0^t g(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}N_u \Big| \ge \varepsilon \Big] \to 0,$$

as $n \to \infty$. For the sum of jumps Σ^n (defined like Σ, but with Y^n instead of Y) we have a.s. uniform convergence $\Sigma^n \to \Sigma$ by Lemma 5.4.4. Hence, the sum $\int_0^t G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u + \int_0^t g(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}N_u + \Sigma^n$ converges in ucp. To conclude on the proceeds, note that at jump times of $\overline{S}$, when cancellation of jumps occurs, one has continuity of Y and hence local uniform convergence of the sequence Y^n. For our setup, Lemmas 5.4.2 and 5.4.3 show continuity of addition on the support of $\big(\int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u + \Sigma, -G(\overline{S}_\cdot, Y_\cdot)\big)$ (that is, along the support of the pairs $\big(\int_0^\cdot G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u + \Sigma^n, -G(\overline{S}_\cdot, Y^n_\cdot)\big)\big)$ for the J_1 and M_1 topologies, respectively. All other terms in (5.15) are continuous processes. So the continuous mapping theorem [Kal02, Lem. 4.3] yields the claim for the proceeds functional L (the uniform topology being stronger than ρ).

It remains to investigate the more general case of $\overline{S}$, N and (Θ^n) being only bounded in $L^0(\mathbb{P})$. Note that the continuity of all terms except the stochastic integrals in the definition of L was proven ω-wise; in this case $\sup_n \sup_{0 \le t \le T} |N_t + \Theta^n_t(\omega)| < \infty$ (by the a.s. convergence of Θ^n to Θ in (D, ρ)) and hence the same arguments carry over here by restricting our attention to compact sets (depending on ω). Hence refinement of the argument above is only needed for the stochastic integral terms. The bound on $\overline{S}$, N and (Θ^n) means that for every $\varepsilon > 0$ there exists $\Omega_\varepsilon \in \mathcal{F}$ with $\mathbb{P}(\Omega_\varepsilon) > 1 - \varepsilon$ and a positive constant K_ε which is a uniform bound for the sequence (together with the limit Θ) on Ω_ε. For the stopping time $\tau := \inf \tau_n$, where $\tau_n := \inf\{t \ge 0 \mid |\Theta^n_t| \vee |N_t| \vee |\overline{S}_t| > K_\varepsilon\} \wedge T$ (τ is a stopping time because the filtration is right-continuous by our assumptions), we then have that $\tau = T$ on Ω_ε. By the arguments above we conclude that $d\big(\int_0^{\cdot \wedge \tau} G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u, \int_0^{\cdot \wedge \tau} G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u\big) \to 0$ in probability. Since $\int_0^{\cdot \wedge \tau} G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u = \int_0^\cdot G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u$ on Ω_ε, we conclude

$$\mathbb{P}\Big[d\Big(\int_0^\cdot G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u, \int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u\Big) \ge \varepsilon\Big] \le 2\varepsilon$$

for all n large enough. Similarly, we find

$$\mathbb{P}\Big[d\Big(\int_0^\cdot g(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}N_u, \int_0^\cdot g(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}N_u\Big) \ge \varepsilon\Big] \le 2\varepsilon$$

for all n large enough. Since N is continuous, the sum of both stochastic integrals also converges in the same manner. This finishes the proof since ε was arbitrary. □

Remark 5.2.8. Inspection of the proof above reveals that predictability of the strategies is only needed to show why the addition map is continuous when there is cancellation of jumps in (5.15); indeed, for predictable Θ the processes Y^Θ and $\overline{S}$ will have no common jump and this was sufficient for the arguments. However, in the case when M (and thus $\overline{S}$) is continuous, only one term in (5.15) might have jumps, namely $G(\overline{S}, Y^{\Theta})$. Hence, in this case the conclusion of Theorem 5.2.7 even holds under the relaxed assumption that the càdlàg strategies are merely adapted, instead of being predictable.

Remark 5.2.9. Our assumption of positive prices (and monotonicity of $x \mapsto g(x,y)$) has been (just) used to prove the M_1-convergence of $G(\overline{S}, Y^n)$ in (5.19). If one would want to consider a model where prices could become negative (like additive impact $S = \overline{S} + f(Y)$, see Example 5.1.1), then M_1-continuity of proceeds would not hold in general, as a simple counter-example can show. Yet, the above proof still shows $L_t(\Theta^n) \to L_t(\Theta)$ in probability, for all $t \in [0,T]$ where $\Delta\Theta_t = 0$. Also note that for continuous Θ^n converging in M_1 to a continuous strategy Θ, hence also uniformly, one obtains that proceeds $L(\Theta^n) \to L(\Theta)$ converge uniformly, in probability.

An important consequence of Theorem 5.2.7 is a stability property for our model. It essentially implies that we can approximate each strategy by a sequence of absolutely continuous strategies, corresponding to small intertemporal shifts of reassigned trades, whose proceeds will approximate the proceeds of the original strategy. More precisely, if we restrict our attention to the class of monotone strategies, then we can restate this stability in terms of the Prokhorov metric on the pathwise proceeds (which are monotone and hence define measures on the time axis). This result on stability of proceeds with respect to small intertemporal Wong-Zakai-type re-allocation of orders may be compared to seminal work by [HHK92] on a different but related problem, who required that for economic reason the utility should be a continuous functional of cumulative consumption with respect to the Lévy-Prokhorov metric d_{LP}, in order to satisfy the sensible property of intertemporal substitution for consumption. Recall for convenience of the reader the definition of d_{LP} in our context: for increasing càdlàg paths on $[0,\tilde{T}]$, $x, y : [0,\tilde{T}] \to \mathbb{R}$ with $x(0-) = y(0-)$ and $x(\tilde{T}) = y(\tilde{T})$,

$$d_{\mathrm{LP}}(x,y) := \inf\{\varepsilon > 0 \mid x(t) \le y((t+\varepsilon)\wedge\tilde{T})+\varepsilon,\ \ y(t) \le x((t+\varepsilon)\wedge\tilde{T})+\varepsilon\ \ \forall t \in [0,\tilde{T}]\}.$$

Corollary 5.2.10. *Let Θ be a predictable process with càdlàg paths defined on the time interval $[0,T]$ (with possible initial and terminal jumps) that is extended to the time interval $[-\varepsilon, T+\varepsilon]$ as in Remark 5.2.6. Consider the sequence of f.v. processes (Θ^n) where*

$$\Theta^n_t := n \int_{t-1/n}^{t} \Theta_s \,\mathrm{d}s, \quad t \ge 0, \tag{5.20}$$

for $n \in \mathbb{N}$ large enough, and let $L := L(\Theta), L^n := L(\Theta^n)$ be the proceeds processes from the respective trading. Then $L^n_t \to L_t$ at all continuity points $t \in [0, T+\varepsilon]$ of L as $n \to \infty$, in probability. In particular, for any bounded monotone strategy Θ the Borel measures $L^n(\mathrm{d}t;\omega)$ and $L(\mathrm{d}t;\omega)$ on $[0,T+\varepsilon]$ are finite (a.s.) and converge in the Lévy-Prokhorov metric $d_{LP}(L^n(\omega), L(\omega))$ in probability, i.e. for any $\eta > 0$,

$$\mathbb{P}\big[d_{LP}(L^n(\omega), L(\omega)) > \eta\big] \to 0 \quad \text{as } n \to \infty.$$

Proof. An application of Proposition 5.2.4 together with Theorem 5.2.7 gives

$$d_{M1}(L^n, L) \xrightarrow{\mathbb{P}} 0.$$

The first part of the claim now follows from the fact that convergence in M_1 implies local uniform convergence at continuity points of the limit, see [Whi02, Lemma 12.5.1]. The same property implies the claim about the Lévy-Prokhorov metric because convergence in this metric is equivalent to weak convergence of the associated measures which on the other hand is equivalent to convergence at all continuity points of the cumulative distribution function (together with the total mass). □

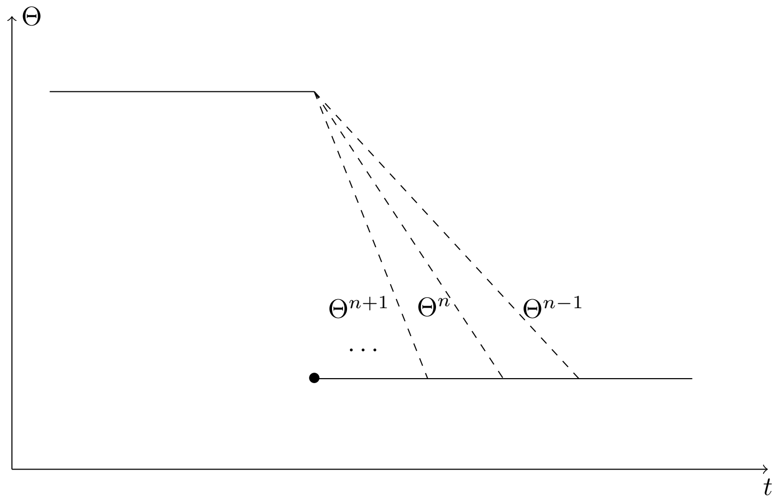


Figure 5.1: The Wong-Zakai approximation in (5.20) for a single jump process.

Note that the sequence (Θ^n) from Corollary 5.2.10 satisfies $\Theta^n \equiv \Theta_T$ on $[T+1/n, T+\varepsilon]$ for all $n \geq \lceil 1/\varepsilon \rceil$, i.e. the approximating strategies arrive at the position Θ_T, however by requiring a bit more time to execute. Based on the Wong-Zakai approximation sequence from (5.20), we next show that each semimartingale strategy on the time interval $[0, T]$ can be approximated by simple adapted strategies with uniformly small jumps that, however, again need slightly more time to be executed.

Proposition 5.2.11. *Let $(\Theta_t)_{t\in[0,T]}$ be a predictable process with càdlàg paths extended to the time interval $[0, T+\varepsilon]$ as in Remark 5.2.6. Then there exists a sequence $(\Theta^n_t)_{t\in[0,T+\varepsilon]}$ of simple predictable càdlàg processes with jumps of size not more than $1/n$ such that $d_{M1}(L(\Theta^n), L(\Theta)) \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$, where d_{M1} denotes the Skorokhod M_1 metric on $D([0, T+\varepsilon]; \mathbb{R})$. Moreover, if Θ is continuous, the same convergence holds true in the uniform metric on $[0, T]$ instead.*

Proof. Consider the Wong-Zakai approximation sequence $\widetilde{\Theta}^n$ from Corollary 5.2.10 for which $d_{M_1}(L(\widetilde{\Theta}^n), L(\Theta)) \xrightarrow{\mathbb{P}} 0$, where the Skorokhod M_1 topology is considered for the extended paths on time-horizon $[0, T+\varepsilon]$, with $n \geq \lceil 1/\varepsilon \rceil$ large enough. Now we approximate each (absolutely) continuous process $\widetilde{\Theta}^n$ by a sequence of simple processes as follows.

For $\delta > 0$, consider the sequence of stopping times with $\sigma^{\delta,n}_0 := 0$ and

$$\sigma^{\delta,n}_{k+1} := \inf\big\{t \bigm| t > \sigma^{\delta,n}_k \text{ and } |\widetilde{\Theta}^n_t - \widetilde{\Theta}^n_{\sigma^{\delta,n}_k}| \geq \delta\big\} \wedge (\sigma^{\delta,n}_k + 1/n) \quad \text{for } k \geq 0.$$

Note that $\sigma_k^{\delta,n}$ are predictable as hitting times of continuous processes and $\sigma_k^{\delta,n} \nearrow \infty$ as $k \to \infty$ because the process $\widetilde{\Theta}^n$ is continuous. When $\delta \to 0$, we have $\Theta^{\delta,n} \xrightarrow{ucp} \widetilde{\Theta}^n$ for

$$\Theta^{\delta,n} := \widetilde{\Theta}_0^n + \sum_{k=1}^{\infty} \big(\widetilde{\Theta}^n_{\sigma_k^{\delta,n}} - \widetilde{\Theta}^n_{\sigma_{k-1}^{\delta,n}}\big) \mathbb{1}_{[\![\sigma_k^{\delta,n},\infty[\![}.$$

Moreover, if for each integer $m \geq 1$ we define the (predictable) process $\Theta^{\delta,n,m}$ by

$$\Theta^{\delta,n,m} := \widetilde{\Theta}_0^n + \sum_{k=1}^{m} \big(\widetilde{\Theta}^n_{\sigma_k^{\delta,n}} - \widetilde{\Theta}^n_{\sigma_{k-1}^{\delta,n}}\big) \mathbb{1}_{[\![\sigma_k^{\delta,n},\infty[\![},$$

then for each fixed δ and n we have $\Theta^{\delta,n,m} \xrightarrow{ucp} \Theta^{\delta,n}$ when $m \to \infty$. Hence, we can choose $\delta = \delta(n)$ small enough and $m = m(n)$ big enough such that

$$d(\widetilde{\Theta}^n, \Theta^{\delta(n),n,m(n)}) < 2^{-n},$$

with $d(\cdot,\cdot)$ denoting a metric that metricizes ucp convergence (cf. e.g. [Pro05, p. 57]). Thus, $\Theta^n := \Theta^{\delta(n),n,m(n)}$ will be close to Θ in the Skorokhod M_1 topology, in probability, because the uniform topology is stronger than the M_1 topology.

Note that if Θ is already continuous, no intermediate Wong-Zakai approximation would be needed, and so we obtain uniform convergence in probability in that case. □

5.3 Case study: expectation constraints on the time to liquidation

In this section, we present a particular example in the framework of multiplicative impact $g(\overline{S}, Y) = f(Y)\overline{S}$, cf. Example 5.1.1, where the M_1 topology is key for identifying the (asymptotically realizable) proceeds) and thus extend the market model from finite variation controls to a larger class of trading strategies. This is key here, because it will turn out that any finite variation control is suboptimal.

Let us investigate an optimal liquidation problem for a variant of the price impact model which features *stochastic liquidity.* The singular control problem exhibits two interesting properties: It still permits an explicit description for the optimal strategy under a new constraint on the *expected* time to (complete) liquidation, but the optimal control is not of finite variation. So the set of admissible strategies needs to accommodate for infinite variation controls. As it is clear how to define the proceeds functional for (continuous) strategies of finite variation (cf. equation (5.4)), and we want (and need) to admit for jumps in the (optimal) control, the M_1 topology is a natural choice to extend the domain continuously. We consider no discounting or drift in the unaffected price process, letting $\overline{S}_t = \overline{S}_0 \mathcal{E}(\sigma W)_t$ with constant $\sigma > 0$. This martingale case will permit to apply convexity arguments in spirit of [PSS11] to construct an optimal control, see Theorem 5.3.2 below. In (2.2), the dynamics of market impact Y (called volume effect in [PSS11]) was a deterministic function of the large trader's strategy Θ. To model liquidity which is stochastic (e.g. by volume imbalances from other large 'noise' traders, cf. Remark 4.1.4), let the impact process Y^Θ solve

$$\mathrm{d}Y_t^\Theta = -\beta Y_t^\Theta \,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t + \mathrm{d}\Theta_t\,, \quad \text{with} \quad Y_{0-}^\Theta = Y_{0-} \in \mathbb{R} \text{ given}, \tag{5.21}$$

for constants $\beta, \hat{\sigma} > 0$ and a Brownian motion B that is independent of W. For the impact function $f \in C^3(\mathbb{R})$, giving the observed price by $S_t = f(Y_t)\overline{S}_t$, we require $f, f' > 0$ with $f(0) = 1$ and that $\lambda(y) := f'(y)/f(y)$ is bounded away from 0 and ∞, i.e. for constants $0 < \lambda_{\min} \leq \lambda_{\max}$ we have $\lambda_{\min} \leq \lambda(y) \leq \lambda_{\max}$ for all $y \in \mathbb{R}$, with bounded derivative λ'. Moreover, we assume that $k(y) := \frac{\hat{\sigma}^2}{2}\frac{f''(y)}{f(y)} - \beta - \beta y \frac{f'(y)}{f(y)}$ is strictly decreasing. An example satisfying these conditions is $f(y) = e^{\lambda y}$ with constant $\lambda > 0$. Let $F(x) := \int_{-\infty}^{x} f(y)\,\mathrm{d}y$, which is positive and of exponential growth due to the bounds on λ: $0 < F(x) \leq \big(e^{\lambda_{\min}} + e^{\lambda_{\max}}\big)/\lambda_{\min}$. The liquidation problem on infinite horizon *with* discounting and *without* intermediate buying in this model is solved in Theorem 4.2.1.

For our problem here, proceeds of general semimartingale strategies Θ should be

$$\begin{aligned} L_T(\Theta) = \int_0^T \overline{S}_t \psi(Y_{t-}^{\Theta})\,\mathrm{d}t + \overline{S}_0 F(Y_{0-}) - \overline{S}_T F(Y_T^{\Theta}) \\ + \int_0^T F(Y_{t-}^{\Theta})\,\mathrm{d}\overline{S}_t + \hat{\sigma}\int_0^T \overline{S}_t f(Y_{t-}^{\Theta})\,\mathrm{d}B_t\,, \end{aligned} \tag{5.22}$$

with $\psi(y) := -\beta y f(y) + \frac{\hat{\sigma}^2}{2} f'(y)$, because (5.22) is the continuous extension (in M_1 in probability, cf. Theorem 5.2.7 and see (5.15) with $g(x,y) = xf(y)$, $\beta_t = \hat{\sigma}^2$ and $\gamma_t = 0$) of the functional $L(\Theta^c) = -\int_0^T S_u\,\mathrm{d}\Theta_u^c$ from continuous f.v. Θ^c to semimartingales Θ that are *bounded in probability* on $[0,\infty)$:

Our goal is to maximize expected proceeds $\mathbb{E}[L_\infty(\Theta)]$ over some suitable set of admissible strategies that we specify now. From an application point of view, it makes sense to impose some bound on the time horizon within which liquidation is to be completed. Indeed, since our control objective here involves no discounting, one needs to restrict the horizon to get a non-trivial solution. Let some $\eta_{\max} \geq 0$ be given. A semimartingale Θ that is bounded in probability on $[0,\infty)$ will be called an *admissible strategy*, if

> there exists a stopping time τ with $\mathbb{E}[\tau] \leq \eta_{\max}$ such that $\Theta_t = \Theta_t \mathbb{1}_{t\leq\tau}$, with $\mathbb{E}[\tau \overline{S}_\tau] < \infty$, $\big(L_\tau(\Theta)\big)^- \in L^1(\mathbb{P})$ and such that the processes $\int_0^{\cdot\wedge\tau} \overline{S}_t F(Y_{t-}^{\Theta})\,\mathrm{d}W_t$, $\int_0^{\cdot\wedge\tau} \overline{S}_t f(Y_{t-}^{\Theta})\,\mathrm{d}B_t$, $\overline{S}_{\cdot\wedge\tau}$ and $(\overline{S}B)_{\cdot\wedge\tau}$ are uniformly integrable (UI) martingales.

The integrability conditions ensure $L_\tau(\Theta) \in L^1(\mathbb{P})$. Indeed, for admissible Θ it suffices to check $\big(\int_0^\tau \overline{S}_t \psi(Y_{t-}^{\Theta})\,\mathrm{d}t\big)^+ \in L^1(\mathbb{P})$. We will show in the proof of Theorem 5.3.2 that ψ attains a maximum $\psi(y^*)$. Thus we can bound $\int_0^\tau \overline{S}_t \psi(Y_{t-}^{\Theta})\,\mathrm{d}t$ from above by $\psi(y^*)\int_0^\tau \overline{S}_t\,\mathrm{d}t$, which is integrable by optional projection [DM82, Thm. VI.57] since $\mathbb{E}[\tau\overline{S}_\tau] < \infty$.

Let $\mathcal{A}_{\eta_{\max}}$ be the set of all admissible strategies with given fixed initial value Θ_{0-}, where $|\Theta_{0-}|$ is the number of shares to be liquidated (sold) if $\Theta_{0-} > 0$, resp. acquired (bought) if $\Theta_{0-} < 0$. The definition of $\mathcal{A}_{\eta_{\max}}$ involves several technical conditions. But the set $\mathcal{A}_{\eta_{\max}}$ is not small, for instance it clearly contains all strategies of finite variation which liquidate until some bounded stopping times τ with $\mathbb{E}[\tau] \leq \eta_{\max}$, and also strategies of infinite variation (see below). Note that intermediate short selling is permitted, and that $\mathcal{A}_0$ contains only the trivial strategy to sell (resp. buy) everything immediately.

We will show that optimal strategies are *impact fixing*. For $\tilde{\Upsilon}, \Upsilon \in \mathbb{R}$ an *impact fixing strategy* $\Theta = \Theta^{\tilde{\Upsilon},\Upsilon}$ is a strategy with liquidation time τ (i.e. $\Theta_t = 0$ for $t \geq \tau$), such that $Y = Y^{\Theta^{\tilde{\Upsilon},\Upsilon}}$ satisfies $Y_t = \tilde{\Upsilon}$ on $[\![0,\tau[\![$ and $Y_\tau = \Upsilon$. More precisely, $\Theta_0 = \Theta_{0-} + \tilde{\Upsilon} - Y_{0-}$, $\mathrm{d}\Theta_t = \beta\tilde{\Upsilon}\,\mathrm{d}t - \hat{\sigma}\,\mathrm{d}B_t$ on $]\!]0,\tau[\![$ until $\tau = \tau^{\tilde{\Upsilon},\Upsilon} := \inf\{t > 0 \mid \Theta_{t-} = \tilde{\Upsilon} - \Upsilon\}$, with final block trade of size $\Delta\Theta_\tau = -\Theta_{\tau-} = \Upsilon - \tilde{\Upsilon}$ and $\Theta = 0$ on $[\![\tau,\infty[\![$. We have the following properties of impact fixing strategies.

Lemma 5.3.1 (Admissibility of impact fixing strategies)**.** *The liquidation time* $\tau = \tau^{\tilde{\Upsilon},\Upsilon}$ *of an impact fixing strategy* $\Theta^{\tilde{\Upsilon},\Upsilon}$ *has expectation* $\mathbb{E}[\tau] = (Y_{0-} - \Theta_{0-} - \Upsilon)/(\beta\tilde{\Upsilon})$ *if* $(Y_{0-} - \Theta_{0-} - \Upsilon)\tilde{\Upsilon} > 0$*, and* $\mathbb{E}[\tau] = 0$ *if* $\Upsilon = Y_{0-} - \Theta_{0-}$*, otherwise* $\mathbb{E}[\tau] = \infty$*. Moreover, if* $\mathbb{E}[\tau^{\tilde{\Upsilon},\Upsilon}] \leq \eta_{\max}$ *then* $\Theta^{\tilde{\Upsilon},\Upsilon} \in \mathcal{A}_{\eta_{\max}}$*.*

Proof. By [BS02, Ch. 2, Sect. 2, eq. (2.0.2) on p. 295], the law of the hitting time H_z of level z by a Brownian motion with drift μ starting in x is for $t \in (0,\infty)$ given by $\mathbb{P}_x[H_z \in \mathrm{d}t] = h^\mu(t, z-x)\,\mathrm{d}t$ with $h^\mu(t,x) := \frac{|x|}{\sqrt{2\pi}t^{3/2}}\exp\big(-\frac{(x-\mu t)^2}{2t}\big)$ and $\mathbb{P}_x[H_z = \infty] = 1 - \exp\big(\mu(z-x) - |\mu|\cdot|z-x|\big)$. With $\mu = \beta\tilde{\Upsilon}/\rho$, $x = (\Theta_{0-} + \tilde{\Upsilon} - Y_{0-})/\rho$ and $z = (\tilde{\Upsilon} - \Upsilon)/\rho$ we obtain the stated terms for $\mathbb{E}[\tau] = \mathbb{E}_x[H_z]$.

Now, let $\tilde{\Upsilon}$, Υ be such that $\mathbb{E}[\tau] \leq \eta_{\max}$. Independence of τ and $\overline{S}$ gives $\mathbb{E}[\overline{S}_\tau] = \overline{S}_0$ and $\mathbb{E}[\tau\overline{S}_\tau] < \infty$. We have $\int_0^\tau \overline{S}_t f(Y_t)\,\mathrm{d}B_t = f(\tilde{\Upsilon})M_\tau$ for $M_T := \int_0^{T\wedge\tau}\overline{S}_t\,\mathrm{d}B_t$ and $\int_0^\tau \overline{S}_t F(Y_t)\,\mathrm{d}W_t = F(\tilde{\Upsilon})\sigma^{-2}\overline{S}_\tau$. Note that $[M]_\tau = \sigma^{-2}[\overline{S}]_\tau$. We will show that M, $\overline{S}_{\cdot\wedge\tau}$ and $(\overline{S}B)_{\cdot\wedge\tau}$ are in $\mathcal{H}^1$ and hence UI martingales. By Burkholder-Davis-Gundy [Pro05, Thm. IV.4.48], there exists $C > 0$ with $\mathbb{E}\big[[\overline{S}]_\tau^{1/2}\big] \leq C\mathbb{E}[\sup_{u\leq\tau}|\overline{S}_u|] = C\mathbb{E}[\exp(\sigma X_\tau)]$ with $X_t := \sup_{u\leq t}(W_u - \frac{\sigma}{2}u)$. Using $\{X_t > z\} = \{H_z < t\}$ for $z, t \geq 0$ with starting point $X_0 = 0$ and drift $\mu = -\sigma/2$ we first obtain

$$
\begin{aligned}
\mathbb{E}[\exp(\sigma X_t)] &= \int_{[0,\infty]} e^{\sigma x}\mathbb{P}[X_t \in \mathrm{d}x] = \int_{[0,\infty]} e^{\sigma x}\,\mathrm{d}\big(1 - \mathbb{P}[X_t > x]\big)_x \\
&= -\int_{[0,\infty]} e^{\sigma x}\,\mathrm{d}\big(\mathbb{P}[X_t > x]\big)_x = -\int_{[0,\infty]} e^{\sigma x}\,\mathrm{d}\big(\mathbb{P}[H_x < t]\big)_x .
\end{aligned}
$$

Since $\mathbb{P}[H_\infty < t] = 0$ we can approximate the Riemann-Stieltjes integral and apply integration by parts twice to get

$$
\mathbb{E}[\exp(\sigma X_t)] = -\lim_{\varepsilon\searrow 0}\int_0^t\int_\varepsilon^{1/\varepsilon} e^{\sigma x}h_x^{-\sigma/2}(u,x)\,\mathrm{d}x\,\mathrm{d}u = -\int_0^t\int_0^\infty e^{\sigma x}h_x^{-\sigma/2}(u,x)\,\mathrm{d}x\,\mathrm{d}u
$$

with $\quad h_x^{-\sigma/2}(t,x) = \dfrac{\mathrm{d}}{\mathrm{d}x}h^{-\sigma/2}(t,x) = -\dfrac{x^2 - t + \frac{\sigma}{2}xt}{\sqrt{2\pi}t^{5/2}}\exp\Big(-\dfrac{(x+\frac{\sigma}{2}t)^2}{2t}\Big).$

So we have $e^{\sigma x}h_x^{-\sigma/2}(t,x) = h_x^{\sigma/2}(t,x) - \sigma h^{\sigma/2}(t,x)$. The contribution from the first summand of the integrand $h_x^{\sigma/2}(t,x) - \sigma h^{\sigma/2}(t,x)$ is zero, since $h^{\sigma/2}(t,x) \to 0$ for $x \to \infty$ and for $x \to 0$. Hence, $\mathbb{E}[\exp(\sigma X_t)]$ equals

$$
\sigma\int_0^t\int_0^\infty h^{\sigma/2}(u,x)\,\mathrm{d}x\,\mathrm{d}u = \sigma\int_0^t\bigg(\frac{\exp\big(-\frac{\sigma^2}{8}u\big)}{\sqrt{2\pi u}} - \frac{\sigma}{2} + \frac{\sigma}{2}\varphi\big(\tfrac{\sigma}{2}\sqrt{u}\big)\bigg)\,\mathrm{d}u
$$

$$= 2\varphi\big(\tfrac{\sigma}{2}\sqrt{t}\big) - 1 + \tfrac{\sigma^2}{2}t\varphi\big(\tfrac{\sigma}{2}\sqrt{t}\big) - \tfrac{\sigma^2}{2}t + \frac{\sigma\sqrt{t}}{\sqrt{2\pi}}\exp\big(-\tfrac{\sigma^2}{8}t\big) \le 1 + \frac{\sigma\sqrt{t}}{\sqrt{2\pi}}\,,$$

where $\varphi(x) = \int_{-\infty}^{x} e^{-z^2/2}\,\mathrm{d}z/\sqrt{2\pi}$. So by independence of X and τ

$$\mathbb{E}[\exp(\sigma X_\tau)] = \mathbb{E}[(t \mapsto \mathbb{E}[e^{\sigma X_t}])(\tau)] \le \mathbb{E}\Big[1 + \frac{\sigma}{\sqrt{2\pi}}\sqrt{\tau}\Big] \le 1 + \frac{\sigma}{\sqrt{2\pi}}(1 + \mathbb{E}[\tau]) < \infty\,.$$

Moreover, $[\overline{S}B]_\tau = \tau[\overline{S}]_\tau$ by independence of $\overline{S}$ and B, so we can bound $\mathbb{E}\big[[\overline{S}B]_\tau^{1/2}\big]$ by $\mathbb{E}\big[\sqrt{\tau}[\overline{S}]_\tau^{1/2}\big] = \mathbb{E}\Big[\sqrt{t}\mathbb{E}\big[[\overline{S}]_t^{1/2}\big]\Big|_{t=\tau}\Big] \le C\mathbb{E}\Big[\sqrt{t}\mathbb{E}[\exp(\sigma X_t)]\Big|_{t=\tau}\Big] \le C\mathbb{E}[\sqrt{\tau} + \tfrac{\sigma}{\sqrt{2\pi}}\tau] < \infty$. Thus, $(\overline{S}B)_{\cdot\wedge\tau}$ is in $\mathcal{H}^1$ and hence a UI martingale.

Finally, $\big(L_\tau(\Theta)\big)^- \in L^1(\mathbb{P})$ follows from $\int_0^\tau \overline{S}_t g(Y_{t-}^\Theta)\,\mathrm{d}t = g(\tilde{\Upsilon})\int_0^\tau \overline{S}_t\,\mathrm{d}t$, which is integrable by optional projection [DM82, Thm. VI.57] since $\mathbb{E}[\tau\overline{S}_\tau] < \infty$, and integrability of $\overline{S}_\tau F(Y_\tau^\Theta) = \overline{S}_\tau F(\Upsilon)$. □

Using convexity arguments we construct the solution for the optimization problem in

Theorem 5.3.2. *For every $\eta_{\max} \in [0,\infty)$ there exist $\hat{\eta} \in [0,\eta_{\max}]$ and $\tilde{\Upsilon}$, $\Upsilon \in \mathbb{R}$ such that the associated impact fixing strategy $\hat{\Theta} := \Theta^{\tilde{\Upsilon},\Upsilon}$ generates maximal expected proceeds in expected time $\mathbb{E}[\tau^{\tilde{\Upsilon},\Upsilon}] = \hat{\eta}$ among all admissible strategies, i.e.*

$$\mathbb{E}[L_\infty(\hat{\Theta})] = \max\big\{\mathbb{E}[L_\infty(\Theta)] \;\big|\; \Theta \in \mathcal{A}_{\eta_{\max}}\big\}\,.$$

Moreover, if $f(y) = e^{\lambda y}$ with $\lambda \in (0,\infty)$, then we have $\hat{\eta} = \eta_{\max}$ and the optimal strategy is unique.

The proof will also show that optimal strategies have to be *impact fixing*. In particular, any non-trivial admissible strategy of finite variation is suboptimal.

Proof. Since f'/f and $(f'/f)'$ are bounded, then f''/f is also bounded and hence there is a unique $y^* \in \mathbb{R}$ with $k(y^*) = 0$. So ψ is strictly increasing on $(-\infty, y^*)$ and decreasing on (y^*,∞), since $\psi'(y) = f(y)k(y)$. Note that ψ is strictly concave on $[y^*,\infty)$ and $\psi(y) > 0$ for $y < 0$. Hence, the concave hull of ψ is

$$\hat{\psi}(y) := \inf\{\ell(y) \mid \ell \text{ is an affine function with } \ell(x) \ge \psi(x)\ \forall x\} = \psi(y \vee y^*)\,.$$

Let $\Theta \in \mathcal{A}_{\eta_{\max}}$ with liquidation time τ. Denote by $\mathbb{Q}$ the measure with $\mathrm{d}\mathbb{Q} = \big(\overline{S}_\tau/\overline{S}_0\big)\,\mathrm{d}\mathbb{P}$. Then by optional projection, as in [DM82, Thm. VI.57], we obtain (taking w.l.o.g. $\overline{S}_0 = 1$):

$$\begin{aligned}
\mathbb{E}[L_\infty] = \mathbb{E}[L_\tau] &= \mathbb{E}\Big[\int_0^\tau \overline{S}_t\psi(Y_t)\,\mathrm{d}t\Big] + F(Y_{0-}) - \mathbb{E}\big[\overline{S}_\tau F(Y_\tau)\big] \\
&= F(Y_{0-}) + \mathbb{E}_\mathbb{Q}\Big[\int_0^\tau \psi(Y_t)\,\mathrm{d}t\Big] - \mathbb{E}_\mathbb{Q}\big[F(Y_\tau)\big] \\
&= F(Y_{0-}) + \int_{\Omega\times[0,\infty)} \psi(Y_t(\omega))\mu(\mathrm{d}\omega,\mathrm{d}t) - \mathbb{E}_\mathbb{Q}\big[F(Y_\tau)\big]\,, \qquad (5.23)
\end{aligned}$$

for the finite measure μ given by $\mu(A\times B) := \int_A \int_0^{\tau(\omega)} \mathbb{1}_B(t)\,\mathrm{d}t\mathbb{Q}[\mathrm{d}\omega]$ with total mass $\mu(\Omega,[0,\infty)) = \mathbb{E}_{\mathbb{Q}}[\tau] = \mathbb{E}[\tau\overline{S}_\tau] < \infty$. For $\tau \neq 0$, Jensen's inequality for $\hat{\psi}$ and F gives

$$\mathbb{E}[L_\infty] \le F(Y_{0-}) + \int_{\Omega\times[0,\infty)} \hat{\psi}(Y_t(\omega))\mu(\mathrm{d}\omega,\mathrm{d}t) - \mathbb{E}_{\mathbb{Q}}\big[F(Y_\tau)\big] \tag{5.24}$$

$$\le F(Y_{0-}) + \mathbb{E}_{\mathbb{Q}}[\tau]\hat{\psi}\bigg(\frac{1}{\mathbb{E}_{\mathbb{Q}}[\tau]}\int_{\Omega\times[0,\infty)} Y_t(\omega)\mu(\mathrm{d}\omega,\mathrm{d}t)\bigg) - \mathbb{E}_{\mathbb{Q}}\big[F(Y_\tau)\big] \tag{5.25}$$

$$= F(Y_{0-}) + \mathbb{E}_{\mathbb{Q}}[\tau]\hat{\psi}\bigg(\frac{1}{\beta\mathbb{E}_{\mathbb{Q}}[\tau]}\mathbb{E}_{\mathbb{Q}}\Big[\int_0^\tau \beta Y_t\,\mathrm{d}t\Big]\bigg) - \mathbb{E}_{\mathbb{Q}}\big[F(Y_\tau)\big] \tag{5.26}$$

$$= F(Y_{0-}) + \mathbb{E}_{\mathbb{Q}}[\tau]\hat{\psi}\bigg(\frac{Y_{0-}-\Theta_{0-}-\mathbb{E}_{\mathbb{Q}}[Y_\tau]}{\beta\mathbb{E}_{\mathbb{Q}}[\tau]}\bigg) - \mathbb{E}_{\mathbb{Q}}\big[F(Y_\tau)\big] \tag{5.27}$$

$$\le F(Y_{0-}) + \mathbb{E}_{\mathbb{Q}}[\tau]\hat{\psi}\bigg(\frac{Y_{0-}-\Theta_{0-}-\mathbb{E}_{\mathbb{Q}}[Y_\tau]}{\beta\mathbb{E}_{\mathbb{Q}}[\tau]}\bigg) - F\big(\mathbb{E}_{\mathbb{Q}}[Y_\tau]\big) \tag{5.28}$$

$$= F(Y_{0-}) + \hat{\Psi}\big(\mathbb{E}_{\mathbb{Q}}[\tau],\mathbb{E}_{\mathbb{Q}}[Y_\tau]\big)\,, \tag{5.29}$$

for $\hat{\Psi}(\eta,\Upsilon) := \eta\hat{\psi}\big(\frac{Y_{0-}-\Theta_{0-}-\Upsilon}{\beta\eta}\big) - F(\Upsilon)$ when $\eta > 0$, while for $\tau = 0$ we get that $\mathbb{E}[L_\infty]$ is given by (5.29) with $\hat{\Psi}(0,\Upsilon) := -F(\Upsilon)$. The step from (5.26) to (5.27) uses that $\mathbb{E}[\overline{S}_\tau B_\tau] = 0$, due to $(\overline{S}B)_{\cdot\wedge\tau}$ being UI, and $\int_0^t \beta Y_s\,\mathrm{d}s = \hat{\sigma}B_t + \Theta_t - \Theta_{0-} - Y_t + Y_{0-}$. Since F is strictly convex, we obtain equality in (5.28) if and only if Y_τ is concentrated at a point $\Upsilon \in \mathbb{R}$ $\mathbb{P}$-a.s. At (5.25) we obtain equality if and only if either $Y_t \in (-\infty, y^*]$ μ-a.e. (where $\hat{\psi}$ is affine) or Y_t is concentrated at a point $\tilde{\Upsilon} \in \mathbb{R}$ μ-a.e. Equality at (5.24) can only happen if $Y \ge y^*$ μ-a.e. Hence, we only get equality

$$\mathbb{E}[L_\infty] = F(Y_{0-}) + \hat{\Psi}\big(\mathbb{E}_{\mathbb{Q}}[\tau],\mathbb{E}_{\mathbb{Q}}[Y_\tau]\big)$$

for *impact fixing strategies* $\Theta = \Theta^{\tilde{\Upsilon},\Upsilon}$ with $\tilde{\Upsilon} \ge y^*$, where $\mathbb{E}[L_\tau] = F(Y_{0-}) + \hat{\Psi}(\mathbb{E}[\tau],\Upsilon)$. Since y^* is the largest maximizer of $\hat{\psi}$, $\lim_{y\to\infty}\psi'(y) = -\infty$ and F is strictly increasing, $\hat{\Psi}(\eta,\cdot)$ has a unique maximizer $\hat{e}(\eta) \in (-\infty, e^*)$ where $e^* = e^*(\eta) = Y_{0-} - \Theta_{0-} - \beta\eta y^*$ for $\eta > 0$ and $\hat{e}(0) = e^*(0) = Y_{0-} - \Theta_{0-}$. Because $\hat{y}(\eta) := \big(Y_{0-} - \Theta_{0-} - \hat{e}(\eta)\big)/(\beta\eta) > y^*$, the impact fixing strategy $\Theta^{\hat{y}(\eta),\hat{e}(\eta)}$ has expected time to liquidation η (cf. Lemma 5.3.1) and generates $F(Y_{0-}) + \hat{\Psi}\big(\eta,\hat{e}(\eta)\big)$ expected proceeds that are optimal among all impact fixing strategies with expected time to liquidation η.

Note that $\hat{e}(\eta)$ is continuous in $\eta \in (0,+\infty)$ by the implicit function theorem; recall that $\hat{e}(\eta)$ solves $0 = \hat{\Psi}_\Upsilon(\eta,\hat{e}(\eta)) = -\hat{\psi}'(\hat{y}(\eta))/\beta - f(\hat{e}(\eta))$, and $\hat{\Psi}_{\Upsilon\Upsilon}(\eta,\Upsilon) < 0$ for $\Upsilon < e^*(\eta)$. Moreover, $\hat{e}(\eta) \to \hat{e}(0)$ when $\eta \to 0$, otherwise $\hat{y}(\eta) \to +\infty$ for a subsequence giving $-\hat{\psi}'(\hat{y}(\eta))/\beta = -(fk)(\hat{y}(\eta))/\beta \to +\infty$ and therefore also $f(\hat{e}(\eta)) \to +\infty$, which would contradict $\limsup_{\eta\to 0}\hat{e}(\eta) \le \lim_{\eta\to 0} e^*(\eta) = Y_{0-} - \Theta_{0-}$.

In particular, the contradiction argument above shows that $\hat{y}(\eta)$ is contained in a compact set for small η. As a consequence, $\hat{\Psi}(\eta,\hat{e}(\eta)) = \eta\hat{\psi}(\hat{y}(\eta)) - F(\hat{e}(\eta)) \to \hat{\Psi}(0,\hat{e}(0))$ as $\eta \to 0$, i.e. the map $\eta \mapsto \hat{\Psi}(\eta,\hat{e}(\eta))$ is continuous on $[0,+\infty)$. Hence, it attains a maximizer $\hat{\eta} \in [0,\eta_{\max}]$ whose associated impact fixing strategy $\hat{\Theta} = \Theta^{\hat{y}(\hat{\eta}),\hat{e}(\hat{\eta})}$ generates maximal expected proceeds in expected time $\mathbb{E}[\tau^{\hat{y}(\hat{\eta}),\hat{e}(\hat{\eta})}] = \hat{\eta}$ among all admissible strategies $\mathcal{A}_{\eta_{\max}}$.

If $f(y) = e^{\lambda y}$ with $\lambda \in (0,\infty)$, one can check by direct calculations that $\hat{\Psi}_\eta(\eta,\Upsilon) > 0$ for $\eta > 0$, $\Upsilon \in \mathbb{R}$, and thus the map $\eta \mapsto \hat{\Psi}(\eta,\hat{e}(\eta))$ is strictly increasing, because using

$\frac{\mathrm{d}}{\mathrm{d}\eta}\hat{\Psi}(\eta,\hat{e}(\eta)) = \hat{\Psi}_\eta(\eta,\hat{e}(\eta)) + \hat{\Psi}_\Upsilon(\eta,\hat{e}(\eta))\hat{e}'(\eta) = \hat{\Psi}_\eta(\eta,\hat{e}(\eta))$. So $\hat{\eta} = \eta_{\max}$ is its unique maximizer in $[0,\eta_{\max}]$ and hence the optimal strategy is unique. □

5.4 Auxiliary results

The next proposition collects known continuity properties of the solution map $\Theta \mapsto Y^\Theta$ on $D([0,T];\mathbb{R})$ from (5.2), with the presentation being adapted to our setup.

Proposition 5.4.1. *Assume that h is Lipschitz continuous. Then the solution map $D([0,T];\mathbb{R}) \mapsto D([0,T];\mathbb{R})$, with $\Theta \mapsto Y^\Theta$ from (5.2), is defined pathwise. The map is continuous when the space $D([0,T];\mathbb{R})$ is endowed with either the uniform topology or the Skorokhod J_1 or M_1 topology. Moreover, if Θ is an adapted càdlàg process, then the process Y^Θ is also adapted.*

Proof. By continuity of N, the pathwise defined map $\Theta \mapsto \Theta + N$ on $D([0,T];\mathbb{R})$ is continuous w.r.t. the uniform and Skorokhod M_1 and J_1 topologies. Now, since $Y^{N,\Theta} = Y^{0,N+\Theta}$, it suffices to show that the deterministic impact solution map $\Theta \mapsto Y^{0,\Theta}$ is pathwise defined and continuous. The proof in the case of the uniform topology and the Skorokhod J_1 topology is given in [PTW07, proof of Thm. 4.1]. For the M_1 topology, cf. [PW10, Thm. 1.1]. That $Y^{0,\Theta}$ is adapted follows from the (pathwise) construction of $Y^{0,\Theta}$ as the (a.s.) limit (in the uniform topology) of adapted processes, the solution processes for a sequence of piecewise-constant controls Θ^n approximating uniformly Θ, cf. [PTW07, proof of Thm. 4.1]. □

In general, we may have $a_n \to a$ and $b_n \to b$ in $D([0,T])$ endowed with J_1 (or M_1), and yet $a_n + b_n \not\to a + b$ when a and b have a common jump time. However, in special cases like in what follows, this does not happen.

Lemma 5.4.2 (Allowed cancellation of jumps for J_1)**.** *Let $a_n \to a_0$ and $b_n \to b_0$ in $(D([0,T]),J_1)$ with the following property: for every $n \geq 0$ and every $t \in (0,T)$*

$$\Delta a_n(t) \neq 0 \quad \textit{implies} \quad \Delta b_n(t) = -\Delta a_n(t).$$

Then $a_n + b_n \to a_0 + b_0$ in $(D([0,T]),J_1)$.

Proof. By [JS03, Prop. VI.2.2, a] it suffices to check that for every $t \in (0,T)$ there exists a sequence $t_n \to t$ such that $\Delta a_n(t_n) \to \Delta a_0(t)$ and $\Delta b_n(t_n) \to \Delta b_0(t)$.

Let $t \in (0,T)$ be arbitrary and first suppose that $\Delta a_0(t) \neq 0$. Then [JS03, Prop. VI.2.1, a] implies the existence of a sequence $t_n \to t$ such that $\Delta a_n(t_n) \to \Delta a_0(t)$. Thus, our assumption on the sequence (b_n) gives $\Delta b_n(t_n) \to \Delta b_0(t)$. For the case $\Delta a_0(t) = 0$, let $t_n \to t$ be such that $\Delta b_n(t_n) \to \Delta b_0(t)$. By [JS03, Prop. VI.2.1, b.5] we conclude that $\Delta a_n(t_n) \to \Delta a_0(t)$ as well, finishing the proof. □

Let us note that the conclusion of Lemma 5.4.2 does not hold for the M_1 topology. Consider for example $a_0 = \mathbb{1}_{[1,\infty)}$ with approximating sequence $a_n(t) := n\int_t^{t+1/n} a_0(s)\,\mathrm{d}s$ and $b_0 = 1 - a_0$ with approximating sequence $b_n(t) := n\int_{t-1/n}^t b_0(s)\,\mathrm{d}s$. Thus we need the following refined statement.

Lemma 5.4.3 (Allowed cancellation of jumps for M_1)**.** *Let $a_n \to a_0$ in $(D([0,T]),\|\cdot\|_\infty)$ and $b_n \to b_0$ in $(D([0,T]),M_1)$ with the following property: $t \in Disc(a_0)$ implies $b_n \to b_0$ locally uniformly in a neighborhood of t. Then $a_n + b_n \to a_0 + b_0$ in $(D([0,T]),M_1)$.*

Proof. We prove the following claim that suffices to deduce M_1-convergence of $a_n + b_n$: For any $t \in [0,T]$ and $\varepsilon > 0$ there are $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$w_s(a_n + b_n, t, \delta) \le w_s(a_n, t, \delta) + w_s(b_n, t, \delta) + \varepsilon \quad \text{for all } n \ge n_0, \tag{5.30}$$

where w_s is the M_1 oscillation function, see [Whi02, Chap. 12, eq. (4.4)]. Indeed, if (5.30) holds, then the second condition in [Whi02, Thm. 12.5.1(v)] would hold, while the first condition there holds because of local uniform convergence at points of continuity of $a_0 + b_0$: Either there is cancellation of jumps and thus local uniform convergence by our assumption, or both paths do not jump which still gives local uniform convergence because M_1-convergence implies such at continuity points of the limit.

To check (5.30), we have $\lim_{\delta \downarrow 0} \limsup_{n\to\infty} v(a_n, a_0, t, \delta) = 0$ at points $t \in [0,T]$ with $\Delta a_0(t) = 0$, where for $x_1, x_2 \in D([0,T])$

$$v(x_1, x_2, t, \delta) := \sup_{0\vee(t-\delta)\le t_1, t_2 \le (t+\delta)\wedge T} |x_1(t_1) - x_2(t_2)|,$$

see [Whi02, Thm. 12.4.1], which implies (5.30) for small δ and large n. Now if $t \in \mathrm{Disc}(a_0)$, $a_n \to a_0$ and $b_n \to b_0$ locally uniformly in a neighborhood of t which implies that for small δ and large n

$$w_s(a_n + b_n, t, \delta) \le w_s(a_0 + b_0, t, \delta) + \varepsilon/2.$$

Because $a_0 + b_0 \in D([0,T])$, we can make $w_s(a_0 + b_0, t, \delta)$ smaller than $\varepsilon/2$, which finishes the proof. □

Lemma 5.4.4 (Uniform convergence of jump term)**.** *Let $a, b_n, b \in D([0,T])$ be such that $[a]^d_T := \sum_{t\le T:\Delta a(t)\neq 0} |\Delta a(t)|^2 < \infty$, b_n are uniformly bounded and at every jump time $t \in [0,T]$ of a, $\Delta a(t) \neq 0$, we have pointwise convergence $b_n(t) \to b(t)$. Let $G \in C^2$ such that $y \mapsto G_{xx}(x,y)$ is Lipschitz continuous on compacts. Then the sum*

$$J(a, b_n)_t := \sum_{\substack{u\le t \\ \Delta a(t)\neq 0}} G\big(a(t), b_n(t)\big) - G\big(a(t-), b_n(t)\big) - G_x\big(a(t-), b_n(t)\big)\Delta a(t)$$

converges uniformly for $t \in [0,T]$ to $J(a,b)_t$, as $n \to \infty$.

Proof. Since a, $[a]^d$, b_n and b are bounded on $[0,T]$ by a constant $C \in \mathbb{R}$, we can assume w.l.o.g. that G_{xx} is globally Lipschitz in y with Lipschitz constant L. Hence $J(a, b_n)_t < \infty$ by Taylor's theorem. Let $H(x, \Delta x, y) := G(x + \Delta x, y) - G(x,y) - G_x(x,y)\Delta x$ and denote by $\tilde{J}^{n,\pm}$ the increasing and decreasing components of $J(a, b_n) - J(a,b)$, respectively, i.e.

$$\tilde{J}^{n,+}_t := \sum_{\substack{u\le t \\ \tilde{H}(\dots)>0}} \tilde{H}\big(a(u-), \Delta a(u), b_n(u), b(u)\big), \quad \tilde{J}^{n,-}_t := \sum_{\substack{u\le t \\ \tilde{H}(\dots)<0}} \tilde{H}\big(a(u-), \Delta a(u), b_n(u), b(u)\big),$$

for $\tilde{H}(x, \Delta x, y, z) := H(x, \Delta x, y) - H(x, \Delta x, z)$. Moreover, take any enumeration of the jump times of a, $\{t_k \mid k \in \mathbb{N}\} = \{t \mid \Delta a(t) \neq 0\}$, and arbitrary $\varepsilon > 0$. Since $[a]^d < \infty$, there exists $K \in \mathbb{N}$ such that $\sum_{k>K} |\Delta a(t_k)|^2 < \varepsilon/(2CL)$. Moreover, we have $|\tilde{H}(x, \Delta x, y, z)| \le \frac{1}{2}|\Delta x|^2 L |y - z|$ and thus

$$|\tilde{J}^{n,\pm}_T| \le \frac{L}{2} \sum_{k=1}^{\infty} |\Delta a(t_k)|^2 |b_n(t_k) - b(t_k)| < \frac{\varepsilon}{2} + \frac{L}{2}\Big(\max_{1\le k\le K} |b_n(t_k) - b(t_k)|\Big) \sum_{k=1}^{K} |\Delta a(t_k)|^2 .$$

By pointwise convergence $b_n(t_k) \to b(t_k)$ at all t_k, there exists $N \in \mathbb{N}$ such that for all $k = 1, \dots, K$ and $n \geq N$ we have $|b_n(t_k) - b(t_k)| < \varepsilon/(L[a]^d_T)$ and therefore $|\tilde{J}^{n,\pm}_T| < \varepsilon$ for $n \geq N$. Hence $J^{n,\pm}_T \to 0$ as $n \to \infty$.

Since $J^{n,\pm}$ are monotone and do not cross zero, we have $\sup_{0 \leq t \leq T} |\tilde{J}^{n,\pm}_t| = |\tilde{J}^{n,\pm}_T|$ and therefore uniform convergence $\tilde{J}^{n,\pm} \to 0$ on $[0, T]$. So in particular $J(a, b_n)$ converges to $J(a, b)$, uniformly on $[0, T]$. □

6 Proportional bid-ask spreads in optimal trading – a double obstacle problem

Building on Chapter 4, we now introduce a bid-ask spread through proportional transaction costs in our market impact model. We consider finite variation strategies Θ. The fundamental ask price $\overline{S}_t$ for selling, that prevails in absence of a large trader, is as in (4.1) a geometric Brownian motion with drift, $\overline{S} = \overline{S}_0 e^{\mu t}\mathcal{E}(\sigma W)_t$. As in Chapter 4, the volume effect process Y given a strategy Θ is a controlled Ornstein-Uhlenbeck process,

$$\mathrm{d}Y_t^{\Theta} = -\beta Y_t^{\Theta}\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t + \mathrm{d}\Theta_t\,, \quad Y_{0-} = y, \tag{6.1}$$

with resilience speed $\beta > 0$. The Brownian motions B and W have correlation $\rho \in [-1, 1]$. In our model for this chapter, bid and ask prices differ by a proportional transaction cost factor $\kappa > 1$, with discounted bid price $e^{-\gamma t} f(Y_t^{\Theta})\overline{S}_t$ and discounted ask price $e^{-\gamma t}\kappa f(Y_t^{\Theta})\overline{S}_t$. Let $\Theta_t = \Theta_t^+ - \Theta_t^-$ with cumulative numbers of assets bought (Θ_t^+) and sold (Θ_t^-) until time t. Then γ-discounted proceeds from trading according to non-decreasing processes Θ^+, Θ^- are

$$\begin{aligned} L_t(\Theta^+, \Theta^-) := &- \int_0^t e^{-\gamma u}\kappa f(Y_u^{\Theta})\overline{S}_u\,\mathrm{d}\Theta_u^{c,+} - \sum_{\substack{0\le u\le t\\ \Delta\Theta_u^+\neq 0}} e^{-\gamma u}\overline{S}_u \int_0^{\Delta\Theta_u^+} \kappa f(Y_{u-}^{\Theta} + x)\,\mathrm{d}x \\ &+ \int_0^t e^{-\gamma u} f(Y_u^{\Theta})\overline{S}_u\,\mathrm{d}\Theta_u^{c,-} + \sum_{\substack{0\le u\le t\\ \Delta\Theta_u^-\neq 0}} e^{-\gamma u}\overline{S}_u \int_0^{\Delta\Theta_u^-} f(Y_{u-}^{\Theta} - x)\,\mathrm{d}x, \end{aligned} \tag{6.2}$$

with decompositions $\Theta_t^{\pm} = \Theta_t^{c,\pm} + \sum_{u\le t}\Delta\Theta_u^{\pm}$ into continuous and pure jump parts.

In Section 6.1, we formulate the optimal liquidation problem with non-monotone strategies. Finite variation optimization with different costs for the increasing and decreasing parts of a strategy is also called a *reversible* investment problem, see e.g. [KW01, DAF14, FP14]. In these articles, the price or cost for infinitesimal buying (resp. selling) depends on time only, differently from our setup where the price depends on the controlled diffusion (as well as time and $\overline{S}$). Our objective has the form of a generalized *finite fuel* problem (see e.g. [KS86]; the term *fuel* here is due to [BC67] and originally indicates a monotone control). Through a heuristic derivation based on smooth pasting techniques, we illustrate in Section 6.1.3 that the optimal strategy should consist of two reflection local times which keep the controlled diffusion Y inside a moving interval $[\mathbb{b}(\Theta), \mathbb{s}(\Theta)]$ for two free boundary curves $\mathbb{b}$, $\mathbb{s}$ that separate the wait region from a buy region ($y < \mathbb{b}(\theta)$), respectively a sell region ($y > \mathbb{s}(\theta)$). See e.g. [LS84] for diffusions with reflecting boundaries. We describe $\mathbb{b}, \mathbb{s}$ rather explicitly by a system of ordinary differential equations (6.14) with boundary condition at infinity. Numerical calculation of $\mathbb{b}$ and $\mathbb{s}$ seems feasible (see Figure 6.1), although existence and uniqueness of $(\mathbb{b}, \mathbb{s})$ are not guaranteed, also due to this kind of boundary condition at infinity, cf. Remark 6.1.6. Section 6.1.4 furthermore discusses the open problem of verification, drawing a connection to Dynkin games (cf. e.g. [KW01]) and double obstacle problems.

Tightly connected to the so-called *finite fuel* problem is its *infinite fuel* variant (cf. e.g. [KOWZ00]), which we formulate and solve in Section 6.2. Herein, the large investor trades indefinitely without constraint on the asset position. This relaxation makes the problem one-dimensional, where we can apply direct verification. The optimal strategy is described by two reflection local times, for buying and selling, that keep the market impact process Y inside a fixed interval. Heuristically, this problem can be understood as a limit of optimal liquidation problems for initial position $\theta \to \infty$, as it is the case in e.g. [KOWZ00] for a different problem.

6.1 Finite fuel variant

In this section, we will investigate the problem (6.6) of maximizing expected proceeds for liquidating a given finite position in the risky asset. In Section 6.1.1, we state the objective and discuss examples of admissible strategies. A heuristic derivation of a candidate optimizer by means of two free boundaries, that separate buy, wait, and sell regions, follows in Section 6.1.3 using smooth pasting. Verification remains an open problem for future research we comment on in Section 6.1.4. Our standing assumptions for the whole chapter are as follows.

Assumption 6.1.1.

- The transaction cost factor is $\kappa \in (1, \infty)$.
- We have $\delta := \gamma - \mu > 0$, that means the drift coefficient $-\delta\overline{S}$ for the γ-discounted fundamental price $e^{-\gamma t}\overline{S}_t$ is negative.
- The impact function $f \in C^3(\mathbb{R})$ satisfies $f, f' > 0$.
- The function $\lambda(y) := f'(y)/f(y)$, $y \in \mathbb{R}$, is bounded from above and away from zero, i.e. there exists $\lambda_{\min}, \lambda_{\max} \in (0, \infty)$ such that $\lambda_{\min} < \lambda(y) \leq \lambda_{\max}$ for all $y \in \mathbb{R}$. Moreover, it satisfies $\lambda' < (\Phi'/\Phi)'$ and $\lambda' < (\Phi''/\Phi')'$, where $\Phi = \Phi_\uparrow$ with

$$\Phi_\uparrow(x) := H_{-\delta/\beta}\Big((\sigma\rho\hat{\sigma} - \beta x)/\big(\sqrt{\beta}\hat{\sigma}\big)\Big), \qquad x \in \mathbb{R}, \tag{6.3}$$

  with Hermite function H_ν (cf. [Leb72, Sect. 10.2]) and $\sigma, \hat{\sigma}, \beta > 0$ and $\rho \in [-1, 1]$.
- The function $k(y) := \frac{\hat{\sigma}^2}{2}\frac{f''(y)}{f(y)} - (\beta + \delta) + (\sigma\rho\hat{\sigma} - \beta y)\frac{f'(y)}{f(y)}$, $y \in \mathbb{R}$, is strictly decreasing and there exists $y^* \in \mathbb{R}$ with $k(y^*) = 0$.

These assumptions are satisfied e.g. for constant λ. Note that boundedness of λ away from zero and infinity implies existence of real numbers y_0 and y_∞ solving $\lambda(y_0) = \Phi'_\uparrow(y_0)/\Phi_\uparrow(y_0)$ and $\lambda(y_\infty) = \Phi''_\uparrow(y_\infty)/\Phi'_\uparrow(y_\infty)$, respectively, since the right-hand sides have range $(0, \infty)$, cf. Lemma 4.4.1. Therefore, Assumption 6.1.1 differs from Assumption 4.1.2 only in the introduction of a constant κ, the existence of y^*, and the existence of a lower bound $\lambda_{\min} > 0$ for λ. We require $\lambda_{\min} > 0$ in order to have a positive antiderivative $F(y) := \int_{-\infty}^{y} f(z)\,\mathrm{d}z > 0$ of f. Existence of y^* is needed in Lemma 6.1.3.

Note that $\Phi_\uparrow$ is (up to a constant factor) *the* positive increasing eigenfunction to the eigenvalue δ of the infinitesimal generator of an Ornstein-Uhlenbeck process, i.e. $\mathcal{L}\Phi_\uparrow = 0$ for $\mathcal{L}\phi(y) = \frac{\hat\sigma^2}{2}\phi''(y) + (\sigma\rho\hat\sigma - \beta y)\phi'(y) - \delta\phi(y)$, cf. [BS02]. In addition to $\Phi_\uparrow$, also the decreasing positive solution $\Phi_\downarrow(y) = H_{-\delta/\beta}((\beta y - \sigma\rho\hat\sigma)/(\sqrt{\beta}\hat\sigma))$ will play a role in this chapter.

6.1.1 Optimal liquidation with proportional transaction costs

Our large trader has an initial position of $\theta \geq 0$ assets that she seeks to liquidate in an infinite time horizon through selling and buying according to non-decreasing processes Θ^- and Θ^+, respectively. We take $\Theta^+_{0-} := \Theta^-_{0-} := 0$ to incorporate a possible jump at time 0. Trading stops at first time

$$\tau_\theta(\Theta^+, \Theta^-) := \inf\big\{t > 0 \;\big|\; \Theta^-_t - \Theta^+_t \geq \theta \text{ or } \Theta^-_{t-} - \Theta^+_{t-} \geq \theta\big\}, \tag{6.4}$$

when the position is liquidated entirely (if $\tau_\theta(\Theta^+, \Theta^-) < \infty$). We seek to maximize total expected proceeds from trading, $\mathbb{E}[L_\infty(\Theta^+, \Theta^-)]$. For this term to make sense, we consider as admissible strategies

$$\begin{aligned}\mathcal{A}(\theta) := \Big\{(\Theta^+, \Theta^-) \;\Big|\; & \Theta^\pm \text{ are non-decreasing adapted càdlàg processes with} \\ & \Theta^\pm_{0-} = 0 \text{ and } \Theta^\pm_t = \Theta^\pm_{t\wedge\tau} \text{ for } \tau = \tau_\theta(\Theta^+, \Theta^-), \text{ such that} \\ & \int_0^\cdot e^{-\delta t}\mathcal{E}(\sigma W)_t F(Y^\Theta_t)\,\mathrm{d}W_t \text{ and } \int_0^\cdot e^{-\delta t}\mathcal{E}(\sigma W)_t f(Y^\Theta_t)\,\mathrm{d}B_t \\ & \text{are supermartingales, the almost sure limit} \\ & L_\infty(\Theta^+, \Theta^-) := \lim_{t\to\infty} L_t(\Theta^+, \Theta^-) \in [-\infty, \infty] \text{ exists, and} \\ & \lim_{t\to\infty} \mathbb{E}[L_t(\Theta^+, \Theta^-)^\pm] = E[L_\infty(\Theta^+, \Theta^-)^\pm]\Big\},\end{aligned} \tag{6.5}$$

where $L_t(\Theta^+, \Theta^-)^+ = 0 \vee L_t(\Theta^+, \Theta^-)$ and $L_t(\Theta^+, \Theta^-)^- = 0 \vee -L_t(\Theta^+, \Theta^-)$. The definition of $\mathcal{A}(\theta)$ guarantees that $\mathbb{E}[L_t(\Theta^+, \Theta^-)] \in [-\infty, \infty)$ is well defined, as we show in Lemma 6.1.3 below. Note that $L_t(\Theta^+, \Theta^-) = L_{\tau\wedge t}(\Theta^+, \Theta^-)$ for $\tau = \tau_\theta(\Theta^+, \Theta^-)$. We have $\mathcal{A}(\theta_1) \subset \mathcal{A}(\theta_2) \subset \mathcal{A}(\infty)$ (using $\tau_\infty(\Theta^+, \Theta^-) = \infty$) for $\theta_1 \leq \theta_2$. Now our optimization objective is

$$\max_{(\Theta^+, \Theta^-)\in\mathcal{A}(\theta)} \mathbb{E}[L_\infty(\Theta^+, \Theta^-)], \tag{6.6}$$

with value function

$$v(y, \theta) := \sup_{(\Theta^+, \Theta^-)\in\mathcal{A}(\theta)} \mathbb{E}[L_\infty(\Theta^+, \Theta^-)] \qquad \text{where } Y_{0-} = y. \tag{6.7}$$

We will now study the class $\mathcal{A}(\theta)$, investigate in Section 6.1.2 the variational inequalities that the value function should satisfy and characterize in Section 6.1.3 a candidate optimizer as a *reflecting strategy*, which will be defined below (see page 107).

The set of admissible strategies, whose definition (6.5) is admittedly technical, includes for instance the set of bounded variation strategies with finite stopping time, as shown in the following lemma. A further important subset of $\mathcal{A}(\theta)$ will be given by Lemma 6.1.4.

Lemma 6.1.2. *Let $\Theta^+, \Theta^- \geq 0$ be bounded non-decreasing adapted càdlàg processes with $\Theta_t^\pm = \Theta_{t\wedge\tau}^\pm$ for a finite time $\tau < \infty$. Let $\theta \in [0,\infty]$. If for some $\theta \in [0,\infty]$ we can bound $\Theta_t^- - \Theta_t^+ \leq \theta$ for all t and $\Theta_\cdot^- - \Theta_\cdot^+ < \theta$ on $[\![0,\tau[\![$, then Θ^+ and Θ^- form an admissible pair, i.e. $(\Theta^+,\Theta^-) \in \mathcal{A}(\theta)$.*

Note that $\tau \leq \tau_\theta(\Theta^+,\Theta^-)$ and we have $\tau = \tau_\theta(\Theta^+,\Theta^-)$ if $\tau_\theta(\Theta^+,\Theta^-) < \infty$.

Proof. Consider the Ornstein-Uhlenbeck process X with $\mathrm{d}X_t = -\beta X_t + \hat{\sigma}\,\mathrm{d}B_t$ and $X_0 = Y_{0-}$. We have $Y_t = X_t + \int_0^t e^{-\beta(t-s)}\,\mathrm{d}\Theta_s$ for $\Theta = \Theta^+ - \Theta^-$, which implies $|Y_t - X_t| \leq c$ for some constant $c \in (0,\infty)$ by boundedness of Θ. By $\tau < \infty$, we have $C(\omega) := \sup_{t\in[0,\tau]} e^{-\delta t} f(X_t + c)\mathcal{E}(\sigma W)_t < \infty$. Hence we can estimate

$$I_t^+ := \int_0^t e^{-\gamma u} f(Y_u^\Theta)\overline{S}_u\,\mathrm{d}\Theta_u^{c,+} + \sum_{\substack{0\leq u\leq t\\ \Delta\Theta_u^+\neq 0}} e^{-\gamma u}\overline{S}_u \int_0^{\Delta\Theta_u^+} f(Y_{u-}^\Theta + x)\,\mathrm{d}x \leq C(\omega)\cdot c,$$

$$I_t^- := \int_0^t e^{-\gamma u} f(Y_u^\Theta)\overline{S}_u\,\mathrm{d}\Theta_u^{c,-} + \sum_{\substack{0\leq u\leq t\\ \Delta\Theta_u^-\neq 0}} e^{-\gamma u}\overline{S}_u \int_0^{\Delta\Theta_u^-} f(Y_{u-}^\Theta - x)\,\mathrm{d}x \leq C(\omega)\cdot c.$$

It follows that $L_t = L_t(\Theta^+,\Theta^-) = I_t^- - \kappa I_t^+$ exists ω-wise. Since $\tau < \infty$, we conclude that $L_\infty = L_\tau \in \mathbb{R}$ exists a.s.

That the stochastic integrals w.r.t. B, W in the definition (6.5) of $\mathcal{A}(\theta)$ are true martingales follows from the bound $|Y_t - X_t| \leq c$ similarly to the proof of Lemma 4.5.2 using the exponential growth of F and f and that X and $\overline{S}$ are Gaussian.

For the expectation $\mathbb{E}[L_t]$, consider $\hat{L}_t := I_t^- - I_t^+$ so that $L_t \leq \hat{L}_t$. With $\Theta_t := \Theta_t^+ - \Theta_t^-$, we can rewrite

$$\hat{L}_t = -\int_0^t e^{-\gamma u} f(Y_u^\Theta)\overline{S}_u\,\mathrm{d}\Theta_u^c - \sum_{\substack{0\leq u\leq t\\ \Delta\Theta_u\neq 0}} e^{-\gamma u}\overline{S}_u \int_0^{\Delta\Theta_u} f(Y_{u-}^\Theta + x)\,\mathrm{d}x. \tag{6.8}$$

By integration by parts, we can rewrite (cf. also (5.15))

$$\begin{aligned}\hat{L}_t &= \overline{S}_0 F(Y_{0-}) - e^{-\gamma t}\overline{S}_t F(Y_t^\Theta) + \int_0^t e^{-\gamma u}\overline{S}_u \psi(Y_u^\Theta)\,\mathrm{d}u \\ &\quad + \int_0^t e^{-\gamma u}\sigma\overline{S}_u F(Y_u^\Theta)\,\mathrm{d}W_u + \int_0^t e^{-\gamma u}\hat{\sigma}\overline{S}_u f(Y_u^\Theta)\,\mathrm{d}B_u\,,\end{aligned} \tag{6.9}$$

for $\psi(y) := \frac{\hat{\sigma}^2}{2} f'(y) + (\sigma\rho\hat{\sigma} - \beta y) f(y) - \delta F(y)$. Due to boundedness of $f'/f \in [\lambda_{\min}, \lambda_{\max}]$, we have $\lim_{y\to\infty}\psi(y) = -\infty$. Moreover, $\psi'(y) = f(y)(k(y)+\beta)$ and $k > 0$ on $(-\infty, y^*)$, so that ψ is increasing on $(-\infty, y^*)$. Hence, there exists a positive upper bound $\hat{c} \geq \sup_y \psi(y)$ and we can conclude

$$\begin{aligned}\hat{L}_t &\leq \overline{S}_0 F(Y_{0-}) + \hat{c}\overline{S}_0 \int_0^t e^{-\delta u}\mathcal{E}(\sigma W)_u\,\mathrm{d}u \\ &\quad + \sigma\overline{S}_0 \int_0^t e^{-\delta u}\mathcal{E}(\sigma W)_u F(Y_u^\Theta)\,\mathrm{d}W_u + \hat{\sigma}\overline{S}_0 \int_0^t e^{-\delta u}\mathcal{E}(\sigma W)_u f(Y_u^\Theta)\,\mathrm{d}B_u\,.\end{aligned}$$

Using that the two stochastic integrals are true martingales, we can find an upper bound $\mathbb{E}[\hat{L}_t] \le \overline{S}_0 F(Y_{0-}) + \hat{c}\overline{S}_0 (1 - e^{-\delta t})/\delta \le \overline{S}_0 F(Y_{0-}) + \hat{c}\overline{S}_0/\delta =: \hat{C}$ for all t. This implies $\mathbb{E}[L_\infty \vee 0] \le \mathbb{E}[\hat{L}_\infty \vee 0] < \infty$, so that $\mathbb{E}[L_\infty] \in [-\infty, \infty)$ is well-defined. □

The objective (6.6) is well-posed due to the following lemma.

Lemma 6.1.3. *For all* $(\Theta^+, \Theta^-) \in \mathcal{A}(\theta)$ *we have* $\mathbb{E}[L_t(\Theta^+, \Theta^-)^+] < \infty$ *and therefore* $\mathbb{E}[L_t(\Theta^+, \Theta^-)] \in [-\infty, \infty)$ *exists for all* $t \in [0, \infty]$.

Proof. As in the proof of Lemma 6.1.2 we get existence of $\mathbb{E}[L_t(\Theta^+, \Theta^-)] \in [-\infty, \infty)$ by considering the upper bound $\hat{L}_t \ge L_t$ from (6.8). For the integrability of $\hat{L}_t$, it suffices that the stochastic integrals w.r.t. B, W in the definition (6.5) of $\mathcal{A}(\theta)$ are supermartingales. □

Reflecting strategies, which keep the controlled impact process Y inside a (moving) interval through local time reflections, will turn out to be of particular importance for our analysis, as this set includes our (candidate) optimizer.

Definition (reflecting strategy). Let $\mathbb{b}, \mathbb{s} \in C^1([0,\infty))$ satisfy $\mathbb{b} < \mathbb{s}$ point-wise. A pair (Θ^+, Θ^-) of non-negative non-decreasing continuous processes starting at $\Theta_0^\pm \ge \Theta_{0-} = 0$ (i.e. with a possible jump at time 0) is called *reflecting strategy* if it keeps $Y_t = Y_t^\Theta$ with $\Theta_t = \theta + \Theta_t^+ - \Theta_t^-$ in the interval $[\mathbb{b}(\Theta_t), \mathbb{s}(\Theta_t)]$ for all $t \in [\![0, \tau_\theta(\Theta^+, \Theta^-)[\![$, with minimal initial jump $|\Theta_0 - \Theta_{0-}|$ to enforce this condition at time 0 and no trading after $\tau := \tau_\theta(\Theta^+, \Theta^-) = \inf\{t > 0 \mid \Theta_t = 0\}$. That is, Θ^+ and Θ^- satisfy

$$\mathrm{d}\Theta_t^+ = \mathbb{1}_{\{Y_t = \mathbb{b}(\Theta_t)\}}\,\mathrm{d}\Theta_t^+ \qquad \text{and} \qquad \mathrm{d}\Theta_t^- = \mathbb{1}_{\{Y_t = \mathbb{s}(\Theta_t)\}}\,\mathrm{d}\Theta_t^-,$$

with $Y_t \in [\mathbb{b}(\Theta_t), \mathbb{s}(\Theta_t)]$, for all $t \in [\![0, \tau[\![$.

Herein, Θ^+ is the reflection local time of Y^Θ at the lower boundary $\mathbb{b}(\Theta)$ and Θ^- is the reflection local time of Y^Θ at the upper boundary $\mathbb{s}(\Theta)$. In particular, the measures $\mathrm{d}\Theta^+$ and $\mathrm{d}\Theta^-$ have disjoint support, so we can reconstruct Θ^+ and Θ^- from Θ and the initial position $\theta = \Theta_{0-}$. We thus identify the pair (Θ^+, Θ^-) with Θ and also call the latter a strategy, writing $\Theta = \Theta^{\mathbb{b},\mathbb{s}}$.

Similarly to Chapter 7 for a single reflection boundary, such Θ^+, Θ^- are unique by known results on the Skorokhod problem, see [DI93] and cf. Remark 7.2.3.

For bounded, non-increasing $\mathbb{b}, \mathbb{s}$, reflecting strategies are admissible by Lemma 6.1.4 below. Compared to the bounded variation strategies from Lemma 6.1.2, reflecting strategies are only of finite variation and liquidation in finite time is not clear a priori.

Lemma 6.1.4. *Let* $\mathbb{b}, \mathbb{s} \in C^1([0,\infty))$ *be non-increasing bounded functions with* $\mathbb{b} + \varepsilon < \mathbb{s}$ *everywhere for some* $\varepsilon > 0$. *Then the corresponding* reflecting strategy *is admissible, i.e.* $(\Theta^+, \Theta^-) \in \mathcal{A}(\theta)$ *for with decomposition* $\Theta_t^{\mathbb{b},\mathbb{s}} = \theta + \Theta_t^+ - \Theta_t^-$ *into continuous increasing and decreasing parts* $\Theta^\pm$.

Proof. Let $C := \|\mathbb{b}\|_\infty \vee \|\mathbb{s}\|_\infty$. Boundedness of $Y^\Theta \le C$ gives the martingale property of the stochastic integrals $\int_0^\cdot e^{-\delta t}\mathcal{E}(\sigma W)_t F(Y^\Theta)_t\,\mathrm{d}W_t$ and $\int_0^\cdot e^{-\delta t}\mathcal{E}(\sigma W)_t f(Y^\Theta)_t\,\mathrm{d}B_t$, since we can bound the quadratic variation process by $\tilde{C}\mathbb{E}[\int_0^t (e^{-\delta u}\mathcal{E}(\sigma W)_u)^2\,\mathrm{d}u] < \infty$ for all t.

Considering the integrals $I_t^\pm = \int_0^t e^{-\gamma u} f(Y_u^\Theta)\overline{S}_u\,\mathrm{d}\Theta_t^\pm$ as in the proof of Lemma 6.1.2, we find that $L_\infty(\Theta^{\mathbb{b},\mathbb{s}}) = L_\infty(\Theta^+, \Theta^-)$ is well-defined. It remains to show existence of the limit $\mathbb{E}[L_t(\Theta^{\mathbb{b},\mathbb{s}})] \to \mathbb{E}[L_\infty(\Theta^{\mathbb{b},\mathbb{s}})] \in [-\infty,\infty)$, for which it suffices to prove integrability, $\mathbb{E}[I^+_\infty] < \infty$ and $\mathbb{E}[I^-_\infty] < \infty$.

By optional projection for $\mathbb{E}[I^\pm_T]$ and monotone convergence, we have

$$\mathbb{E}[I^\pm_\infty] = \overline{S}_0 \mathbb{E}\Big[\int_0^\infty e^{-\delta t} f(Y^\Theta_t)\mathcal{E}(\sigma W)_t \,\mathrm{d}\Theta^\pm_t\Big] = \overline{S}_0 \mathbb{E}\Big[\int_0^\infty e^{-\delta t} f(\tilde{Y}_t)\,\mathrm{d}\tilde{\Theta}^\pm_t\Big].$$

where $\mathrm{d}\tilde{Y}_t = (\sigma\rho\hat{\sigma} - \beta\tilde{Y}_t)\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t + \mathrm{d}\tilde{\Theta}_t$, $\tilde{Y}_0 = Y_0$, with $\tilde{\Theta} = \Theta_0 + \tilde{K}^+_t - \tilde{K}^-_t$ for continuous reflection local times $\tilde{K}^\pm$ satisfying $\tilde{Y}_t = \mathbb{b}(\tilde{\Theta}_t)$ $\mathrm{d}K^+$-a.e. and $\tilde{Y}_t = \mathbb{s}(\tilde{\Theta}_t)$ $\mathrm{d}K^-$-a.e., such that $\tilde{Y}_t \in [\mathbb{b}(\tilde{\Theta}_t), \mathbb{s}(\tilde{\Theta}_t)]$ for all t. That is, considering the shifted impact $\tilde{Y}$ instead of Y, we can disregard W. It follows $\mathbb{E}[I^\pm_\infty] \le f(C)\overline{S}_0 \int_0^\infty \mathbb{E}[e^{-\delta\tau^\pm(\ell)}]\,\mathrm{d}\ell$ for stopping times $\tau^\pm(\ell) := \inf\{t > 0 \mid \tilde{K}^\pm_t > \ell\}$.

Let $\sigma_0 = 0$ and define inductively $\sigma_n := \sigma^+_n \vee \sigma^-_n$ with $\sigma^+_{n+1} := \inf\{t > \sigma_n \mid \tilde{K}^+_t > \tilde{K}^+_{\sigma_n}\}$ and similarly $\sigma^-_{n+1} := \inf\{t > \sigma_n \mid \tilde{K}^-_t > \tilde{K}^-_{\sigma_n}\}$. Note that $\sigma_n \to \infty$ for $n \to \infty$ since both boundaries $\mathbb{b}$ and $\mathbb{s}$ are at least ε apart. Moreover, on each $[\![\sigma_n, \sigma_{n+1}[\![$ one of $\tilde{K}^+$ and $\tilde{K}^-$ is constant so that $\tilde{Y}$ on $[\![\sigma_n, \sigma_{n+1}[\![$ is a reflected diffusion with one-sided reflection only. Conditioning on time σ_n and $\theta_n := \Theta_{\sigma_n}$, $\ell_n := \tilde{K}^-_{\sigma_n}$, we can thus apply Theorem 7.2.2 with local-time dependent boundary $g_-(\ell) = \mathbb{s}(\theta_n + \ell_n - \ell)$ for $\ell \ge \ell_n$. Let X solve $\mathrm{d}X_t = (\sigma\rho\hat{\sigma} - \beta X_t)\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}\hat{B}_t - \mathrm{d}A_t$, $X_0 = Y_{\sigma_n}$, with $(\mathcal{F}_{\sigma_n+t})_{t\ge0}$-Brownian motion $\hat{B}_t := B_{\sigma_n+t}$ and reflection local time A satisfying $A_0 = 0$, $X_t = g_-(A_t)$ $\mathrm{d}A_t$-a.e. such that $X_t \le g_-(A_t)$ for all $t \ge 0$. Denote the inverse local time by $\tau_A(\ell) := \inf\{t > 0 \mid A_t > \ell\}$. On $\{\sigma_n \le \sigma_n + t \le \sigma_{n+1}\}$, we have $\tilde{Y}_{\sigma_n+t} = X_t$ and $\tilde{K}^-_{\sigma_n+t} = A_t$ so that also $\tau^-(\ell) = \sigma_n + \tau_A(\ell - \ell_n)$ whenever $\sigma_n \le \tau^-(\ell) \le \sigma_{n+1}$. Hence,

$$\begin{aligned}
&\mathbb{E}\big[e^{-\delta\tau^-(\ell)}\mathbb{1}_{\{\sigma_n\le\tau^-(\ell)\le\sigma_{n+1}\}} \bigm| \mathcal{F}_{\sigma_n}\big]\\
&\quad= e^{-\delta\sigma_n}\mathbb{E}\big[e^{-\delta\tau_A(\ell-\ell_n)}\mathbb{1}_{\{\tau_A(\ell-\ell_n)\le\sigma_{n+1}-\sigma_n\}} \bigm| \mathcal{F}_{\sigma_n}\big]\\
&\quad\le e^{-\delta\sigma_n}\mathbb{E}\big[e^{-\delta\tau_A(\ell-\ell_n)} \bigm| \mathcal{F}_{\sigma_n}\big]\\
&\quad= e^{-\delta\sigma_n}\exp\Big(-\int_0^{\ell-\ell_n}\big(g_-'(a)+1\big)\frac{\Phi_\uparrow'(g_-(a))}{\Phi_\uparrow(g_-(a))}\,\mathrm{d}a\Big)\\
&\quad\le e^{-\delta\sigma_n}\exp\Big(-\frac{\Phi_\uparrow'(-C)}{\Phi_\uparrow(-C)}(\ell-\ell_n)\Big) = e^{-\delta\sigma_n}e^{-c(\ell-\ell_n)}\,,
\end{aligned}$$

with $c := \Phi_\uparrow'(-C)/\Phi_\uparrow(-C) > 0$, using that $\Phi_\uparrow'/\Phi_\uparrow$ is positive and increasing by Lemma 4.4.1. Moreover, we have $\tau_A(\ell_{n+1} - \ell_n) = \tau^-(\ell_{n+1}) - \sigma_n \le \tau^-(\ell_{n+1}) - \tau^-(\ell_n)$, since $\sigma_n \ge \tau^-(\ell_n)$. Now considering an independent copy of $\hat{B}$ we get a copy of τ_A which is in particular independent of ℓ_{n+1}, so that $\mathbb{E}[e^{-\delta\tau_A(\ell_{n+1}-\ell_n)} \mid \ell_n, \ell_{n+1}] \le e^{-c(\ell_{n+1}-\ell_n)}$. Using $\tau^-(\ell_n) = \sum_{k=1}^n (\tau^-(\ell_k) - \tau^-(\ell_{k-1}))$ we conclude $\mathbb{E}[e^{-\delta\tau^-(\ell)}] \le e^{-c\ell}$.

Similarly, for $\tilde{K}^+$ reflecting $-\tilde{Y}^\Theta$ downwards at $-g_+(\tilde{\Theta})$ until σ_{n+1} with boundary $g_+(\ell) = \mathbb{b}(\theta_n - \ell_n + \ell)$ where now $\ell \ge \ell_n := \tilde{K}^+_{\sigma_n}$ we first get

$$\begin{aligned}
&\mathbb{E}\big[e^{-\delta\tau^+(\ell)}\mathbb{1}_{\{\sigma_n\le\tau^+(\ell)\le\sigma_{n+1}\}} \bigm| \mathcal{F}_{\sigma_n}\big]\\
&\quad\le e^{-\delta\sigma_n}\exp\Big(\int_0^{\ell-\ell_n^+}\big(1-g_+'(a)\big)\frac{\Phi_\downarrow'(g_+(a))}{\Phi_\downarrow(g_+(a))}\,\mathrm{d}a\Big)\\
&\quad\le e^{-\delta\sigma_n}\exp\Big(\frac{\Phi_\downarrow'(C)}{\Phi_\downarrow(C)}(\ell-\ell_n^+)\Big)
\end{aligned}$$

and then conclude $\mathbb{E}[e^{-\delta\tau^+(\ell)}] \le e^{-\tilde{c}\ell}$ for $\tilde{c} := -\Phi'_\downarrow(C)/\Phi_\downarrow(C) > 0$. Together, setting $\hat{c} := c \vee \tilde{c} > 0$, we find $\int_0^\infty \mathbb{E}[e^{-\delta\tau^\pm(\ell)}]\,\mathrm{d}\ell \le \int_0^\infty e^{-\hat{c}\ell}\,\mathrm{d}\ell = 1/\hat{c} < \infty$, so $\mathbb{E}[I^\pm_\infty] < \infty$. □

6.1.2 Variational inequalities for the value function

The Hamilton-Jacobi-Bellman equation for the maximization problem $\max_\Theta \mathbb{E}[L_\infty(\Theta)]$ suggests that the state space $\{(y,\theta) \in \mathbb{R} \times [0,\infty)\}$ separates into three regions $\mathcal{B}$ (buying), $\mathcal{W}$ (waiting) and $\mathcal{S}$ (selling) such that the value function $V(y,\theta) = v(y,\theta)$ satisfies

$$\begin{cases} \quad f(y) = V_y(y,\theta) + V_\theta(y,\theta) & \text{for } (y,\theta) \in \mathcal{S}, \\ \quad f(y) < V_y(y,\theta) + V_\theta(y,\theta) < \kappa f(y) & \text{for } (y,\theta) \in \mathcal{W}, \\ \quad \kappa f(y) = V_y(y,\theta) + V_\theta(y,\theta) & \text{for } (y,\theta) \in \mathcal{B}, \\ \mathcal{L}V(y,\theta) < 0 & \text{for } (y,\theta) \in \mathcal{B} \cup \mathcal{S}, \\ \mathcal{L}V(y,\theta) = 0 & \text{for } (y,\theta) \in \mathcal{W}, \end{cases} \tag{6.10}$$

where $\mathcal{L}\phi(y) = \frac{\hat{\sigma}^2}{2}\phi''(y) + (\sigma\rho\hat{\sigma} - \beta y)\phi'(y) - \delta\phi(y)$ and $\mathcal{L}\phi(y,\theta) := \mathcal{L}\big(\phi(\cdot,\theta)\big)(y)$.

In Section 6.1.3, we make the ansatz that there exist smooth functions $\mathbb{b} \in C^1([0,\infty))$ (buy boundary) and $\mathbb{s} \in C^1([0,\infty))$ (sell boundary) with $\mathbb{b} < \mathbb{s}$, such that

$$\begin{aligned} \mathcal{B} &= \{(y,\theta) : y < \mathbb{b}(\theta)\}, \\ \mathcal{W} &= \{(y,\theta) : \mathbb{b}(\theta) < y < \mathbb{s}(\theta)\}, \\ \mathcal{S} &= \{(y,\theta) : \mathbb{s}(\theta) < y\}. \end{aligned}$$

Since $\mathcal{L}V = 0$ in $\mathcal{W}$, we would have

$$V(y,\theta) = A(\theta)\Phi_\downarrow(y) + B(\theta)\Phi_\uparrow(y), \qquad \text{for } (y,\theta) \in \mathcal{W}, \tag{6.11}$$

with suitable factors $A(\theta)$, $B(\theta)$ and the positive increasing and decreasing solutions of $\mathcal{L}\phi = 0$, $\Phi_\uparrow$ and $\Phi_\downarrow$, respectively.

After a possible initial jump to reach $\mathcal{W}$, the strategy corresponding to $\mathbb{b}, \mathbb{s}$ consists continuous reflecting local times Θ^+ and Θ^- where Θ^+ reflects (Y^Θ, Θ) obliquely in direction $(+1,+1)$ at the lower boundary $\mathbb{b}(\Theta)$ and Θ^- reflects (Y^Θ, Θ) obliquely at $\mathbb{s}(\Theta)$ in direction $(-1,-1)$, until $\Theta_\tau = 0$ is reached at liquidation time τ with $Y^\Theta_\tau = \mathbb{s}(\Theta_\tau)$, if $\tau > 0$.

6.1.3 Heuristic construction of the candidate optimal strategy

In this section, we derive heuristically a candidate optimal strategy $\Theta^{\mathbb{b},\mathbb{s}}$ with a smooth pasting approach based on the ansatz discussed in Section 6.1.2 that the optimizer should be characterized by two smooth boundary curves $\mathbb{b}, \mathbb{s}$ that separate buy, wait, and sell regions. In the course of this heuristic derivation, we will make additional assumptions. We do not verify these assumptions, as direct verification seems out of reach as discussed in Section 6.1.4.

Our candidate $V(y,\theta)$ for the value function $v(y,\theta)$ should satisfy the system of variational inequalities and equalities (6.10). Let us assume that V is given in $\mathcal{W}$ by (6.11) for (yet unknown) factors $A(\theta), B(\theta)$ and C^1 boundary curves $\mathbb{b}(\theta), \mathbb{s}(\theta)$ defining $\mathcal{W} = \{(y,\theta) \in \mathbb{R} \times (0,\infty) \mid \mathbb{b}(\theta) < y < \mathbb{s}(\theta)\}$. With finite fuel singular control problems, reflecting boundaries are often accompanied by a "C^2–smooth-fit principle" while repelling or absorbing boundaries correspond to "C^1–smooth-fit", see e.g. [KOWZ00]. Since our candidate optimizer $\Theta^{\mathbb{b},\mathbb{s}}$ would stay in $\overline{\mathcal{W}} = \{(y,\theta) \mid \mathbb{b}(\theta) \le y \le \mathbb{s}(\theta)\}$ by means of reflection at both, $\mathbb{b}$ and $\mathbb{s}$, it is thus reasonable to expect a "C^2–smooth-fit principle" principle, so that we assume $V \in C^{2,1}$.

Remark 6.1.5 (Difference to deterministic impact dynamics)**.** Note that this setup of two reflecting boundaries would be different from an analogous proportional transaction cost modification of our deterministic impact model ($\hat{\sigma} = 0$) from Chapter 2. There, buy and sell boundaries, $\mathbb{b}$ and $\mathbb{s}$, should lie in $\{y < 0\}$ in order to have liquidation in finite time, so no stochasticity would push the impact process towards $\mathbb{b}$ which should be to the left $\mathbb{s}$, $\mathbb{b}(\theta) < \mathbb{s}(\theta)$ for all $\theta \ge 0$. Thus, $\mathbb{b}$ would be "repelling" if we apply the language of stochastic processes to this deterministic problem.

The assumption $V \in C^{2,1}$ means that the directional derivatives of V paste together at $y = \mathbb{b}(\theta)$ and $y = \mathbb{s}(\theta)$. Inside $\overline{\mathcal{W}}$, we have the directional derivatives

$$\begin{aligned} V_y + V_\theta &= A\Phi_\downarrow' + B\Phi_\uparrow' + A'\Phi_\downarrow + B'\Phi_\uparrow \,, \\ V_{yy} + V_{y\theta} &= A\Phi_\downarrow'' + B\Phi_\uparrow'' + A'\Phi_\downarrow' + B'\Phi_\uparrow' \,, \end{aligned}$$

with abbreviations $A = A(\theta)$, $\Phi_\uparrow = \Phi_\uparrow(y)$, $V_y = V_y(y,\theta)$, etc. Assuming that $V \in C^{2,1}$ satisfies (6.10), C^1-smooth pasting means $V_y(y,\theta) + V_\theta(y,\theta) = \kappa f(y)$ for $y = \mathbb{b}(\theta)$ and $V_y(y,\theta) + V_\theta(y,\theta) = f(y)$ at $y = \mathbb{s}(\theta)$. Similarly, C^2-smooth pasting gives $V_{yy}(y,\theta) + V_{y\theta}(y,\theta) = \kappa f'(y)$ for $y = \mathbb{b}(\theta)$ and $V_{yy}(y,\theta) + V_{y\theta}(y,\theta) = f'(y)$ for $y = \mathbb{s}(\theta)$.

For fixed position θ, these are four linear equations in the four unknown variables $(A, B, A', B') = (A(\theta), B(\theta), A'(\theta), B'(\theta))$. Let us rewrite this system in matrix form. Given $b = \mathbb{b}(\theta)$ and $s = \mathbb{s}(\theta)$, we can write $\mathbf{M}(b,s) \cdot (A, B, A', B')^\top = \boldsymbol{f}(b,s)$ with vector $\boldsymbol{f}(b,s) := (\kappa f(b)\,,\, f(s)\,,\, \kappa f'(b)\,,\, f'(s))^\top$ and block Toeplitz matrix $\mathbf{M}(b,s) \in \mathbb{R}^{4\times 4}$ given by

$$\mathbf{M}(b,s) := \begin{pmatrix} \mathbf{M}_1(b,s) & \mathbf{M}_0(b,s) \\ \mathbf{M}_2(b,s) & \mathbf{M}_1(b,s) \end{pmatrix} \text{ where } \mathbf{M}_n(b,s) := \begin{pmatrix} \Phi_\downarrow^{(n)}(b) & \Phi_\uparrow^{(n)}(b) \\ \Phi_\downarrow^{(n)}(s) & \Phi_\uparrow^{(n)}(s) \end{pmatrix}. \quad (6.12)$$

Note that for $\zeta(x,y) := \Phi_\uparrow(x)\Phi_\downarrow(y) - \Phi_\downarrow(x)\Phi_\uparrow(y) = -\det \mathbf{M}_0(b,s)$ we have $\zeta_x > 0$, $\zeta_{xxy} > 0$ etc. so that $\zeta(b,s) \lesseqgtr 0$ for $b \lesseqgtr s$, $\zeta(b,s) \gtreqless 0$ for $b \lesseqgtr s$ etc. using that $\Phi_\uparrow^{(n)} > 0$ and $(-1)^n\Phi_\downarrow^{(n)} > 0$ for all n, by Lemma 4.4.1. Hence the 2×2 blocks $\mathbf{M}_n(b,s)$ of $\mathbf{M}(b,s)$ are all invertible whenever $b \ne s$.

Assume that moreover the whole block Toeplitz matrix $\mathbf{M}(b,s)$ is non-singular. Let the symbols $\boldsymbol{A}, \boldsymbol{B}, \dot{\boldsymbol{A}}, \dot{\boldsymbol{B}}$ to denote the four components of the solution vector as functions of b, s instead of θ, $(\boldsymbol{A}(b,s), \boldsymbol{B}(b,s), \dot{\boldsymbol{A}}(b,s), \dot{\boldsymbol{B}}(b,s))^\top := \mathbf{M}(b,s)^{-1}\boldsymbol{f}(b,s)$, so that

$$\begin{aligned} A(\theta) &= \boldsymbol{A}(\mathbb{b}(\theta), \mathbb{s}(\theta)), & B(\theta) &= \boldsymbol{B}(\mathbb{b}(\theta), \mathbb{s}(\theta)), \\ A'(\theta) &= \dot{\boldsymbol{A}}(\mathbb{b}(\theta), \mathbb{s}(\theta)), & B'(\theta) &= \dot{\boldsymbol{B}}(\mathbb{b}(\theta), \mathbb{s}(\theta)). \end{aligned} \quad (6.13)$$

Now chain rule gives $\dot{\boldsymbol{A}} = \boldsymbol{A}_s\,\mathbb{s}' + \boldsymbol{A}_b\,\mathbb{b}'$ and $\dot{\boldsymbol{B}} = \boldsymbol{B}_s\,\mathbb{s}' + \boldsymbol{B}_b\,\mathbb{b}'$. Therefore, the candidate $(\mathbb{b}, \mathbb{s}) : [0,\infty) \to \mathbb{R}^2$ for our two free boundaries should solve the autonomous differential equation

$$\begin{cases} \mathbb{b}'(\theta) = \left(\dfrac{\dot{\boldsymbol{A}}\boldsymbol{B}_s - \dot{\boldsymbol{B}}\boldsymbol{A}_s}{\boldsymbol{A}_b\boldsymbol{B}_s - \boldsymbol{B}_b\boldsymbol{A}_s}\right)(\mathbb{b}(\theta), \mathbb{s}(\theta)), \\ \mathbb{s}'(\theta) = \left(\dfrac{\dot{\boldsymbol{B}}\boldsymbol{A}_b - \dot{\boldsymbol{A}}\boldsymbol{B}_b}{\boldsymbol{A}_b\boldsymbol{B}_s - \boldsymbol{B}_b\boldsymbol{A}_s}\right)(\mathbb{b}(\theta), \mathbb{s}(\theta)). \end{cases} \quad (6.14)$$

The requirement of stopping as soon as $\Theta = 0$ is reached at time at $\tau_\theta(\Theta^+, \Theta^-)$ yields an (implicit) boundary condition at $\theta = 0$, $V(\mathbb{s}(0), 0) = 0$, i.e.

$$\boldsymbol{A}(b,s)\Phi_\downarrow(s) + \boldsymbol{B}(b,s)\Phi_\uparrow(s) = 0, \qquad \text{for } b = \mathbb{b}(0) \text{ and } s = \mathbb{s}(0). \tag{6.15}$$

This single equation does not suffice to obtain both, $\mathbb{b}(0)$ and $\mathbb{s}(0)$. Assume that $b_\infty = \lim_{\theta\to\infty} \mathbb{b}(\theta)$ and $s_\infty = \lim_{\theta\to\infty} \mathbb{s}(\theta)$ exist and also that $\tilde{V}(y) := \lim_{\theta\to\infty} V(y,\theta)$ exists. Regarding the literature on finite fuel singular control, e.g. discussed in [KOWZ00], it is natural to expect that $\tilde{V}$ solves an infinite fuel problem. We solve that problem in Section 6.2 and characterize the infinite fuel optimizer by two levels b, s solving (6.23). Then $b_\infty = b$ and $s_\infty = s$ would give an additional boundary condition at $\theta = \infty$ for the differential equation (6.14), which, together with (6.15), may suffice to find $\mathbb{b}, \mathbb{s}$, cf. Remark 6.1.6.

6.1.4 Outlook and open problems

Through our heuristic discussion so far we reached at a quite constructive ODE characterization (6.14) of a candidate optimal strategy $\Theta^{\mathbb{b},\mathbb{s}}$. Having an ODE is useful e.g. for simulations. However, at this point we are stuck with our analysis. Existence and uniqueness of the two non-constant free boundary curves $\mathbb{b}$ and $\mathbb{s}$ in $\mathbb{R}^2$ which separate the wait region from the action regions of buying and selling, respectively, is not clear (cf. Remark 6.1.6). Moreover, our previous verification scheme for the related problem in Chapter 4 crucially depended on the monotonicity of strategies in (4.16) to formulate a calculus of variations problem.

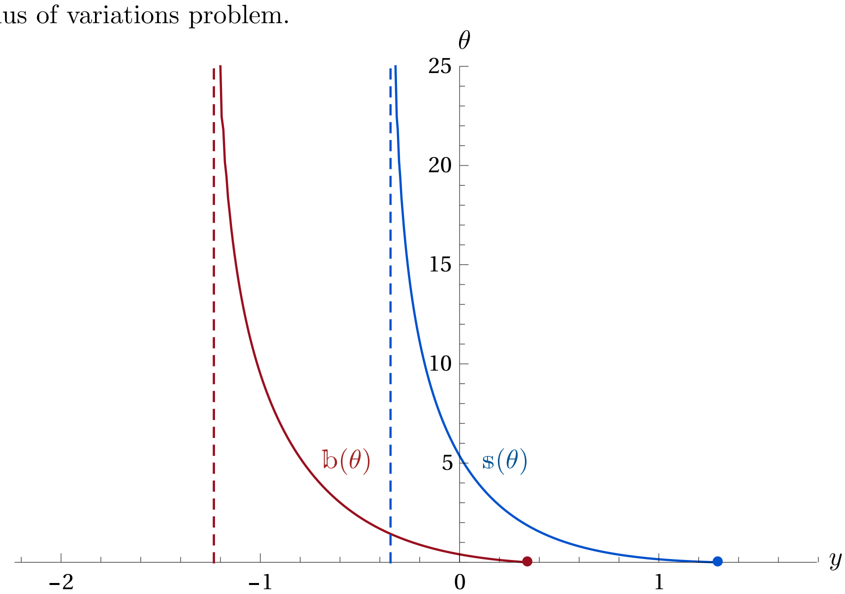


Figure 6.1: Possible solution $(\mathbb{b}, \mathbb{s})$ of (6.14) with boundary conditions (6.15) at zero and (6.23) for $(\mathbb{b}(\infty), \mathbb{s}(\infty))$. Parameters are $\delta = 0.1$, $\beta = 1$, $\rho = 0$, $\sigma = \hat{\sigma} = 1$, $f'/f = 1$, $\kappa = 1.1$.

Remark 6.1.6. Existence and uniqueness of a solution $(\mathbb{b}, \mathbb{s})$ to the autonomous differential equation (6.14) with boundary conditions at infinity given by (6.23) and at zero via (6.15) is in general not clear. However, numerical investigations suggest that the system is well-behaved and a simple Euler scheme "starting at infinity" gives promising results, cf. Figure 6.1. For this, denote by $D(b,s)$ the vector field of (6.14), so that $(\mathbb{b}'(\theta), \mathbb{s}'(\theta)) = D(\mathbb{b}(\theta), \mathbb{s}(\theta))$. Let (b_∞, s_∞) solve (6.23). We search an integral curve which asymptotically reaches the point (b_∞, s_∞).

Let us shortly explain how to calculate Figure 6.1. By construction (cf. Section 6.2), we have $D(b_\infty, s_\infty) = 0$. Numerical investigation suggests that (b_∞, s_∞) is not an attractor, but only two integral curves $(\mathbb{b}, \mathbb{s})$ converge to this point. To find their asymptotic directions $\lim_{\theta\to\infty} D(\mathbb{b}(\theta), \mathbb{s}(\theta))/\|D(\mathbb{b}(\theta), \mathbb{s}(\theta))\|$, choose $\varepsilon > 0$ and solve $\arg(-D(b_\infty + \varepsilon\cos\alpha, s_\infty + \varepsilon\sin\alpha)) = \alpha$ in $(-\pi, \pi]$, where $\arg(x,y) = \arg(x+iy)$ is the angular direction of a vector $(x,y) \in \mathbb{R}^2$. Since we expect $\mathbb{b}, \mathbb{s}$ to be decreasing, take a solution α^* in $[0, \pi/4]$ (the other solution $\hat{\alpha}$ seems to always lie in $(-\pi, -\pi/4]$). Starting in $\tilde{\mathbb{b}}(0) := b_\infty + \varepsilon\cos\alpha^*$, $\tilde{\mathbb{s}}(0) := s_\infty + \varepsilon\sin\alpha^*$, follow the trajectory backwards, solving $(\tilde{\mathbb{b}}', \tilde{\mathbb{s}}') = -D(\tilde{\mathbb{b}}, \tilde{\mathbb{s}})$ e.g. by Euler scheme until $w(\tilde{\mathbb{b}}(x^*), \tilde{\mathbb{s}}(x^*)) = 0$ at some $x^* > 0$ for $w(b,s) := \boldsymbol{A}(b,s)\Phi_\downarrow(s) + \boldsymbol{B}(b,s)\Phi_\uparrow(s)$. In fact, $x \mapsto w(\tilde{\mathbb{b}}(x), \tilde{\mathbb{s}}(x))$ seems to decrease, while it would increase following the other trajectory corresponding to $\hat{\alpha}$. Now $\mathbb{b}(\theta) := \tilde{\mathbb{b}}(x-\theta)$, $\mathbb{s}(\theta) := \tilde{\mathbb{s}}(x-\theta)$ for $\theta \in [0, x^*]$ solves (6.14) with (6.15) at zero and, ε-approximately, (6.23) at x^*. Decreasing $\varepsilon \to 0$ would give $x^* \to \infty$.

Remark 6.1.7 (A double obstacle problem)**.** Since we know little about the structure of $\mathbb{b}, \mathbb{s}$ solving (6.14), a direct verification of smoothness and of the variational (in-)equalities for our candidate value function V seems rather ambitious. A different approach might be to draw a connection to optimal stopping as in Section 4.6 for the sell-only problem. In the case of non-monotone strategies, the analogue would be a double obstacle problem or Dynkin game, cf. [KW01, DAF14, FP14]. In these articles, the connection to Dynkin games uses a convexity structure for the singular control problem, that I cannot see in our setup.

Assume we would already know $A(\theta)$ and $B(\theta)$. Regarding Corollary 4.6.1, it would be natural to expect that the directional derivative $w(y) = V_y(y,\theta) + V_\theta(y,\theta)$ of the value function should solve the double obstacle problem

$$\begin{cases} \mathcal{L}w(y) - \beta\big(A(\theta)\Phi_\downarrow'(y) + B(\theta)\Phi_\uparrow'(y)\big) = 0 & \text{if } f(y) < w(y) < \kappa f(y), \\ \mathcal{L}w(y) - \beta\big(A(\theta)\Phi_\downarrow'(y) + B(\theta)\Phi_\uparrow'(y)\big) \le 0 & \text{if } w(y) = f(y), \\ \mathcal{L}w(y) - \beta\big(A(\theta)\Phi_\downarrow'(y) + B(\theta)\Phi_\uparrow'(y)\big) \ge 0 & \text{if } w(y) = \kappa f(y). \end{cases} \tag{6.16}$$

This is closely related to the Dynkin game where two players choose stopping times τ, σ when to trade, one trying to maximize gains (sell at time τ), the other minimizes costs (buy at time σ),

$$\sup_\tau \inf_\sigma \mathbb{E}\bigg[e^{-\delta\tau} f(X_\tau)\mathbb{1}_{\tau<\sigma} + e^{-\delta\sigma}\kappa f(X_\sigma)\mathbb{1}_{\sigma<\tau} \tag{6.17}$$

$$- \int_0^{\sigma\wedge\tau} \beta e^{-\delta t}\big(A(\theta)\Phi_\downarrow'(X_t) + B(\theta)\Phi_\uparrow'(X_t)\big)\,\mathrm{d}t\bigg], \tag{6.18}$$

with shifted uncontrolled impact process $\mathrm{d}X_t = (\sigma\rho\hat{\sigma} - \beta X_t)\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t$, $X_0 = y$.

Such reformulations are used in [DY09] and the recent [FR19] to solve different non-monotone singular control problems. Therein however, the problems decouple, i.e. they can be rewritten such that $w = V_y$ has no dependence on θ (in our notation).

6.2 Infinite fuel variant

As discussed in Section 6.1.4, the finite fuel problem of liquidating a given amount of assets is rather involved. Moreover, our description (6.14) of a candidate optimizer for that problem involves a rather intricate boundary condition at infinity. For this reason, we now study the (simpler) the *infinite fuel* variant. Because the current amount of assets ("fuel") is no longer a relevant state variable of the control problem for this variant, the problem becomes simpler, being a singular control problem in one dimension (instead of two). We solve this one-dimensional problem with Theorem 6.2.8 using a smooth pasting approach and direct verification of the variational inequalities.

Intuitively, since $\tau_\theta(\Theta^+, \Theta^-) \to \infty$ for $\theta \to \infty$, the finite fuel optimal strategy should asymptotically equal the infinite fuel optimizer and the finite fuel value function $V(y, \theta)$ should converge to the infinite fuel value function $\tilde{V}(y)$ for initial position $\theta \to \infty$. The same interpretation may be expected for the asymptote y_∞ of the boundary surface $\mathbb{y}$ in Chapter 4. See also [KOWZ00] for a discussion of the connection between finite and infinite fuel solutions, or the recent [FK19], which is tightly related to our model from Chapter 4.

6.2.1 A one-dimensional problem

Our large investor trades indefinitely without any short sale restriction. We impose the same Assumption 6.1.1 as in the previous section. Note that formally $\tau_\infty(\Theta^+, \Theta^-) = \infty$ for the "time to liquidation" from (6.4). Hence, our objective is (6.6) with $\theta = \infty$, i.e.

$$\max_{(\Theta^+,\Theta^-)\in\mathcal{A}(\infty)} \mathbb{E}[L_\infty(\Theta^+, \Theta^-)], \tag{6.19}$$

for the set of admissible strategies $\mathcal{A}(\infty)$ from (6.5), with value function

$$v(y) := \sup_{(\Theta^+,\Theta^-)\in\mathcal{A}(\infty)} \mathbb{E}[L_\infty(\Theta^+, \Theta^-)] \qquad \text{where } Y_{0-} = y. \tag{6.20}$$

In spirit of the martingale optimality principle Proposition 4.5.1 from Chapter 4, we will construct a C^2 function $y \mapsto \tilde{V}(y)$ and a strategy $(\hat{\Theta}^+, \hat{\Theta}^-)$ such that the process $G_t := L_t(\Theta^+, \Theta^-) + e^{-\gamma t}\overline{S}_t \tilde{V}(Y_t^{\Theta^+-\Theta^-})$ is a supermartingale for all admissible strategies (Θ^+, Θ^-) and a martingale for $(\hat{\Theta}^+, \hat{\Theta}^-)$, which proves optimality of the latter and that $\tilde{V}$ is indeed the value function, $\tilde{V} = v$. We will verify in Section 6.2.2 our assumptions on $\tilde{V}$ and about the structure of the state space, which we describe next.

The supermartingale property for G suggests that the state space $\mathbb{R}$ of possible impact levels $y = Y_t$ should separate into three regions – the buy region $\mathcal{B}_\infty$, wait region $\mathcal{W}_\infty$, and sell region $\mathcal{S}_\infty$ – such that $\tilde{V}$ satisfies the variational (in-)equalities

$$\begin{cases} f(y) = \tilde{V}'(y) & \text{for } y \in \mathcal{S}_\infty, \\ f(y) < \tilde{V}'(y) < \kappa f(y) & \text{for } y \in \mathcal{W}_\infty, \\ \kappa f(y) = \tilde{V}'(y) & \text{for } y \in \mathcal{B}_\infty, \\ \mathcal{L}\tilde{V}(y) < 0 & \text{for } y \in \mathcal{B}_\infty \cup \mathcal{S}_\infty, \\ \mathcal{L}\tilde{V}(y) = 0 & \text{for } y \in \mathcal{W}_\infty, \end{cases} \tag{6.21}$$

with differential operator $\mathcal{L}\phi(y) = \frac{\hat{\sigma}^2}{2}\phi''(y) + (\sigma\rho\hat{\sigma} - \beta y)\phi'(y) - \delta\phi(y)$.

We make the ansatz that these are intervals $\mathcal{B}_\infty = (-\infty, b_\infty)$, $\mathcal{W}_\infty = (b_\infty, s_\infty)$, and $\mathcal{S}_\infty = (s_\infty, \infty)$, separated by two yet unknown boundary points $b_\infty < s_\infty$. The optimal strategy should then consist of the two reflection local times which keep the market impact process Y inside $\overline{\mathcal{W}_\infty}$.

Since $\Phi_\downarrow$ and $\Phi_\uparrow$ span the space of solutions $\phi \in C^2(\mathbb{R})$ of $\mathcal{L}\phi = 0$, imposing $\mathcal{L}\tilde{V} = 0$ in (b_∞, s_∞) means that there exist factors $A, B \in \mathbb{R}$ such that $\tilde{V}(y) = A\Phi_\downarrow(y) + B\Phi_\uparrow(y)$, for $b_\infty < y < s_\infty$. The variational equalities inside $\mathcal{B}_\infty$ and $\mathcal{S}_\infty$, cf. (6.21), hence suggest

$$\tilde{V}(y) = \begin{cases} A\Phi_\downarrow(y) + B\Phi_\uparrow(y), & \text{for } y \in [b_\infty, s_\infty] \\ A\Phi_\downarrow(b_\infty) + B\Phi_\uparrow(b_\infty) + \int_{b_\infty}^{y} \kappa f(x)\,\mathrm{d}x, & \text{for } y < b_\infty, \\ A\Phi_\downarrow(s_\infty) + B\Phi_\uparrow(s_\infty) + \int_{s_\infty}^{y} f(x)\,\mathrm{d}x, & \text{for } y > s_\infty, \end{cases} \tag{6.22}$$

which already gives continuity for all A, B and $b_\infty < s_\infty$. Now $\tilde{V} \in C^2$ means that first and second derivatives from the left and right paste together at b_∞ and s_∞. Using the matrices $\mathbf{M}_n$ from (6.12), this gives the two linear systems of equations for (A, B), namely $\mathbf{M}_1(b_\infty, s_\infty) \cdot \binom{A}{B} = \binom{\kappa f(b_\infty)}{f(s_\infty)}$ by C^1-pasting and $\mathbf{M}_2(b_\infty, s_\infty) \cdot \binom{A}{B} = \binom{\kappa f'(b_\infty)}{f'(s_\infty)}$ by C^2-pasting. Since $\mathbf{M}_n(b_\infty, s_\infty)$ are non-singular for $b_\infty \neq s_\infty$, we may combine both equations, eliminate (A, B), and obtain a non-linear system of equations

$$\mathbf{M}_2(b, s) \cdot \mathbf{M}_1(b, s)^{-1} \cdot \binom{\kappa f(b)}{f(s)} - \binom{\kappa f'(b)}{f'(s)} = \binom{0}{0}, \tag{6.23}$$

for $b_\infty = b$ and $s_\infty = s$. Given a solution (b_∞, s_∞) of (6.23) with $b_\infty < s_\infty$, we have with (6.22) a candidate value function $\tilde{V} \in C^2(\mathbb{R})$ that satisfies the variational equalities in (6.21). It then remains to prove the variational inequalities, $\mathcal{L}\tilde{V} < 0$ in $(-\infty, b_\infty] \cup [s_\infty, \infty)$ and $f < \tilde{V} < \kappa f$ in (b_∞, s_∞). We verify these in Section 6.2.2. Let us now comment on (6.23).

Remark 6.2.1 (Existence of a solution)**.** In the same way existence of y_∞ is just a model assumption in Chapter 4 for general $\lambda = f'/f$, that we verified e.g. for the particular case of constant λ, we may simply impose existence of b, s with $b < s$ solving (6.23) as an additional model assumption. Again, for λ being constant, we can indeed prove existence of such a solution, see Lemma 6.2.2 below.

Lemma 6.2.2. *Let $f(y) = e^{\lambda y}$ with constant $\lambda > 0$. Then there exists a solution $(b, s) \in \mathbb{R}^2$ of* (6.23) *with $b < s$.*

Proof. For constant λ, (6.23) is an eigenvalue problem: We search $b, s \in \mathbb{R}$ such that λ is an eigenvalue of $\mathbf{M}_2(b, s)\mathbf{M}_1(b, s)^{-1}$ with eigenvector $v := \binom{\kappa f(b)}{f(s)}$. The eigenvalue property reads $\det(\mathbf{M}_2\mathbf{M}_1^{-1} - \lambda I_2) = 0$. For $\zeta(x, y) = \Phi_\uparrow(x)\Phi_\downarrow(y) - \Phi_\downarrow(x)\Phi_\uparrow(y)$ we have

$$\begin{aligned} \det(\mathbf{M}_2\mathbf{M}_1^{-1} - \lambda I_2) &= \det(\mathbf{M}_2 - \lambda\mathbf{M}_1)\det(\mathbf{M}_1)^{-1} \\ &= \frac{(\Phi_\downarrow'' - \lambda\Phi_\downarrow')(s)\,(\Phi_\uparrow'' - \lambda\Phi_\uparrow')(b) - (\Phi_\downarrow'' - \lambda\Phi_\downarrow')(b)\,(\Phi_\uparrow'' - \lambda\Phi_\uparrow')(s)}{\zeta_{xy}(b, s)}. \end{aligned}$$

Note that $\Phi_\downarrow'' - \lambda\Phi_\downarrow' > 0$ for $\lambda \geq 0$. Hence for $b \neq s$ we have $\det(\mathbf{M}_2\mathbf{M}_1^{-1} - \lambda I_2) = 0$ if and only if

$$\left(\frac{\Phi_\uparrow'' - \lambda\Phi_\uparrow'}{\Phi_\downarrow'' - \lambda\Phi_\downarrow'}\right)(b) = \left(\frac{\Phi_\uparrow'' - \lambda\Phi_\uparrow'}{\Phi_\downarrow'' - \lambda\Phi_\downarrow'}\right)(s). \tag{6.24}$$

We know that $g(y) := (\Phi''_\uparrow - \lambda\Phi'_\uparrow)/(\Phi''_\downarrow - \lambda\Phi'_\downarrow)$ is negative on $(-\infty, y_\infty)$, with $g(y) \to 0$ for $y \to -\infty$, and positive on (y_∞, ∞), where $\lambda = \Phi''_\uparrow(y_\infty)/\Phi'_\uparrow(y_\infty)$ as in Assumption 4.1.2. differentiating gives

$$g' = \frac{(\Phi'''_\uparrow - \lambda\Phi''_\uparrow)(\Phi''_\downarrow - \lambda\Phi'_\downarrow) - (\Phi''_\uparrow - \lambda\Phi'_\uparrow)(\Phi'''_\downarrow - \lambda\Phi''_\downarrow)}{(\Phi''_\downarrow - \lambda\Phi'_\downarrow)^2}.$$

By Lemma 4.4.1, there exists a unique solution z to $\Phi'''_\uparrow(z) = \lambda\Phi''_\uparrow(z)$. Moreover, $z < y_\infty$ and $\Phi''''_\uparrow - \lambda\Phi'''_\uparrow > 0$ on $[z, \infty)$. So we get $g' > 0$ on (y_∞, ∞) and $g' < 0$ on $(-\infty, z)$. Let $y_* \in [z, y_\infty]$ solve $g'(y_*) = 0$. Using monotonicity of $\Phi^{(n+1)}_\uparrow/\Phi^{(n)}_\uparrow$ and the definitions of z and y_∞, we find

$$g''(y_*) = \left(\frac{(\Phi''''_\uparrow - \lambda\Phi'''_\uparrow)(\Phi''_\downarrow - \lambda\Phi'_\downarrow) - (\Phi''_\uparrow - \lambda\Phi'_\uparrow)(\Phi''''_\downarrow - \lambda\Phi'''_\downarrow)}{(\Phi''_\downarrow - \lambda\Phi'_\downarrow)^2}\right)(y_*) > 0,$$

which implies uniqueness of y_*. So for every $s \in [y_*, y_\infty)$ there exists exactly one $b \in (-\infty, y_*]$ (and vice versa) such that $g(b) = g(s)$, i.e. such that λ is an eigenvalue of $\mathbf{M}_2(b,s)\mathbf{M}_1(b,s)^{-1}$, and all (b, s) with $g(b) = g(s)$ satisfy $b, s < y_\infty$. The implied function $b(s) : (y_*, y_\infty) \to (-\infty, y_*)$ with $g(b(s)) = g(s)$ is moreover decreasing, since $g' > 0$ on (y_*, y_∞) and $g' < 0$ on $(-\infty, y_*)$.

It remains to check for which $b = b(s)$ the vector $v := \binom{\kappa f(b)}{f(s)}$ is an eigenvector of

$$\mathbf{M}_2(b,s)\mathbf{M}_1(b,s)^{-1} = \frac{1}{\zeta_{xy}(b,s)} \begin{pmatrix} \zeta_{xxy}(b,s) & \zeta_{xyy}(b,b) \\ \zeta_{xxy}(s,s) & \zeta_{xyy}(b,s) \end{pmatrix}$$

to the eigenvalue λ. Since we already know that λ is an eigenvalue when $b = b(s)$, the eigenvector property reduces to the single equation $\kappa = \kappa_*(b,s)$ with

$$\kappa_*(b,s) := \frac{-\zeta_{xyy}(b,b)}{\zeta_{xxy}(b,s) - \lambda\zeta_{xy}(b,s)} \frac{f(b)}{f(s)} = \frac{\frac{\Phi''_\uparrow(b)}{\Phi''_\downarrow(b)}\Phi'_\downarrow(b) - \Phi'_\uparrow(b)}{\left(1 - \lambda\frac{\Phi'_\downarrow(b)}{\Phi''_\downarrow(b)}\right)\left(g(b)\Phi'_\downarrow(s) - \Phi'_\uparrow(s)\right)} \frac{f(b)}{f(s)}.$$

From Lemma 4.4.1 it follows that $\Phi'_\downarrow(b)/\Phi''_\downarrow(b) \to 0$ for $b \to -\infty$. Since moreover $g(b) \to 0$ and $f(b) \to 0$ for $b \to -\infty$ and $b(s) \to -\infty$ for $s \nearrow y_\infty$ we find that $\lim_{s \nearrow y_\infty} \kappa_*(b(s), s) = +\infty$. On the other hand, we immediately see $\kappa_*(y_*, y_*) = 1$. Hence by continuity of $s \mapsto \kappa_*(b(s), s)$ there exist $s_* \in (y_*, y_\infty)$ and $b_* := b(s_*) \in (-\infty, y_*)$ such that $\kappa = \kappa_*(b_*, s_*)$, i.e. the pair (b_*, s_*) solves (6.23). ☐

Remark 6.2.3 (Transient impact is necessary)**.** Note that for purely permanent impact, i.e. formally $\beta = 0$, the equation (6.23) would have no solution with $b < s$ in general. If $\beta = 0$, the market impact process Y is a controlled Brownian motion, so that the corresponding increasing and decreasing positive solutions of $\mathcal{L}\phi = 0$ would be $\Phi_\uparrow(y) = \exp(-y\sigma\rho/\hat\sigma + y\sqrt{(\sigma\rho)^2 + 2\delta}/\hat\sigma)$ and $\Phi_\downarrow(y) = \exp(-y\sigma\rho/\hat\sigma - y\sqrt{(\sigma\rho)^2 + 2\delta}/\hat\sigma)$, respectively. Now proceeding as above for constant λ, (6.24) implies $b = s$.

6.2.2 Direct verification

Throughout this section, assume that $(b_\infty, s_\infty) \in \mathbb{R}^2$ solves (6.23) with $b_\infty < s_\infty$. Such solution exists e.g. for constant $\lambda = f'/f$, cf. Lemma 6.2.2. In particular, we know $\tilde V \in C^2$ for $\tilde V$ from (6.22). Now, we can moreover show that $\tilde V$ is increasing and convex on (b_∞, s_∞).

Lemma 6.2.4. *The function $\tilde{V}$ from (6.22) for $b_\infty < s_\infty$ solving (6.23) is increasing and convex, $\tilde{V}', \tilde{V}'' > 0$, on the interval (b_∞, s_∞).*

Proof. From C^1-smooth pasting, we have $\binom{A}{B} = \mathbf{M}_1(b_\infty, s_\infty)^{-1} \cdot \binom{\kappa f(b_\infty)}{f(s_\infty)}$. Consider the auxiliary function $\zeta(x,y) = \Phi_\uparrow(x)\Phi_\downarrow(y) - \Phi_\downarrow(x)\Phi_\uparrow(y)$ to express the determinant of matrices $\mathbf{M}_n(b_\infty, s_\infty)$. Then we may rewrite

$$\tilde{V}(y) = \kappa f(b_\infty)\frac{\zeta_y(y, s_\infty)}{\zeta_{xy}(b_\infty, s_\infty)} + f(s_\infty)\frac{\zeta_x(b_\infty, y)}{\zeta_{xy}(b_\infty, s_\infty)}, \qquad \text{for } y \in [b_\infty, s_\infty]. \tag{6.25}$$

Remember that by Lemma 4.4.1 we know $\Phi_\uparrow^{(n)}(y) > 0$ and $(-1)^n\Phi_\downarrow^{(n)}(y) > 0$ for integer $n \geq 0$ and $y \in \mathbb{R}$. So ζ has partial derivatives $\zeta_{xxy} < 0$ and $\zeta_{xyy} > 0$ and therefore $-\det \mathbf{M}_1(x,y) = \zeta_{xy}(x,y) \gtreqless 0$ for $x \lesseqgtr y$. Hence for $y \in (b_\infty, s_\infty)$, we find $\zeta_{xy}(b_\infty, s_\infty) > \zeta_{xy}(y, s_\infty) > 0$ and $\zeta_{xy}(b_\infty, s_\infty) > \zeta_{xy}(b_\infty, y) > 0$, and therefore $\tilde{V}'(y) > 0$.

Similarly, we get $\zeta_{xxxyy} > 0$, $\zeta_{xxyyy} < 0$ and hence $-\det \mathbf{M}_2(x,y) = \zeta_{xxyy}(x,y) \lesseqgtr 0$ for $x \lesseqgtr y$. Now, C^2-smooth pasting yields yet another representation of $\tilde{V}$ which gives as derivative

$$\tilde{V}'(y) = \frac{\kappa f'(b_\infty)\zeta_{xyy}(y, s_\infty) + f'(s_\infty)\zeta_{xxy}(b_\infty, y)}{\zeta_{xxyy}(b_\infty, s_\infty)}, \tag{6.26}$$

for $y \in [b_\infty, s_\infty]$. From (6.26), we thus see $\tilde{V}'' > 0$ on (b_∞, s_∞). □

We are now able to prove the variational inequalities for $\tilde{V}$. Remember that the function k, $k(y) = \frac{\hat{\sigma}^2}{2}\frac{f''(y)}{f(y)} - (\beta + \delta) + (\sigma\rho\hat{\sigma} - \beta y)\frac{f'(y)}{f(y)}$ is strictly decreasing by Assumption 6.1.1.

Lemma 6.2.5. *We have $k(b_\infty) > 0 > k(s_\infty)$ and $y \mapsto \tilde{V}'(y)/f(y)$ is strictly decreasing on $[b_\infty, s_\infty]$. Therefore $\kappa f > \tilde{V}' > f$ on the interval (b_∞, s_∞).*

Proof. For notational convenience, denote $b := b_\infty$ and $s := s_\infty$. The function $g := \tilde{V}'/f$ satisfies the ordinary differential equation $\frac{\hat{\sigma}^2}{2}g''(y) = -k(y)g(y) + c(y)g'(y)$ with $c(y) := \beta y - \sigma\rho\hat{\sigma} - \hat{\sigma}^2\frac{f'(y)}{f(y)}$.

We have $g(b) = \kappa > 1 = g(s)$ and by C^2-smooth fit also $g'(b) = 0 = g'(s)$, because

$$f(b)^2 g'(b) = \tilde{V}''(b)f(b) - \tilde{V}'(b)f'(b) = \kappa f'(b)f(b) - \kappa f(b)f'(b) = 0,$$
$$f(s)^2 g'(s) = \tilde{V}''(s)f(s) - \tilde{V}'(s)f'(s) = f'(s)f(s) - f(s)f'(s) = 0.$$

Denote by $y_1 := b \vee \inf\{y \in \mathbb{R} : k(y) < 0\} \wedge s$ the zero of k in $[b,s]$ (if it exists) such that $k(\cdot) > 0$ on $[b, y_1)$ and $k(\cdot) < 0$ on $(y_1, s]$. By [SLG84, Lemma 4.1], we have $(y-s)g(y)g'(y) > 0$ for all $y \in [y_1, s)$. This implies $g'(\cdot) < 0$ on $[y_1, s)$, since $g > 0$ on $[b,s]$, as shown before. In particular, we must have $k(b) > 0$, i.e. $y_1 > b$, because otherwise $g'(b) < 0$.

For the interval $[b, y_1]$, we employ a change of variable $\varphi(y) := \int_0^y \exp(\int_0^u \frac{2}{\hat{\sigma}^2}c(v)\,\mathrm{d}v)\,\mathrm{d}u$ like in the proof of [SLG84, Lemma 4.1], so that $l(x) := g(\varphi^{-1}(x))$ solves the differential equation $\frac{\hat{\sigma}^2}{2}l''(x) = -k(\varphi^{-1}(x))\big(\varphi'(\varphi^{-1}(x))\big)^2 l(x)$. Let $\tilde{b} := \varphi(b)$ and $x_1 := \varphi(y_1)$. Since $k(\varphi^{-1}(x)) > 0$ and $l(x) > 0$ for $x \in [\tilde{b}, x_1)$, l is strictly concave there. Hence, $l'(\tilde{b}) = g'(b)/\varphi'(b) = 0$ implies that l is strictly decreasing on $[\tilde{b}, x_1)$. Thus, $g'(\cdot) < 0$

on (b, y_1), because $l'(\varphi(y))$ and $g'(y)$ have the same sign. In particular, we must have $k(s) < 0$, i.e. $y_1 < s$, because otherwise $g'(s) < 0$ by strict convexity of l. By continuity of g', it follows $g'(\cdot) < 0$ on the whole interval (b, s). □

The variational inequalities inside buy and sell regions turn out to be direct consequences of Lemma 6.2.5.

Lemma 6.2.6. *We have $\mathcal{L}\tilde{V}(y) < 0$ for $y > s_\infty$.*

Proof. Note that $\tilde{V}'(y) = f(y)$ and $\tilde{V}''(y) = f'(y)$ for $y > s_\infty$. By C^2-smooth fit, we get $h(s_\infty) = 0$ for $h(y) := \mathcal{L}\tilde{V}(y) = \frac{\hat{\sigma}^2}{2} f'(y) + (\sigma\rho\hat{\sigma} - \beta y) f(y) - \delta\tilde{V}(s_\infty) - \delta\int_{s_\infty}^{y} f(x)\,\mathrm{d}x$. Moreover, $h'(y) = f(y)k(y)$. Since $k(\cdot) < 0$ on $[s_\infty, \infty)$ by Lemma 6.2.5, it follows $h(y) < 0$ for all $y > s_\infty$. □

Lemma 6.2.7. *We have $\mathcal{L}\tilde{V}(y) < 0$ for $y < b_\infty$.*

Proof. Note that $\tilde{V}'(y) = \kappa f(y)$ and $\tilde{V}''(y) = \kappa f'(y)$ for $y < b_\infty$. By C^2-smooth fit, we get $h(b_\infty) = 0$ for $h(y) := \mathcal{L}\tilde{V}(y) = \frac{\hat{\sigma}^2}{2}\kappa f'(y) + (\sigma\rho\hat{\sigma} - \beta y)\kappa f(y) - \delta\tilde{V}(b_\infty) - \delta\int_{b_\infty}^{y} \kappa f(x)\,\mathrm{d}x$. Moreover, $h'(y) = \kappa f(y)k(y)$. Since $k(\cdot) > 0$ on $(-\infty, b_\infty]$ by Lemma 6.2.5, it follows $h(y) < 0$ for all $y < b_\infty$. □

With Lemmas 6.2.5 to 6.2.7, we now have all ingredients to prove our main result for this section.

Theorem 6.2.8. *Let $Y_{0-} = y \in \mathbb{R}$ and let $(b_\infty, s_\infty) \in \mathbb{R}^2$ with $b_\infty < s_\infty$ solve (6.23). Then the strategy $(\hat{\Theta}^+, \hat{\Theta}^-)$ given by $\hat{\Theta}_t^+ = (0 \vee \Delta) + K_t^+$ and $\hat{\Theta}_t^- = (0 \vee -\Delta) + K_t^-$ maximizes $\mathbb{E}[L_\infty(\Theta^+, \Theta^-)]$ among all $(\Theta^+, \Theta^-) \in \mathcal{A}(\infty)$ and $\tilde{V}(y) = \mathbb{E}[L_\infty(\hat{\Theta}^+, \Theta^-)]$, where the initial jump Δ is*

$$\Delta = \begin{cases} s_\infty - y & \text{if } y > s_\infty, \\ b_\infty - y & \text{if } y < b_\infty, \\ 0 & \text{if } y \in [b_\infty, s_\infty], \end{cases}$$

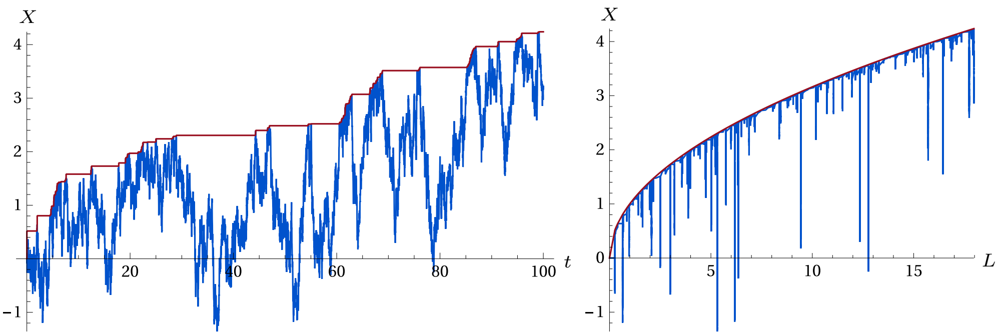
and K^+ and K^- are the minimal continuous increasing adapted non-negative processes that keep $Y_t = Y_t^{K^+ - K^-}$ (which starts in $Y_0 = y + \Delta \in [b_\infty, s_\infty]$) inside the interval $[b_\infty, s_\infty]$ for all $t \geq 0$. That is, K^+ is the upward reflection local time of Y at the lower boundary b_∞ and K^- is the downward reflection local time of Y at the upper boundary s_∞.

Proof. Admissibility $(\hat{\Theta}^+, \hat{\Theta}^-) \in \mathcal{A}_\infty$ essentially follows in the same way as Lemma 6.1.4. By construction $\tilde{V}$ satisfies the variational equalities $\tilde{V}' = f$ in $\mathcal{S}_\infty := [s_\infty, \infty)$, $\mathcal{L}\tilde{V} = 0$ in $\mathcal{W}_\infty := [b_\infty, s_\infty]$ and $\tilde{V}' = \kappa f$ in $\mathcal{B}_\infty := (-\infty, b_\infty]$. Lemmas 6.2.5 to 6.2.7 guarantee that $\tilde{V}$ also satisfies the corresponding variational inequalities, so that the process $G_t := L_t(\Theta^+, \Theta^-) + e^{-\gamma t}\overline{S}_t\tilde{V}(Y_t^{\Theta^+ - \Theta^-})$ is a supermartingale for every $(\Theta^+, \Theta^-) \in \mathcal{A}_\infty$ and a martingale for $(\hat{\Theta}^+, \hat{\Theta}^-)$, which implies optimality like in Proposition 4.5.1. □

Note that existence of $b_\infty < s_\infty$ as required in Theorem 6.2.8 is clear e.g. for constant λ, cf. Lemma 6.2.2. For general λ this existence requirement is just another assumption in addition to Assumption 6.1.1 as discussed in Remark 6.2.1.

7 Approximating diffusion reflections at elastic boundaries

This chapter corresponds to to the article [BBF18a]. Here, we show a probabilistic functional limit result for one-dimensional diffusion processes that are reflected at an elastic boundary which is a function of the reflection local time. Such processes are constructed as limits of a sequence of diffusions which are discretely reflected by small jumps at an elastic boundary, with reflection local times being approximated by ε-step processes. The construction yields an alternative proof for the Laplace transform formula of the inverse local time for reflection which is crucial in Chapter 4. The approximation scheme has a natural interpretation as a small block approximation of the optimal strategy from Chapter 4, cf. Remark 7.2.4.



(a) X against real time t.

(b) X against local time L.

Figure 7.1: Example. Brownian motion X_t (blue) reflected at the elastic boundary $g(L) = \sqrt{L}$ (red), where L is the reflection local time of X at boundary $g(L)$.

7.1 Elastic reflection – model and notation

Let W be a one-dimensional $(\mathcal{F}_t)$-Brownian motion. Consider Lipschitz-continuous functions $\sigma : \mathbb{R} \to (0, \infty)$ and $b : \mathbb{R} \to \mathbb{R}$ such that the continuous $\mathbb{R}$-valued (b, σ)-diffusion $\mathrm{d}Z_t = b(Z_t)\,\mathrm{d}t + \sigma(Z_t)\,\mathrm{d}W_t$ with generator $\mathcal{G} := \frac{1}{2}\sigma(x)^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} + b(x)\frac{\mathrm{d}}{\mathrm{d}x}$ is regular and recurrent. Moreover, let X be a (b, σ)-diffusion with reflection at an elastic boundary. This means that for a given non-decreasing $g \in C^1([0, \infty))$, the processes (X, L) satisfy

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t - \mathrm{d}L_t\,, \qquad X_0 = g(0)\,, \tag{7.1}$$

with the reflection local time L being a continuous non-decreasing process L that only grows when X is at the (local-time-dependent) boundary $g(L)$, i.e.

$$\mathrm{d}L_t = \mathbb{1}_{\{X_t = g(L_t)\}}\,\mathrm{d}L_t\,, \quad L_0 = 0\,, \quad \text{with } X_t \le g(L_t) \text{ for all } t \ge 0. \tag{7.2}$$

Note that the reflecting boundary is not deterministic in real time and space coordinates. Instead, the boundary $g(L)$, at which the diffusion X is being reflected, is elastic in the sense that it is itself a stochastic process which retracts when being hit, cf. Figure 7.1b. Strong existence and uniqueness of (X, L) follow from classical results (cf. Remark 7.2.3) and are also an outcome of our explicit construction below, see Lemma 7.3.9.

We are particularly interested (see Remark 7.2.4) in the inverse local time

$$\tau_\ell := \inf\{t > 0 \mid L_t > \ell\}. \tag{7.3}$$

Remark 7.1.1. Note that $\{t \geq 0 \mid X_t = g(L_t)\}$ is a.s. of Lebesgue measure zero by [RY99, ex. VI.1.16]. For a constant boundary $g(\ell) \equiv a$, Tanaka's formula for symmetric local times [RY99, ex. VI.1.25] hence shows that the process L, that we obtain as a solution to the SDE with reflection (7.1) – (7.2), is the symmetric local time of the continuous semimartingale X at given level $a \in \mathbb{R}$, i.e. $L_t = \lim_{\varepsilon \searrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{(a-\varepsilon, a+\varepsilon)}(X_s)\, \mathrm{d}\langle X, X\rangle_s$.

We denote by H^y the first hitting time of a point y by a (b, σ)-diffusion, and write $H^{x \to y}$ for the hitting time when the diffusion starts in x. Note that $\mathbb{P}[H^{x\to y} < \infty] = 1$ for all x, y by our assumption on the diffusion being regular and recurrent.

7.2 Approximation by small ε-reflections

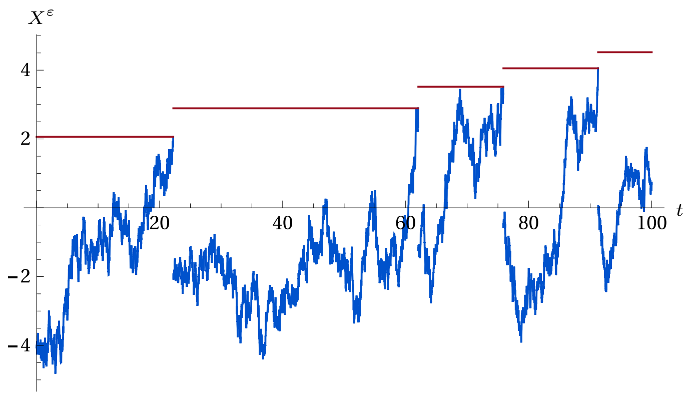


Figure 7.2: Approximating processes X^ε and $g(L^\varepsilon) = \sqrt{L^\varepsilon}$ for $\varepsilon = 4$.

We construct solutions to (7.1) – (7.2) and derive an explicit representation (7.12) of the Laplace transform of the inverse local time at boundary g by approximating reflection by jumps in the following system of SDEs:

$$\mathrm{d}X_t^\varepsilon = b(X_t^\varepsilon)\,\mathrm{d}t + \sigma(X_t^\varepsilon)\,\mathrm{d}W_t - \mathrm{d}L_t^\varepsilon\,, \qquad X_{0-}^\varepsilon := g(0)\,, \tag{7.4}$$

$$L_t^\varepsilon := \sum_{0\leq s\leq t} \Delta L_s^\varepsilon \quad \text{with } \Delta L_t^\varepsilon := \begin{cases} \varepsilon & \text{if } X_{t-}^\varepsilon = g(L_{t-}^\varepsilon), \\ 0 & \text{otherwise,} \end{cases} \qquad L_{0-}^\varepsilon := 0\,, \tag{7.5}$$

$$\tau_\ell^\varepsilon := \inf\{t > 0 \mid L_t^\varepsilon > \ell\} \quad \text{for } \ell \geq 0. \tag{7.6}$$

As soon as process X^ε hits the boundary, it is reflected by a jump of fixed size $\varepsilon > 0$. We will speak of L^ε as discrete local time, as it is approximating L in the sense of

Theorem 7.2.2. Since the target reflected diffusion X starts at the boundary g, we now have $X^\varepsilon_0 = g(0) - \varepsilon$ after an initial jump $\Delta L^\varepsilon_0 = \varepsilon$ away from $X^\varepsilon_{0-} := g(0)$.

Lemma 7.2.1. *For any $\varepsilon > 0$, the SDE (7.4)–(7.5) has a unique (up to indistinguishability) strong global solution $(X^\varepsilon_t, L^\varepsilon_t)_{t\geq 0}$. Moreover, uniqueness in law holds.*

Proof. Indeed, one can argue by results [RW87, V.9–11, V.17] for classical diffusion SDEs with Lipschitz coefficients (b,σ) by inductive construction on $[\![0,\tau_n[\![$ where for $n \geq 1$, $\tau_n := \inf\{t > \tau_{n-1} \mid X^\varepsilon_{t-} = g(n\varepsilon)\} = \tau^\varepsilon_{\varepsilon n}$ with $\tau_0 := 0$. Clearly L^ε_t equals $L^\varepsilon_{\tau_{n-1}}$ for $t \in [\![\tau_{n-1},\tau_n[\![$ and $L^\varepsilon_{\tau_n} = L^\varepsilon_{\tau_{n-1}} + \varepsilon$, while $X^\varepsilon_u = F(X^\varepsilon_{\tau_{n-1}}, (W_{\tau_{n-1}+s})_{s\geq 0})_{u-\tau_{n-1}}$ on $[\![\tau_{n-1},\tau_n[\![$ holds for a suitable functional representation F of strong solutions to (b,σ)-diffusions [RW87, Theorem V.10.4]. Such construction extends to $[\![0,\tau_\infty[\![$ for $\tau_\infty := \lim_n \tau_n$.

It suffices to show $\tau_\infty = \infty$ (a.s.). To this end, let $g_\infty := \lim_n g(n\varepsilon) \in \mathbb{R}\cup\{\infty\}$. In the case $g_\infty < \infty$, one can find $x, y \in \mathbb{R}$ with $g_\infty - \varepsilon < x < y < g_\infty$. By recurrence of (b,σ)-diffusions, we have (a.s.) finite times $\tau^y_0 := \inf\{t > 0 \mid X^\varepsilon_t = y\}$, $\tau^x_n := \inf\{t > \tau^y_{n-1} \mid X^\varepsilon_t = x\}$, $\tau^y_n := \inf\{t > \tau^x_n \mid X^\varepsilon_t = y\}$, for $n \in \mathbb{N}$. The durations $\tau^y_n - \tau^x_n$, $n \in \mathbb{N}$, for upcrossings of the interval $[x,y]$ are i.i.d., by the strong Markov property of the time-homogeneous diffusion. Moreover, X^ε is continuous on all $[\![\tau^x_n,\tau^y_n]\!]$. By the law of large numbers, $\frac{1}{n}\sum_{i=1}^n \exp(-\lambda(\tau^y_i - \tau^x_i))$ converges almost surely for $n \to \infty$ to the Laplace transform $\mathbb{E}_x[\exp(-\lambda H^y)]$, $\lambda \geq 0$, of the time H^y for hitting y by the (b,σ)-diffusion process (started at x). This expectation is strictly less than 1 for $\lambda > 0$, as $H^y > 0$ P_x-a.s. for $y > x$, whereas the limit of $\frac{1}{n}\sum_{i=1}^n \exp(-\lambda(\tau^y_i - \tau^x_i))$ equals 1 on $\{\tau_\infty < \infty\}$, where $\lim_{i\to\infty}(\tau^y_i - \tau^x_i) = 0$. Hence $P[\tau_\infty < \infty] = 0$.

If $g_\infty = \infty$, let $\tau'_n := \inf\{t > \tau_{n-1} \mid X^\varepsilon_{t-} = g((n-1)\varepsilon)\}$, for $n \geq 1$, so that $\tau_{n-1} < \tau'_n \leq \tau_n$ and $X^\varepsilon_{\tau'_n -} = g((n-1)\varepsilon) = X^\varepsilon_{(\tau_{n-1})-}$. Using the time change $\varphi_t := \int_0^t \sum_{n=1}^\infty 1_{[\![\tau'_n,\tau_n[\![}\,\mathrm{d}u$ with inverse $s_t := \inf\{u \mid \varphi_u > t\}$, we get (cf. [RW87, IV.30.10]) that $X'_t := X^\varepsilon_{s_t}$, $t \geq 0$, solves the SDE $\mathrm{d}X'_t = b(X'_t)\,\mathrm{d}t + \sigma(X'_t)\,\mathrm{d}W'_t$, $X'_0 = g(0)$, on $[\![0,\varphi_\infty[\![$ for $\varphi_\infty := \sup_t \varphi_t$, with respect to $W'_t = \int_0^{s_t}\sum_{n=1}^\infty 1_{[\![\tau'_n,\tau_n[\![}\,\mathrm{d}W_u$. We have $W'_t = B_{t\wedge\varphi_\infty}$ for some Brownian motion B on $[0,\infty)$ by the Dambis-Dubins-Schwarz theorem, cf. [KS91, Thm. 3.4.6, Prob. 3.4.7]. So X' solves the (b,σ)-diffusion SDE w.r.t. B on $[\![0,\varphi_\infty[\![$. Consider a (b,σ)-diffusion $\tilde X$ w.r.t. B on $[0,\infty)$. By the usual Gronwall argument for uniqueness of SDE solutions, we get $X' = \tilde X$ on all $[\![0,\varphi_{\tau_n}]\!]$ and hence $X' = \tilde X$ on $[\![0,\varphi_\infty[\![$. In particular, X' remains a.s. bounded on any finite time interval $[\![0,T[\![$ with $T \leq \varphi_\infty$. However, in the event $\{\tau_\infty < \infty\} \subset \{\varphi_\infty < \infty\}$, we get from $X'_{\varphi_{\tau_n}} = g(n\varepsilon) \to \infty$ that $\sup_{t<\varphi_\infty} X'_t = \infty$. Hence, we must have $\mathbb{P}[\tau_\infty < \infty] = 0$. □

By (7.4) – (7.6), we have $\tau^\varepsilon_0 = \tau^\varepsilon_{0-} = 0$ and $\tau^\varepsilon_\ell = \tau^\varepsilon_{(k-1)\varepsilon}$ for $\ell \in [(k-1)\varepsilon, k\varepsilon)$ with $k \in \mathbb{N}$, and $\tau^\varepsilon_{k\varepsilon}$ is the k-th jump time of X^ε and L^ε within period $(0,\infty)$. For $\ell = k\varepsilon$, the approximating process X^ε is a continuous (b,σ)-diffusion on stochastic intervals $[\![\tau^\varepsilon_{\ell-},\tau^\varepsilon_\ell[\![$, and $X^\varepsilon_{\tau^\varepsilon_\ell} = X^\varepsilon_{\tau^\varepsilon_{(\ell-)}} - \varepsilon = g(L^\varepsilon_{\tau^\varepsilon_{(\ell-)}}) - \varepsilon = g(\ell-\varepsilon) - \varepsilon$. For such $\ell = k\varepsilon$, we shall call $\tau^\varepsilon_\ell - \tau^\varepsilon_{\ell-}$ the length of the (k-th) excursion of X^ε away from the boundary. Note that this excursion length is independent of $\mathcal{F}^\varepsilon_{\tau^\varepsilon_{(\ell-)}}$ and its (conditional) distribution is

$$\tau^\varepsilon_\ell - \tau^\varepsilon_{\ell-} \sim H^{g(\ell)} \quad \text{under } \mathbb{P}_{g(\ell-\varepsilon)-\varepsilon}\,, \tag{7.7}$$

what is also denoted as $\tau^\varepsilon_\ell - \tau^\varepsilon_{\ell-} \overset{d}{=} H^{g(\ell-\varepsilon)-\varepsilon\,\to\,g(\ell)}$. The Laplace transform of first hitting times $H^{x\to z}$ is well-known, see e.g. [RW87, V.50]: for $x, z \in \mathbb{R}$ and $\lambda > 0$,

$$\mathbb{E}\big[e^{-\lambda H^{x\to z}}\big] \equiv \mathbb{E}_x\big[e^{-\lambda H^z}\big] = \begin{cases} \Phi_{\lambda,\uparrow}(x)/\Phi_{\lambda,\uparrow}(z) & \text{if } x < z, \\ \Phi_{\lambda,\downarrow}(x)/\Phi_{\lambda,\downarrow}(z) & \text{if } x > z, \end{cases} \tag{7.8}$$

where functions $\Phi_{\lambda,\uparrow}$ and $\Phi_{\lambda,\downarrow}$ are uniquely determined up to a constant factor as the increasing and decreasing, respectively, positive solutions ϕ of the differential equation $\mathcal{G}\phi = \lambda\phi$ with generator $\mathcal{G} = \frac{1}{2}\sigma(x)^2\frac{\mathrm{d}^2}{\mathrm{d}x^2} + b(x)\frac{\mathrm{d}}{\mathrm{d}x}$ of the (b,σ)-diffusion. Since we assume the boundary function g to be non-decreasing, only $\Phi_{\lambda,\uparrow}$ is of interest for our purpose and we abbreviate $\Phi_\lambda = \Phi_{\lambda,\uparrow}$.

Due to independence of Brownian increments over disjoint time intervals, the Laplace transform of the inverse local time can be calculated from a sum of (independent) excursion lengths at (discrete) local times $\ell_n := \varepsilon n$ as

$$\begin{aligned}\mathbb{E}\big[\exp\big(-\lambda\tau^\varepsilon_\ell\big)\big] &= \mathbb{E}\Bigg[\exp\Bigg(-\lambda\sum_{n=1}^{\lfloor \ell/\varepsilon\rfloor}\big(\tau^\varepsilon_{\ell_n} - \tau^\varepsilon_{\ell_n -}\big)\Bigg)\Bigg] = \prod_{n=1}^{\lfloor \ell/\varepsilon\rfloor}\mathbb{E}\Big[\exp\Big(-\lambda\big(\tau^\varepsilon_{\ell_n} - \tau^\varepsilon_{\ell_n -}\big)\Big)\Big] \\ &= \prod_{n=1}^{\lfloor \ell/\varepsilon\rfloor}\mathbb{E}_{g(\ell_n-\varepsilon)-\varepsilon}\big[\exp\big(-\lambda H^{g(\ell_n)}\big)\big] = \prod_{n=1}^{\lfloor \ell/\varepsilon\rfloor}\frac{\Phi_\lambda\big(g(\ell_n-\varepsilon)-\varepsilon\big)}{\Phi_\lambda\big(g(\ell_n)\big)} \\ &= \exp\Bigg(\sum_{n=1}^{\lfloor \ell/\varepsilon\rfloor}\log\bigg(\frac{\Phi_\lambda\big(g(\ell_n-\varepsilon)-\varepsilon\big)}{\Phi_\lambda\big(g(\ell_n)\big)}\bigg)\Bigg),\end{aligned} \tag{7.9}$$

for $\ell \geq 0$ and $\lambda > 0$. With $h_n(\xi) := \Phi_\lambda\big(g(\ell_n - \xi) - \xi\big)$, each summand in (7.9) equals

$$\begin{aligned}\log h_n(\varepsilon) - \log h_n(0) &= \int_0^\varepsilon \frac{h_n'(\xi)}{h_n(\xi)}\,\mathrm{d}\xi = -\int_0^\varepsilon \big(g'(\ell_n-\xi)+1\big)\frac{\Phi_\lambda'\big(g(\ell_n-\xi)-\xi\big)}{\Phi_\lambda\big(g(\ell_n-\xi)-\xi\big)}\,\mathrm{d}\xi \\ &= -\int_{\ell_{n-1}}^{\ell_n}\big(g'(a)+1\big)\frac{\Phi_\lambda'\big(g(a)+a-\ell_n\big)}{\Phi_\lambda\big(g(a)+a-\ell_n\big)}\,\mathrm{d}a\,.\end{aligned} \tag{7.10}$$

Therefore, we obtain

$$\mathbb{E}\big[\exp\big(-\lambda\tau^\varepsilon_\ell\big)\big] = \exp\Bigg(-\int_0^{\varepsilon\lfloor \ell/\varepsilon\rfloor}\big(g'(a)+1\big)\frac{\Phi_\lambda'\big(g(a)+a-\varepsilon\lceil a/\varepsilon\rceil\big)}{\Phi_\lambda\big(g(a)+a-\varepsilon\lceil a/\varepsilon\rceil\big)}\,\mathrm{d}a\Bigg). \tag{7.11}$$

Intuitively, this already suggests the formula (7.12) when taking $\varepsilon \to 0$.

Theorem 7.2.2. *The approximations $(X^\varepsilon_t, L^\varepsilon_t)_{t\geq 0}$ from* (7.4)–(7.5) *converge uniformly in probability for $\varepsilon \to 0$ to a pair $(X_t, L_t)_{t\geq 0}$ of continuous adapted processes with non-decreasing L, which is the unique strong solution (globally on $[0,\infty)$) to the reflected SDE* (7.1)–(7.2)*. The inverse local time $\tau_\ell := \inf\{t > 0 \mid L_t > \ell\}$ has the Laplace transform*

$$\mathbb{E}\big[e^{-\lambda\tau_\ell}\big] = \exp\bigg(-\int_0^\ell \big(g'(a)+1\big)\frac{\Phi_\lambda'\big(g(a)\big)}{\Phi_\lambda\big(g(a)\big)}\,\mathrm{d}a\bigg) \quad \text{for } \lambda > 0,\ \ell \geq 0, \tag{7.12}$$

where Φ_λ is the (up to a constant factor) unique positive increasing solution of the differential equation $\mathcal{G}\phi = \lambda\phi$, for $\mathcal{G}$ denoting the generator of the (b,σ)-diffusion.

Proof. Existence and uniqueness of (X, L) are shown in Lemma 7.3.9 below. Corollary 7.3.10 gives uniform convergence in probability. Using dominated convergence for the right-hand side of (7.11), we find $\lim_{\varepsilon\to 0}\mathbb{E}[e^{-\lambda\tau_\ell^\varepsilon}] = \exp\big(-\int_0^\ell (g'(a)+1)\frac{\Phi_\lambda'(g(a))}{\Phi_\lambda(g(a))}\,\mathrm{d}a\big)$. For the left-hand side, it suffices to prove weak convergence $\tau_\ell^\varepsilon \Rightarrow \tau_\ell$ as $\varepsilon \to 0$ for all $\ell \geq 0$. This is done in Corollary 7.3.11 below. □

Remark 7.2.3. Existence and uniqueness for (X, L) can also be concluded from classical results, cf. [DI93, suitably extended to non-bounded domains], by considering the pair (X, L) as a degenerate diffusion in $\mathbb{R}^2$ with oblique reflection in direction $(-1, +1)$ at a smooth boundary, see Figure 7.1b. This uses an iteration argument involving the Skorokhod-map and yields another approximation by a sequence of continuous processes. Yet, these do not satisfy the target diffusive dynamics inside the domain, except at the limiting fixed point (unless (b, σ) are constant). In contrast, $(X^\varepsilon, L^\varepsilon)$ adheres to the same dynamics as (X, L) between jump times, cf. (7.1) and (7.4), is Markovian and has a natural interpretation.

Remark 7.2.4. An application example for (7.12) and elastically reflected diffusions is the optimal execution for the sale of a financial asset position if liquidity is stochastic, see Chapter 4. A large trader with adverse price impact seeks to maximize expected proceeds from selling θ risky assets in an illiquid market. Her trading strategy A (predictable, càdlàg, non-decreasing) affects the asset price $S_t = f(Y_t^A)\overline{S}_t$ via a volume impact process $\mathrm{d}Y_t^A = -\beta Y_t^A\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t - \mathrm{d}A_t$ with $\overline{S}_t = \mathcal{E}(\sigma W)_t$ for an increasing function f, and Brownian motions (B, W) with correlation ρ. The gains to maximize in expectation are

$$G_T(A) := \int_0^T e^{-\delta t} f(Y_t^A)\overline{S}_t\,\mathrm{d}A_t^c + \sum_{\substack{0\leq t\leq T\\ \Delta A_t\neq 0}} e^{-\delta t}\overline{S}_t \int_0^{\Delta A_t} f(Y_{t-}^A - x)\,\mathrm{d}x.$$

The optimal strategy turns out to be the local time L of a reflected Ornstein-Uhlenbeck process X (with $b(x) := \rho\sigma\hat{\sigma} - \beta x$ and $\sigma(x) = \sigma > 0$) at a suitable elastic boundary g, as in (7.1)–(7.2), see Section 4.2. After a change of measure argument, one can write the expected proceeds from such strategies as $\mathbb{E}[G_\infty(L)] = \int_0^\theta f\big(g(\ell)\big)\mathbb{E}\big[e^{-\delta\tau_\ell}\big]\,\mathrm{d}\ell$. To find the optimal free boundary g, one can then apply (7.12) to express the proceeds as a functional of the boundary g, and optimize over all possible boundaries by solving a calculus of variations problem. This is key to the proof in Chapter 4. The discrete local time L^ε has a natural interpretation as the step process which approximates the continuous optimal strategy L by doing small block trades, as they would be realistic in an actual implementation, with identical (no-)action region. The approximation is asymptotically optimal for the control problem. Indeed, straightforward calculations similar to the derivation of (7.11) show that L^ε is asymptotically optimal in first order, i.e. $\mathbb{E}[G_\infty(L)] = \mathbb{E}[G_\infty(L^\varepsilon)] + \mathcal{O}(\varepsilon)$.

7.3 Tightness and convergence

To show convergence of $(\tau_\ell^\varepsilon)_\varepsilon$, we will prove that the pair of càdlàg processes $(X^\varepsilon, L^\varepsilon)$ forms a tight sequence in $\varepsilon \to 0$. Applying weak convergence theory for SDEs by Kurtz and Protter [KP96], we show that any limit point (for $\varepsilon \to 0$) satisfies (7.1) and (7.2). Uniqueness in law for solutions of (7.1) – (7.2) will then allow to conclude Theorem 7.2.2.

Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence with $\varepsilon_n \to 0$ and consider the sequence $(X^{\varepsilon_n}, L^{\varepsilon_n})_n$. To show tightness, we will apply the following criterion due to Aldous.

Proposition 7.3.1 ([Bil99, Cor. to Thm. 16.10]). *Let $(E, |\cdot|)$ be a separable Banach space. If a sequence $(Y^n)_{n\in\mathbb{N}}$ of adapted, E-valued càdlàg processes satisfies the following two conditions, then it is tight.*

(a) The sequences $\big(J_T(Y^n)\big)_n$ and $(Y_0^n)_n$ are tight (in $\mathbb{R}$, resp. E) for any $T \in (0,\infty)$, with $J_T(Y^n) := \sup_{0<t\le T}\big|Y_t^n - Y_{t-}^n\big|$ denoting the largest jump until time T.

(b) For any $T \in (0,\infty)$ and $\varepsilon_0, \eta > 0$ there exist $\delta_0 > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$, all (discrete) Y^n-stopping times $\hat\tau \le T$ and all $\delta \in (0, \delta_0]$ we have

$$\mathbb{P}\big[|Y^n_{\hat\tau+\delta} - Y^n_{\hat\tau}| \ge \varepsilon_0\big] \le \eta\,.$$

To get tightness one needs to control both jump size and, regarding $(L_n^{\varepsilon})_n$, the frequency of jumps simultaneously. As we are considering processes with jumps of size $\pm\varepsilon_n \to 0$, only the latter is not yet clear. To this end, the next lemma provides a technical bound on X^{ε_n}, L^{ε_n}, while a second lemma constricts the probability that X^{ε_n} (respectively L^{ε_n}) performs a number of N_n jumps in a time interval of fixed length.

Lemma 7.3.2 (Upper bound). *Fix a time horizon $T \in (0,\infty)$ and $\eta \in (0,1]$. Then there exists a constant $M \in \mathbb{R}$ such that $\mathbb{P}[\exists n : g(L_T^{\varepsilon_n} - \varepsilon_n) > M] \le \eta$, with the domain of definition for the function g being extended by $g(-x) := g(0)$ for $-x < 0$.*

Proof. Consider a continuous (b,σ)-diffusion Y that starts at time $t = 0$ at $g(0)$. For $n \in \mathbb{N}$ and $k = 0, 1, 2, \ldots$, let $\ell(n,k) := k\varepsilon_n$. By induction over k, using comparison for diffusion SDEs, cf. [KS91, Theorem 5.2.18], one obtains that (a.s.) $X_t^{\varepsilon_n} \le Y_t$ for $t \in [\![0, \tau^{\varepsilon_n}_{\ell(n,k)}[\![$ for all $k \ge 1$, and hence $X^{\varepsilon_n} \le Y$ on $[0,\infty)$ (a.s.) because $\lim_{k\to\infty}\tau^{\varepsilon_n}_{\ell(n,k)} = \infty$ for any n by Lemma 7.2.1. Therefore, on the event $\{\exists n : g(L_T^{\varepsilon_n} - \varepsilon_n) > M\}$ we have $\sup_{t\in[0,T]} Y_t \ge M$, and hence $H^{g(0)\to M} \le T$. Thus $\mathbb{P}[\exists n : g(L_T^{\varepsilon_n} - \varepsilon_n) > M] \le \mathbb{P}[H^{g(0)\to M} \le T]$. Now the claim follows because $\lim_{M\to\infty}\mathbb{P}[H^{g(0)\to M} \le T] = 0$. □

Lemma 7.3.3 (Frequency of jumps). *Fix $T \in (0,\infty)$, $\varepsilon_0, \eta > 0$, and set $N_n := \lceil \varepsilon_0/\varepsilon_n \rceil$. Then there exists $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for every bounded stopping time $\hat\tau \le T$ we have $\mathbb{P}\big[J^{\varepsilon_n}_{\hat\tau,\delta} \ge N_n\big] \le \eta$ for all $n \ge n_0$, where $J^{\varepsilon_n}_{\hat\tau,\delta} := \inf\{k \mid L^{\varepsilon_n}_{\hat\tau} + k\varepsilon_n \ge L^{\varepsilon_n}_{\hat\tau+\delta}\}$ is the number of jumps of X^{ε_n}, respectively L^{ε_n}, in time $]\!]\hat\tau, \hat\tau+\delta]\!]$.*

Proof. We will first find an estimate for the jump count probability for arbitrary but fixed $\delta > 0$, $n \in \mathbb{N}$, $N_n \in \mathbb{N}$ and $\hat\tau \le T$. Only in part 2) of the proof we will consider $(N_n)_{n\in\mathbb{N}}$ as stated, to study the limit $n \to \infty$. More precisely, we will show in part 1) that, given $\mathcal{F}_{\hat\tau}$, for every $\lambda > 0$ there exist $k_{n,\lambda} \in \{0, 1, \ldots, N_n - 1\}$ s.t. for $x_n := g(L^{\varepsilon_n}_{\hat\tau} + \varepsilon_n k_{n,\lambda})$,

$$\mathbb{P}\big[J^{\varepsilon_n}_{\hat\tau,\delta} \ge N_n \;\big|\; \mathcal{F}_{\hat\tau}\big] \le e^{\lambda\delta}\left(\frac{\Phi_\lambda(x_n - \varepsilon_n)}{\Phi_\lambda(x_n)}\right)^{N_n - 1}. \tag{7.13}$$

1) In this part, fix arbitrary $\delta > 0$, $n \in \mathbb{N}$, $N_n \in \mathbb{N}$ and $\hat\tau \le T$. We enumerate the jumps and estimate the sum of excursion lengths by δ. Let $\ell_k := L^{\varepsilon_n}_{\hat\tau} + k\varepsilon_n$ be the (discrete) local time at the k-th jump after time $\hat\tau$. If X^{ε_n} has at least N_n jumps in the

interval $]\!]\hat\tau, \hat\tau+\delta]\!]$, it is doing at least $N_n - 1$ complete excursions (cf. (7.7)), so that, noting that $\tau^{\varepsilon_n}_{L^{\varepsilon_n}_t - \varepsilon_n} \le t < \tau^{\varepsilon_n}_{L^{\varepsilon_n}_t}$ (for all $t \ge 0$) and $\ell_{N_n-1} + \varepsilon_n \le L^{\varepsilon_n}_{\hat\tau+\delta}$, we have

$$\delta = (\hat\tau+\delta) - \hat\tau \ge \tau^{\varepsilon_n}_{L^{\varepsilon_n}_{\hat\tau+\delta}-\varepsilon_n} - \tau^{\varepsilon_n}_{L^{\varepsilon_n}_{\hat\tau}} \ge \sum_{k=1}^{N_n-1}\big(\tau^{\varepsilon_n}_{\ell_k} - \tau^{\varepsilon_n}_{\ell_{k-1}}\big) \stackrel{d}{=} \sum_{k=1}^{N_n-1} H_k$$

with the last equality being in distribution conditionally on $\mathcal{F}_{\hat\tau}$, for H_k being conditionally independent and distributed as $H^{g(\ell_{k-1})-\varepsilon_n \to g(\ell_k)}$. Clearly, ℓ_k is $\mathcal{F}_{\hat\tau}$-measurable. By the Laplace transform (7.8) of H_k and the Markov inequality, we get for $\lambda > 0$

$$\begin{aligned}
\mathbb{P}\big[J^{\varepsilon_n}_{\hat\tau,\delta} \ge N_n \,\big|\, \mathcal{F}_{\hat\tau}\big] &\le \mathbb{P}\Big[\sum_{k=1}^{N_n-1} H_k \le \delta \,\Big|\, \mathcal{F}_{\hat\tau}\Big] \le e^{\lambda\delta}\mathbb{E}\Big[\exp\Big(-\lambda\sum_{k=1}^{N_n-1} H_k\Big) \,\Big|\, \mathcal{F}_{\hat\tau}\Big] \\
&= e^{\lambda\delta}\prod_{k=1}^{N_n-1}\mathbb{E}\Big[\exp\Big(-\lambda H^{g(\ell_{k-1})-\varepsilon_n\to g(\ell_k)}\Big) \,\Big|\, \mathcal{F}_{\hat\tau}\Big] \\
&= e^{\lambda\delta}\prod_{k=1}^{N_n-1}\frac{\Phi_\lambda\big(g(\ell_{k-1})-\varepsilon_n\big)}{\Phi_\lambda\big(g(\ell_k)\big)} \le e^{\lambda\delta}\prod_{k=1}^{N_n-1}\frac{\Phi_\lambda\big(g(\ell_k)-\varepsilon_n\big)}{\Phi_\lambda\big(g(\ell_k)\big)} \\
&\le e^{\lambda\delta}\Big(\max_{0\le k<N_n}\frac{\Phi_\lambda\big(g(\ell_k)-\varepsilon_n\big)}{\Phi_\lambda\big(g(\ell_k)\big)}\Big)^{N_n-1} = e^{\lambda\delta}\Big(\frac{\Phi_\lambda(x_n-\varepsilon_n)}{\Phi_\lambda(x_n)}\Big)^{N_n-1}
\end{aligned}$$

where $x_n := g(\ell_k)$ for the index $k = k_{n,\lambda}$ attaining the maximum.

2) For given $\delta > 0$ and $\hat\tau \le T$, let us now consider the sequence $N_n = \lceil \varepsilon_0/\varepsilon_n \rceil$, $n \in \mathbb{N}$. To investigate the limit $n \to \infty$, first observe that by Taylor expansion

$$\log\frac{\Phi_\lambda(x-\varepsilon_n)}{\Phi_\lambda(x)} = -\varepsilon_n\frac{\Phi'_\lambda(x)}{\Phi_\lambda(x)} + \varepsilon_n r(x,\varepsilon_n),$$

where $r(\cdot,\varepsilon_n) \to 0$ converges uniformly on compacts for $\varepsilon_n \to 0$. Since $\hat\tau + \delta \le T + \delta$ is bounded, Lemma 7.3.2 yields a constant $M \in \mathbb{R}$ such that $\mathbb{P}\big[\exists n : x_n > M\big] \le \frac{\eta}{2}$ for the x_n from above. On the event $\{\forall n : x_n \in I\}$ with compact $I := [g(0), M]$, we have uniform convergence of $r(x_n, \varepsilon_n)$ and thereby get

$$\begin{aligned}
\limsup_{n\to\infty} e^{\lambda\delta}\Big(\frac{\Phi_\lambda(x_n-\varepsilon_n)}{\Phi_\lambda(x_n)}\Big)^{N_n-1} &= \exp\Big(\lambda\delta + \limsup_{n\to\infty}(N_n-1)\log\frac{\Phi_\lambda(x_n-\varepsilon_n)}{\Phi_\lambda(x_n)}\Big) \\
&= \exp\Big(\lambda\delta + \limsup_{n\to\infty}(N_n\varepsilon_n - \varepsilon_n)\Big(r(x_n,\varepsilon_n) - \frac{\Phi'_\lambda(x_n)}{\Phi_\lambda(x_n)}\Big)\Big) \\
&\le \exp\Big(\lambda\delta - \varepsilon_0\inf_{x\in I}\frac{\Phi'_\lambda(x)}{\Phi_\lambda(x)}\Big) = \sup_{x\in I}\exp\Big(\lambda\delta - \varepsilon_0\frac{\Phi'_\lambda(x)}{\Phi_\lambda(x)}\Big).
\end{aligned}$$

By [PY03, Theorem 1], $\psi^x(\lambda) := \frac{1}{2}\Phi'_\lambda(x)/\Phi_\lambda(x)$ is the Laplace exponent of $A^x(\kappa^x_\cdot)$, where κ^x_ℓ is the inverse local time at constant level x of a (b,σ)-diffusion Z^x starting at x, and $A^x(t)$ is the occupation time $A^x(t) := \int_0^t \mathbb{1}_{\{Z^x_s \le x\}}\,\mathrm{d}s$. So we get for $\lambda \to \infty$ that $\exp\big(-2\varepsilon_0\psi^x(\lambda)\big) = \mathbb{E}_x\big[\exp\big(-\lambda A^x(\kappa^x_{2\varepsilon_0})\big)\big] \to 0$. By compactness of I and Dini's theorem there exists $\lambda = \lambda_{\varepsilon_0,\eta,M}$ such that for $\delta := 1/\lambda$ we have

$$\limsup_{n\to\infty} e^{\lambda\delta}\Big(\frac{\Phi_\lambda(x_n-\varepsilon_n)}{\Phi_\lambda(x_n)}\Big)^{N_n-1} \le e^{\lambda\delta}\sup_{x\in I}\exp\big(-2\varepsilon_0\psi^x(\lambda)\big) \le \frac{\eta}{2} \tag{7.14}$$

on the event $\{x_n \le M \text{ for all } n\}$. By equation (7.13) and $\mathbb{P}[\exists n : x_n > M] \le \eta/2$, this completes the proof. □

Using the preceding two lemmas, we will first prove tightness of $(L^{\varepsilon_n})_n$ and of $(X^{\varepsilon_n})_n$ separately. Tightness of the pair $(X^{\varepsilon_n}, L^{\varepsilon_n})_n$ is handled afterwards.

Lemma 7.3.4 (Tightness of the local time approximations)**.** *The sequence $(L^{\varepsilon_n})_n$ of càdlàg processes defined by* (7.4) *and* (7.5) *satisfies Aldous' criterion and thus is tight.*

Proof. Part (a) of Proposition 7.3.1 is clear, as the initial value $L_0^{\varepsilon_n} = \varepsilon_n$ is deterministic and $J_T(L^{\varepsilon_n}) \le \varepsilon_n$. For part (b), consider $T, \eta, \varepsilon_0 > 0$ and any bounded L^{ε_n}-stopping time $\hat\tau \le T$. The event $|L^{\varepsilon_n}_{\hat\tau+\delta} - L^{\varepsilon_n}_{\hat\tau}| \ge \varepsilon_0$ means that L^{ε_n} performs at least $N_n := \lceil \varepsilon_0/\varepsilon_n \rceil$ jumps in the stochastic interval $\rrbracket \hat\tau, \hat\tau+\delta \rrbracket$. Lemma 7.3.3 yields some n_0 and $\delta_0 = \delta_0(\varepsilon_0)$ such that Aldous' criterion is satisfied for all $n \ge n_0$. Hence, $(L^{\varepsilon_n})_n$ is tight by Proposition 7.3.1. □

Next we show boundedness of $(X^{\varepsilon_n})_n$, needed for Lemma 7.3.6 to prove tightness.

Lemma 7.3.5 (Bounding the diffusion approximations)**.** *Let $T \in (0,\infty)$ and $\eta > 0$. Then there exists $M \in \mathbb{R}$ such that $\mathbb{P}[\sup_{t\in[0,T]} |X_t^{\varepsilon_n}| > M] < \eta$ for all $n \in \mathbb{N}$.*

Proof. By Lemma 7.3.2, for every $n \in \mathbb{N}$ the process X^{ε_n} on $[0,T]$ is bounded from above by a constant M with probability at least $1-\eta/2$. It remains to show that it is also bounded from below with high probability. To this end, we will construct a process Y that is a lower bound for all X^{ε_n} and then argue for Y.

For $\hat\varepsilon := \sup_n \varepsilon_n$ consider a (b,σ)-diffusion Y which is discretely reflected by jumps of size $-\hat\varepsilon$ at a constant boundary $c := g(0)-\hat\varepsilon$, with $Y_0 = y := g(0)-2\hat\varepsilon$. Such Y is a special case of (7.4)–(7.5), for a constant boundary function: $\mathrm{d}Y_t = b(Y_t)\,\mathrm{d}t + \sigma(Y_t)\,\mathrm{d}W_t - L_t^Y$ with $L_t^Y := \sum_{0\le s\le t} \Delta L_t^Y$ and $\Delta L_t^Y := \hat\varepsilon \mathbb{1}_{\{Y_{t-}=c\}}$. Let $\tau_k^Y := \inf\{t > 0 \mid L_t^Y > k\hat\varepsilon\}$ be the k-th hitting time of Y at the boundary c. Thus on all intervals $\llbracket \tau_k^Y, \tau_{k+1}^Y \llbracket$, Y is a continuous (b,σ)-diffusion starting in y. Now for fixed n, $\varepsilon := \varepsilon_n$, note that $X^\varepsilon_{\tau^\varepsilon_{m\varepsilon}} = g((m-1)\varepsilon) - \varepsilon \ge c \ge Y_{\tau^\varepsilon_{m\varepsilon}}$ by monotonicity of g. As $\tau^\varepsilon_{m\varepsilon} \to \infty$ for $m \to \infty$ by Lemma 7.2.1, induction over the inverse (discrete) local times $\tau^\varepsilon_{m\varepsilon}$, $m \in \mathbb{N}$, yields $X^\varepsilon \ge Y$ on $\llbracket \tau_k^Y, \tau_{k+1}^Y \rrbracket$ if $X^\varepsilon_{\tau_k^Y} \ge Y_{\tau_k^Y}$ by comparison results [KS91, Thm. 5.2.18]. Since $X_0^\varepsilon \ge Y_0$, the latter follows by induction over k. As $\tau_k^Y \to \infty$ for $k \to \infty$ by Lemma 7.2.1, we get $X^{\varepsilon_n} \ge Y$ on $[0,\infty)$ for all n. So it suffices to show $\mathbb{P}[\inf_{t\in[0,T]} Y_t < -M] < \eta/2$ for some M, which directly follows from the càdlàg property of Y. □

Lemma 7.3.6 (Tightness of the reflected diffusion approximations)**.** *The sequence $(X^{\varepsilon_n})_n$ of càdlàg processes from* (7.4) *and* (7.5) *satisfies Aldous' criterion and thus is tight.*

Proof. Condition (a) of Proposition 7.3.1 holds. To verify part (b), let $\eta > 0$, $T \in (0,\infty)$, and $\hat\tau \le T$ be a stopping time. By Lemma 7.3.5, $|X^{\varepsilon_n}_{\hat\tau}|$ is with a probability of at least $1-\eta/4$ bounded by some constant M (not depending on n and $\hat\tau$). Let us consider the events $\{X^{\varepsilon_n}_{\hat\tau+\delta} \le X^{\varepsilon_n}_{\hat\tau} - \varepsilon_0\} \cup \{X^{\varepsilon_n}_{\hat\tau+\delta} \ge X^{\varepsilon_n}_{\hat\tau} + \varepsilon_0\} = \{|X^{\varepsilon_n}_{\hat\tau+\delta} - X^{\varepsilon_n}_{\hat\tau}| \ge \varepsilon_0\}$ separately.

1) First consider $\{X^{\varepsilon_n}_{\hat\tau+\delta} \le X^{\varepsilon_n}_{\hat\tau} - \varepsilon_0\}$. For $\xi := X^{\varepsilon_n}_{\hat\tau}$ we construct a reflected process Y^ξ such that $Y_t^\xi \le X^{\varepsilon_n}_{\hat\tau+t}$ for all $t \ge 0$. We then estimate $\mathbb{P}[X^{\varepsilon_n}_{\hat\tau+\delta} \le X^{\varepsilon_n}_{\hat\tau} - \varepsilon_0]$ by means of $\mathbb{P}[Y_\delta^x \le x - \varepsilon_0]$ in (7.15), uniformly for all n large enough. We estimate the latter in (7.16) using the probability of a down-crossing in time δ of intervals $[x-\varepsilon_0, x-2\hat\varepsilon]$ by a

continuous diffusion. Covering $\bigcup_x [x-\varepsilon_0, x-2\hat{\varepsilon}]$ by finitely many intervals $[y_k, y_{k+1}]$ in (7.17) then allows us to choose $\delta > 0$ sufficiently small.

To this end, choose $\hat{\varepsilon} \leq \varepsilon_0/4$ and n large enough such that $\varepsilon_n \leq \hat{\varepsilon}$, and let $(Y_t^\xi)_{t\geq 0}$ be the (b,σ)-diffusion w.r.t. the Brownian motion $(W_{\hat{\tau}+t} - W_{\hat{\tau}})_{t\geq 0}$ with $Y_0^\xi = \xi - 2\hat{\varepsilon}$, which is discretely reflected by jumps of size $-\hat{\varepsilon}$ at a constant boundary at level $\xi - \hat{\varepsilon}$. More precisely, $\mathrm{d}Y_t^\xi = b(Y_t^\xi)\,\mathrm{d}t + \sigma(Y_t^\xi)\,\mathrm{d}W_{\hat{\tau}+t} - K_t^\xi$ with (discrete) local time $K_t^\xi := \sum_{0\leq s\leq t} \Delta K_s^\xi$ for $\Delta K_t^\xi := \hat{\varepsilon}\mathbb{1}_{\{Y_{t-}^\xi = \xi - \hat{\varepsilon}\}}$. Global existence and uniqueness of (Y^ξ, K^ξ) follows from the proof of Lemma 7.2.1. By comparison arguments and induction as in the proof of Lemma 7.3.5, one verifies $Y_t^\xi \leq X_{\hat{\tau}+t}^{\varepsilon_n}$ for $t \in [0,\infty)$. Indeed, [KS91, Theorem 5.2.18] gives $Y_\cdot^\xi \leq X_{\hat{\tau}+\cdot}^{\varepsilon_n}$ on $[\![0,\tau_1[\![$ until the first jump of either $Y_\cdot^\xi$ or $X_{\hat{\tau}+\cdot}^{\varepsilon_n}$ at time $\tau_1 > 0$. If only Y^ξ jumps, we have $Y_{\tau_1}^\xi = Y_{(\tau_1)-}^\xi - \hat{\varepsilon} \leq X_{(\tau_1)-}^{\varepsilon_n} - \hat{\varepsilon} = X_{\tau_1}^{\varepsilon_n} - \hat{\varepsilon}$, but if $X_{\hat{\tau}+\cdot}^{\varepsilon_n}$ jumps, we have $X_{\hat{\tau}+\tau_1}^{\varepsilon_n} = g(L_{(\hat{\tau}+\tau_1)-}^{\varepsilon_n}) - \varepsilon_n \geq g(L_{\hat{\tau}}^{\varepsilon_n}) - \varepsilon_n = \xi \geq Y_{\tau_1}^\xi$. Now $Y_{\tau_1}^\xi \leq X_{\hat{\tau}+\tau_1}^{\varepsilon_n}$, so we get $Y_\cdot^\xi \leq X_{\hat{\tau}+\cdot}^{\varepsilon_n}$ on $[\![\tau_k, \tau_{k+1}[\![$ by induction for all jump times τ_k of $(Y_\cdot^\xi, X_{\hat{\tau}+\cdot}^{\varepsilon_n})$.

Using $Y_\delta^\xi \leq X_{\hat{\tau}+\delta}^{\varepsilon_n}$ and the strong Markov property of Y^ξ w.r.t. $(\mathcal{F}_{\hat{\tau}+t})_{t\geq 0}$, we get

$$\mathbb{P}\big[X_{\hat{\tau}+\delta}^{\varepsilon_n} \leq X_{\hat{\tau}}^{\varepsilon_n} - \varepsilon_0, |X_{\hat{\tau}}^{\varepsilon_n}| \leq M\big] \leq \sup_{-M\leq x\leq M} \mathbb{P}[Y_\delta^x \leq x - \varepsilon_0]\,. \tag{7.15}$$

By construction Y^ξ depends on n and τ (through ξ), while the right-hand side of (7.15) does not. Thus one only needs to bound the probability of an $(\varepsilon_0 - 2\hat{\varepsilon})$-displacement of diffusions Y^x with starting points $x - 2\hat{\varepsilon}$ from a compact set, which are reflected (by $(-\hat{\varepsilon})$-jumps) at constant boundaries $x - \hat{\varepsilon}$. By the arguments in the proof of Lemma 7.3.3 (here applied for Y^x which is reflected at a constant boundary), for $\delta = \delta_0 > 0$ there exists $N \in \mathbb{N}$ with the following property: for every $x \in [-M, M]$, the number $J_\delta^x := \inf\{k \mid k\hat{\varepsilon} \geq K_\delta^x\}$ of jumps of Y^x until time δ is bounded by $N-1$ with probability at least $1 - \eta/8$.

Indeed, by (7.13), fixing $\delta > 0$, $\lambda := 1/\delta$, one gets for any x that $\mathbb{P}[J_\delta^x \geq \lceil N(x)\rceil] \leq \eta/8$ where $N(x) := 1 + \big(\log(\eta/8) - 1\big)/\big(\log\Phi_\lambda(x-\hat{\varepsilon}) - \log\Phi_\lambda(x)\big) \in \mathbb{R}$. Compactness of $[-M, M]$ and continuity of $N(x)$ gives $N := \lceil\sup_{x\in[-M,M]} N(x)\rceil < \infty$. Hence,

$$\sup_{x\in[-M,M]} \mathbb{P}[Y_\delta^x \leq x - \varepsilon_0, J_\delta^x \leq N-1] \leq N \sup_{x\in[-M,M]} \mathbb{P}[H^{x-2\hat{\varepsilon}\,\to\, x-\varepsilon_0} \leq \delta], \tag{7.16}$$

since for the event under consideration, the process Y^x would have to move at least once (in at most N occasions) continuously from the point $x - 2\hat{\varepsilon}$ to $x - \varepsilon_0$. Let $d := (\varepsilon_0 - 2\hat{\varepsilon})/2 \geq \varepsilon_0/4 > 0$, $K := \lfloor 2M/d\rfloor$ and $y_k := kd - M$. For $x \in [y_k, y_{k+1}]$, we have $H^{y_{k-2}\,\to\, y_{k-2}-d} \leq H^{x-\varepsilon_0\,\to\, x-2\hat{\varepsilon}}$ since $[y_{k-2} - d, y_{k-2}] \subset [x - \varepsilon_0, x - 2\hat{\varepsilon}]$, so by $[-M, M] \subset [y_0, y_{K+1}]$ we get

$$\begin{aligned}
\mathbb{P}\big[H^{X_{\hat{\tau}}^{\varepsilon_n} - \varepsilon_n \,\to\, X^{\varepsilon_n} - \varepsilon_0} \leq \delta, |X_{\hat{\tau}}^{\varepsilon_n}| \leq M\big] &\leq \eta/8 + N \sup_{x\in[-M,M]} \mathbb{P}[H^{x-2\hat{\varepsilon}\,\to\, x-\varepsilon_0} \leq \delta] \\
&= \eta/8 + N \max_{k=0,\dots,K} \sup_{x\in[kd-M,(k+1)d-M]} \mathbb{P}\big[H^{x-2\hat{\varepsilon}\,\to\, x-\varepsilon_0} \leq \delta\big] \\
&\leq \eta/8 + N \max_{k=-2,\dots,K} \mathbb{P}\big[H^{y_k\,\to\, y_k - d} \leq \delta\big]\,.
\end{aligned} \tag{7.17}$$

For a sufficiently small $\delta = \delta_1 \in (0, \delta_0]$ the right-hand side of (7.17) can be made smaller than $\eta/4$. The above holds for all n such that $\varepsilon_n \leq \hat{\varepsilon}$, meaning that there is some n_0

such that is holds for all $n \geq n_0$. Note that δ_1 only depends on T (via M and K) and on n_0 but not on n. Hence, for all $\delta \in (0, \delta_1]$, all $n \geq n_0$ and all $\hat{\tau} \leq T$ we have

$$\mathbb{P}[X^{\varepsilon_n}_{\hat{\tau}+\delta} \leq X^{\varepsilon_n}_{\hat{\tau}} - \varepsilon_0] \leq \frac{\eta}{2}\,. \tag{7.18}$$

2) For the alternative second case $X^{\varepsilon_n}_{\hat{\tau}+\delta} \geq X^{\varepsilon_n}_{\hat{\tau}} + \varepsilon_0$, consider the solution $(Y_t)_{t\geq\hat{\tau}}$ on $[\![\hat{\tau}, \infty[\![$ of $\mathrm{d}Y_t = b(Y_t)\,\mathrm{d}t + \sigma(Y_t)\,\mathrm{d}W_t$ with $Y_{\hat{\tau}} = X^{\varepsilon_n}_{\hat{\tau}}$. Using comparison results for continuous diffusions [KS91, Theorem 5.2.18] inductively over times $[\![\tau^{\varepsilon_n}_{(k-1)\varepsilon_n}, \tau^{\varepsilon_n}_{k\varepsilon_n}[\![$, we find $Y_t \geq X^{\varepsilon_n}_t$ for all $t \in [\![\hat{\tau}, \infty[\![$, a.s. Hence, arguing like in the previous case

$$\begin{aligned}\mathbb{P}\big[X^{\varepsilon_n}_{\hat{\tau}+\delta} \geq X^{\varepsilon_n}_{\hat{\tau}} + \varepsilon_0, |X^{\varepsilon_n}_{\hat{\tau}}| \leq M\big] &\leq \mathbb{P}\big[Y_{\hat{\tau}+\delta} \geq Y_{\hat{\tau}} + \varepsilon_0, |Y_{\hat{\tau}}| \leq M\big] \\ &\leq \sup_{-M\leq y\leq M} \mathbb{P}\big[H^{y\,\to\,y+\varepsilon_0} \leq \delta\big]. \end{aligned}\tag{7.19}$$

As in (7.17) we find a $\delta_2 > 0$ such that for all $\delta \in (0, \delta_2]$ the right side of (7.19) is bounded by $\eta/4$. Hence we have $\mathbb{P}[X^{\varepsilon_n}_{\hat{\tau}+\delta} \geq X^{\varepsilon_n}_{\hat{\tau}} + \varepsilon_0] \leq \eta/2$, so with (7.18), Proposition 7.3.1 applies. □

Now, to prove joint tightness of $(X^{\varepsilon_n}, L^{\varepsilon_n})_n$, we can utilize the fact that both processes satisfy Aldous' criterion and that their jump times and jump magnitudes are identical.

Lemma 7.3.7 (Tightness of joint approximations)**.** *The sequence $(X^{\varepsilon_n}, L^{\varepsilon_n})_n$ of càdlàg $\mathbb{R}^2$-valued processes defined by* (7.4) *and* (7.5) *is tight.*

Proof. In view of Proposition 7.3.1, choose the space $E := \mathbb{R}^2$ equipped with Euclidean norm $|\cdot|$ and let $Y^n := (X^{\varepsilon_n}, L^{\varepsilon_n}) \in D\big([0,\infty), E\big)$. Then $Y^n_0 = (g(0) - \varepsilon_n, \varepsilon_n)$ and $J_T(Y^n) = \sqrt{2}\varepsilon_n$ form tight sequences in E and $\mathbb{R}$, respectively. Furthermore,

$$\mathbb{P}\big[|Y^n_{\hat{\tau}+\delta} - Y^n_{\hat{\tau}}| \geq \varepsilon_0\big] \leq \mathbb{P}\Big[|X^{\varepsilon_n}_{\hat{\tau}+\delta} - X^{\varepsilon_n}_{\hat{\tau}}| \geq \frac{\varepsilon_0}{2}\Big] + \mathbb{P}\Big[|L^{\varepsilon_n}_{\hat{\tau}+\delta} - L^{\varepsilon_n}_{\hat{\tau}}| \geq \frac{\varepsilon_0}{2}\Big]\,.$$

Hence Y^n also satisfies Aldous's criterion and therefore is tight. □

Tightness only implies weak convergence of a subsequence. It remains to show (in Lemma 7.3.9) that every limit point satisfies (7.1) and (7.2) and that uniqueness in law holds. The latter will follow from pathwise uniqueness results for SDEs with reflection, while for the former we apply results from [KP96] on weak converges of SDEs. For that purpose, note that the approximated local times form a *good* sequence of semimartingales (cf. [KP96, Definition 7.3]), as shown in the following lemma.

Lemma 7.3.8. *The sequence $(L^{\varepsilon_n})_n$ is of uniformly controlled variation and thus* good*.*

Proof. Let $\delta := \sup_n \varepsilon_n$. Then all processes L^{ε_n} have jumps of size at most $\delta < \infty$. Fix some $\alpha > 0$. By tightness, there exists some $C \in \mathbb{R}$ such that $\mathbb{P}[L^{\varepsilon_n}_\alpha > C] \leq 1/\alpha$. So the stopping time $\tau_{n,\alpha} := \inf\{t \geq 0 \mid L^{\varepsilon_n}_t > C\}$ satisfies $\mathbb{P}[\tau_{n,\alpha} \leq \alpha] = \mathbb{P}[L^{\varepsilon_n}_\alpha > C] \leq 1/\alpha\,.$ Moreover, by monotonicity of L^{ε_n} we have $\mathbb{E}\Big[\int_0^{t\wedge\tau_{n,\alpha}} \mathrm{d}|L^{\varepsilon_n}|_s\Big] = \mathbb{E}[L^{\varepsilon_n}_{t\wedge\tau_{n,\alpha}}] \leq C < \infty\,.$ Hence (L^{ε_n}) is of uniformly controlled variation in the sense of [KP96, Definition 7.5]. So by [KP96, Theorem 7.10] it is a *good* sequence of semimartingales. □

We have gathered all necessary results to prove convergence of our approximating diffusions and local times to the continuous counterpart.

Lemma 7.3.9 (Weak convergence of the approximations)**.** *The sequence* $(X^{\varepsilon_n}, L^{\varepsilon_n})_n$ *of càdlàg processes defined by* (7.4) – (7.5) *converges weakly to the unique continuous strong solution* (X, L) *of* (7.1) – (7.2)*.*

Proof. By Prokhorov's theorem, tightness of $(X^{\varepsilon_n}, L^{\varepsilon_n}, W)_n$ implies weak convergence of a subsequence to some limit point, $(X^{\varepsilon_{n_k}}, L^{\varepsilon_{n_k}}, W)_k \Rightarrow (\tilde{X}, \tilde{L}, \tilde{W}) \in D\big([0,\infty), \mathbb{R}^3\big)$. Continuity of $(\tilde{X}, \tilde{L})$ is clear since $\varepsilon_n \to 0$ is the maximum jump size. First we prove that $(\tilde{X}, \tilde{L})$ satisfies the asserted SDEs. Afterwards, we will prove uniqueness of the limit point. To ease notation, let w.l.o.g. the subsequence (n_k) be (n).

By [KP96, Theorem 8.1] we get that $(\tilde{X}, \tilde{L})$ satisfy (7.1) for the semimartingale $\tilde{W}$. That $\tilde{W}$ is a Brownian motion follows from standard arguments, cf. [NO10, proof of Theorem 1.9]. As $D\big([0,\infty), \mathbb{R}^3\big)$ is separable we find, by an application of the Skorokhod representation theorem, that $\tilde{L}$ is non-decreasing and $\tilde{X}_t \le g(\tilde{L}_t)$ for all $t \ge 0$, $\mathbb{P}$-a.s. because these properties already hold for $(X^{\varepsilon_n}, L^{\varepsilon_n})$.

To prove that $\tilde{L}$ grows only at times t with $\tilde{X}_t = g(\tilde{L}_t)$, we have to approximate the indicator function by continuous functions. For $\delta > 0$ define

$$h_\delta(x,\ell) := \begin{cases} \big(x - g(\ell)\big)/\delta + 1 & \text{for } g(\ell) - \delta \le x \le g(\ell), \\ 1 - \big(x - g(\ell)\big)/\delta & \text{for } g(\ell) \le x \le g(\ell) + \delta, \\ 0 & \text{otherwise,} \end{cases}$$

$$h_0(x,\ell) := \mathbb{1}_{\{x = g(\ell)\}} \text{ and } H^{\delta,n}_t := h_\delta(X^{\varepsilon_n}_t, L^{\varepsilon_n}_t) \text{ and } \tilde{H}^\delta_t := h_\delta(\tilde{X}_t, \tilde{L}_t)\,.$$

For $\delta \searrow 0$ the functions $h_\delta \searrow h_0$ converge pointwise monotonically. Continuity of h_δ implies weak convergence $(H^{\delta,n}, L^{\varepsilon_n}) \Rightarrow (\tilde{H}^\delta, \tilde{L})$. By Lemma 7.3.8, (L^{ε_n}) is a good sequence. So for every $\delta > 0$ the stochastic integrals $\int_0^\cdot H^{\delta,n}_{s-}\,\mathrm{d}L^{\varepsilon_n}_s \Rightarrow \int_0^\cdot \tilde{H}^\delta_{s-}\,\mathrm{d}\tilde{L}_s$ converge weakly. Note that $\mathrm{d}L^{\varepsilon_n}_t = H^{0,n}_{t-}\,\mathrm{d}L^{\varepsilon_n}_t$. Hence, for every $\delta > 0$ we have

$$\int_0^\cdot H^{\delta,n}_{s-}\,\mathrm{d}L^{\varepsilon_n}_s = \int_0^\cdot H^{\delta,n}_{s-} H^{0,n}_{s-}\,\mathrm{d}L^{\varepsilon_n}_s = \int_0^\cdot H^{0,n}_{s-}\,\mathrm{d}L^{\varepsilon_n}_s = L^{\varepsilon_n}\,.$$

By the weak convergence $L^{\varepsilon_n} \Rightarrow \tilde{L}$ it follows for every $\delta > 0$ that $\tilde{L}_t = \int_0^t \tilde{H}^\delta_{s-}\,\mathrm{d}\tilde{L}_s$. By monotonicity of $\tilde{L}$, $\mathrm{d}\tilde{L}_t$ defines a random measure on $[0,\infty)$. Hence monotone convergence of $\tilde{H}^\delta_t \searrow \tilde{H}^0_t$ yields $\mathrm{d}\tilde{L}_t = h_0(\tilde{X}_t, \tilde{L}_t)\,\mathrm{d}\tilde{L}_t$.

Thus, we showed that $(X^\varepsilon, L^\varepsilon)$ converges in distribution to a weak solution $(\tilde{X}, \tilde{L})$ of the reflected SDE, i.e. it might be defined on a different probability space with its own Brownian motion. Note that $(\tilde{X}, \tilde{L})$ is continuous on $[0,\infty)$ and that $\tilde{\tau}_\infty := \sup_k \tilde{\tau}_k = \infty$ a.s., where $\tilde{\tau}_k := \inf\{t > 0 \mid |\tilde{X}_t| \vee \tilde{L}_t > k\}$. To show the existence and uniqueness of a strong solution as stated in the theorem, we will use the results from [DI93]. Consider the domain $\bar{G} := \{(x,\ell) \in \mathbb{R}^2 \mid x \le g(\ell), \ell \ge 0\}$. We may interpret the process (X_t, L_t) as a continuous diffusion in $\bar{G}$ with oblique reflection in direction $(-1,+1)$ at the boundary, although the notion of a two-dimensional reflection seems unusual here, because (X, L) only varies in one dimension in the interior of G. The unbounded domain G can be exhausted by bounded domains $G_k := \big\{(x,\ell) \in G \,\big|\, |x|, |\ell| < k\big\}$, which might have a non-smooth boundary especially at $(g(0), 0)$, but still satisfy [DI93, Cond. (3.2)]. Hence, by [DI93, Cor. 5.2] the process (X, L) exists (up to explosion time) on the initial probability space and is (strongly) unique on $[\![0, \tau_k[\![$ with exit time $\tau_k := \inf\{t > 0 \mid |X_t| \vee L_t > k\}$, for all $k \in \mathbb{N}$. So (X, L) is unique until explosion time $\tau_\infty := \sup_k \tau_k$. Moreover, by [DI93, Theorem 5.1] we have the following pathwise uniqueness result: for any two continuous solutions (X^1, L^1) and (X^2, L^2) with explosion times τ^1_∞ and τ^2_∞, respectively defined on the same probability space with the same Brownian motion and the same initial condition, we have that $X^1 = X^2$ and $L^1 = L^2$ on $[\![0, \tau^1_k \wedge \tau^2_k]\!]$ for every $k \in \mathbb{N}$ a.s. Using a known argument due to Yamada and Watanabe, ideas being as in [KS91, Ch. 5.3.D], one can bring the two (weak) solutions $(\tilde{X}, \tilde{L}, \tilde{W})$ and (X, L, W) to a canonical space with a common Brownian motion. By pathwise uniqueness there, one concludes that $\tau_\infty = \infty$ a.s. (as $\tilde{\tau}_\infty = \infty$). Hence the strong solution (X, L) does not explode in finite time. In addition, we conclude uniqueness in law like in [KS91, Prop. 5.3.20] and thus any weak limit point of the approximating sequence $(X^\varepsilon, L^\varepsilon)$ will have the same law as (X, L). □

This convergence result can be strengthened as follows.

Corollary 7.3.10 (Convergence in probability)**.** *The sequence $(X^{\varepsilon_n}, L^{\varepsilon_n})_n$ of càdlàg processes defined by* (7.4)–(7.5) *converges in probability to* (X, L) *defined by* (7.1)–(7.2).

Proof. Following the proof of [KP91, Cor. 5.6], we will strengthen weak convergence $(X^{\varepsilon_n}, L^{\varepsilon_n}) \Rightarrow (X, L)$ to convergence in probability. First, note that Lemma 7.3.9 implies weak convergence of the triple $(X^{\varepsilon_n}, L^{\varepsilon_n}, W) \Rightarrow (X, L, W)$ by e.g. [SK85, Corollary 3.1]. Hence, for every bounded continuous $F : D([0,\infty); \mathbb{R}^2) \to \mathbb{R}$ and every bounded continuous $G : C([0,\infty); \mathbb{R}) \to \mathbb{R}$, we have $\lim_{n\to\infty} \mathbb{E}[F(X^{\varepsilon_n}, L^{\varepsilon_n})G(W)] = \mathbb{E}[F(X, L)G(W)]$. Now, the previous equation even holds for all bounded measurable G by L^1-approximation of measurable functions by continuous functions. By strong uniqueness of (X, L), there exists a measurable function $H : C([0,\infty); \mathbb{R}) \to C([0,\infty); \mathbb{R}^2)$ such that $(X, L) = H(W)$. In particular, $G(W) := F(H(W)) = F(X, L)$ is bounded and measurable, so we conclude

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}\big[(F(X^{\varepsilon_n}, L^{\varepsilon_n}) - F(X, L))^2\big] \\
&\qquad = \lim_{n\to\infty} \big(\mathbb{E}\big[F(X^{\varepsilon_n}, L^{\varepsilon_n})^2\big] - 2\mathbb{E}\big[F(X^{\varepsilon_n}, L^{\varepsilon_n})F(X, L)\big] + \mathbb{E}\big[F(X, L)^2\big]\big) = 0
\end{aligned}$$

and hence convergence in probability follows.

To show this, first consider $D([0, T])$ with Skorokhod-metric d_T, instead of $D([0,\infty))$, by restricting paths on $[0, T]$. Fix an arbitrary $\eta > 0$ and write $Y^n := (X^{\varepsilon_n}, L^{\varepsilon_n})$ and $Y := (X, L)$. Since $\mathbb{P}[\|Y\|_\infty \le K] \to 1$ and $\mathbb{P}[\|Y^n\|_\infty \le K] \to 1$ as $K \to \infty$, we have that Y, Y^n are bounded with high probability.

By compactness of $[0, T] \times [-K, K]^2$ and separability of $D([0, T])$ (cf. e.g. [Bil99, Thm. 12.2]), there exists a finite covering of $D([0, T]; [-K, K]^2)$ with balls $B_{\eta/4}(q)$. For each of these finitely many $q \in D([0, T])$ let $F_q : D([0, T]) \to [0, 1]$ be a continuous function with value 1 on $B_{\eta/4}(q)$ and 0 outside of $B_{\eta/2}(q)$.

For two bounded paths $x, y \in D([0, T]; [-K, K]^2)$ we have that $d_T(x, y) > \eta$ implies $(F_q(x) - F_q(y))^2 = 1$ for at least one of the chosen q. We can therefore estimate the probability $\mathbb{P}[d_T(Y, Y^n) \ge \eta, \|Y\| \vee \|Y^n\| \le K] \le \sum_q \mathbb{E}[(F_q(Y) - F_q(Y^n))^2]$. Since the finite sum on the right-hand side converges to 0 as $n \to \infty$, we get that $d_T(Y, Y^n) < \eta$ with high probability. By repeating this argument for multiple time horizons T, we can also bound the $D([0,\infty); \mathbb{R}^2)$-distance $d(Y, Y^n) = \sum_{T=1}^\infty 2^{-T}(1 \wedge d_T(Y, Y^n)) \le \eta$ with high probability. □

Corollary 7.3.11 (Weak convergence of the inverse local times)**.** *For any $\ell > 0$, the sequence $(\tau^{\varepsilon_n}_\ell)_n$ from* (7.6) *converges in law to the inverse local time τ_ℓ defined by* (7.3).

Proof. Convergence $L^{\varepsilon_n} \Rightarrow L$ implies $L_t^{\varepsilon_n} \Rightarrow L_t$ at all continuity points of L, i.e. at all points, hence $\mathbb{P}\big[\tau_\ell^{\varepsilon_n} \leq t\big] = \mathbb{P}\big[L_t^{\varepsilon_n} \geq \ell\big] \to \mathbb{P}[L_t \geq \ell] = \mathbb{P}[\tau_\ell \leq t]\,.$ □

This completes the proof of Theorem 7.2.2.

Bibliography

[AC00] Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *J. Risk*, 3(2):5–39, 2000.

[AFS10] Aurélien Alfonsi, Antje Fruth, and Alexander Schied. Optimal execution strategies in limit order books with general shape functions. *Quant. Finance*, 10(2):143–157, 2010.

[AKS16] Aurélien Alfonsi, Florian Klöck, and Alexander Schied. Multivariate transient price impact and matrix-valued positive definite functions. *Math. Oper. Res.*, 41(3):914–934, 2016.

[Alm12] Robert Almgren. Optimal trading with stochastic liquidity and volatility. *SIAM J. Financial Math.*, 3(1):163–181, 2012.

[AS10] Aurélien Alfonsi and Alexander Schied. Optimal trade execution and absence of price manipulations in limit order book models. *SIAM J. Financial Math.*, 1(1):490–522, 2010.

[ASS12] Aurélien Alfonsi, Alexander Schied, and Alla Slynko. Order book resilience, price manipulation, and the positive portfolio problem. *SIAM J. Financial Math.*, 3(1):511–533, 2012.

[AW95] Joseph Abate and Ward Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA J. Comput.*, 7(1):36–43, 1995.

[BB04] Peter Bank and Dietmar Baum. Hedging and portfolio optimization in financial markets with a large trader. *Math. Finance*, 14(1):1–18, 2004.

[BBF18a] Dirk Becherer, Todor Bilarev, and Peter Frentrup. Approximating diffusion reflections at elastic boundaries. *Electron. Commun. Probab.*, 23:1–12, 2018.

[BBF18b] Dirk Becherer, Todor Bilarev, and Peter Frentrup. Optimal asset liquidation with multiplicative transient price impact. *Appl. Math. Optim.*, 78(3):643–676, 2018.

[BBF18c] Dirk Becherer, Todor Bilarev, and Peter Frentrup. Optimal liquidation under stochastic liquidity. *Finance Stoch.*, 22(1):39–68, 2018.

[BBF19] Dirk Becherer, Todor Bilarev, and Peter Frentrup. Stability for gains from large investors' strategies in M_1/J_1 topologies. *Bernoulli*, 25(2):1105–1140, 2019.

[BC67] John Bather and Herman Chernoff. Sequential decisions in the control of a space-ship (finite fuel). *J. Appl. Probab.*, 4(3):584–604, 1967.

[BE08] Erhan Bayraktar and Masahiko Egami. An analysis of monotone follower problems for diffusion processes. *Math. Oper. Res.*, 33(2):336–350, 2008.

[Bil99] Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999.

[BL98] Dimitris Bertsimas and Andrew W. Lo. Optimal control of execution costs. *J. Financial Markets*, 1(1):1–50, 1998.

[BLZ16] Bruno Bouchard, Grégoire Loeper, and Yiyi Zou. Almost-sure hedging with permanent price impact. *Finance Stoch.*, 20(3):741–771, 2016.

[BR17] Tilmann Blümmel and Thorsten Rheinländer. Financial markets with a large trader. *Ann. Appl. Probab.*, 27(6):3735–3786, 2017.

[BS02] Andrei N. Borodin and Paavo Salminen. *Handbook of Brownian motion—facts and formulae.* Probability and its Applications. Birkhäuser Verlag, Basel, second edition, 2002.

[CdL13] Rama Cont and Adrien de Larrard. Price dynamics in a Markovian limit order market. *SIAM J. Financial Math.*, 4(1):1–25, 2013.

[ÇJP04] Umut Çetin, Robert A. Jarrow, and Philip E. Protter. Liquidity risk and arbitrage pricing theory. *Finance Stoch.*, 8(3):311–341, 2004.

[CL95] Louis K. C. Chan and Josef Lakonishok. The behavior of stock prices around institutional trades. *J. Finance*, 50(4):1147–1174, 1995.

[ÇST10] Umut Çetin, H. Mete Soner, and Nizar Touzi. Option hedging for small investors under liquidity costs. *Finance Stoch.*, 14(3):317–341, 2010.

[CT04] Rama Cont and Peter Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004.

[DAF14] Tiziano De Angelis and Giorgio Ferrari. A stochastic partially reversible investment problem on a finite time-horizon: Free-boundary analysis. *Stochastic Process. Appl.*, 124(12):4080 – 4119, 2014.

[DAFF17] Tiziano De Angelis, Salvatore Federico, and Giorgio Ferrari. Optimal boundary surface for irreversible investment with stochastic costs. *Math. Oper. Res.*, 42(4):1135–1161, 2017.

[DI93] Paul Dupuis and Hitoshi Ishii. SDEs with oblique reflection on nonsmooth domains. *Ann. Probab.*, 21(1):554–580, 1993.

[DM82] Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential. B.* North-Holland, Amsterdam, 1982.

[DM04] François Dufour and Boris Miller. Singular stochastic control problems. *SIAM J. Control Optim.*, 43(2):708–730, 2004.

[DS98] Freddy Delbaen and Walter Schachermayer. The fundamental theorem of asset pricing for unbounded stochastic processes. *Math. Ann.*, 312(2):215–250, 1998.

[DY09] Min Dai and Fahuai Yi. Finite-horizon optimal investment with transaction costs: A parabolic double obstacle problem. *J. Differential Equations*, 246(4):1445–1469, 2009.

[DZ98] Mark H. A. Davis and Mihail Zervos. A pair of explicitly solvable singular stochastic control problems. *Appl. Math. Optim.*, 38(3):327–352, 1998.

[EKK91] Nicole El Karoui and Ioannis Karatzas. A new approach to the Skorohod problem, and its applications. *Stochastics Stochastics Rep.*, 34(1–2):57–82, 1991.

[FGL+04] J. Doyne Farmer, László Gillemot, Fabrizio Lillo, Szabolcs Mike, and Anindya Sen. What really causes large price changes? *Quant. Finance*, 4(4):383–397, 2004.

[FJ02] J. Doyne Farmer and Shareen Joshi. The price dynamics of common trading strategies. *J. Econ. Behav. Organ.*, 49(2):149–171, 2002.

[FK19] Giorgio Ferrari and Torben Koch. An optimal extraction problem with price impact. *Appl. Math. Optim.*, 2019. DOI: 10.1007/s00245-019-09615-9.

[FKTW12] Peter A. Forsyth, J. Shannon Kennedy, Sue T. Tse, and Heath Windcliff. Optimal trade execution: A mean quadratic variation approach. *J. Econom. Dynam. Control*, 36(12):1971–1991, 2012.

[FP14] Salvatore Federico and Huyên Pham. Characterization of the optimal boundaries in reversible investment problems. *SIAM J. Control Optim.*, 52(4):2180–2223, 2014.

[FR19] Giorgio Ferrari and Neofytos Rodosthenous. Optimal control of debt-to-gdp ratio in an n-state regime switching economy. Technical report, Bielefeld Center for Mathematical Economics, 2019.

[Fre98] Rüdiger Frey. Perfect option hedging for a large trader. *Finance Stoch.*, 2(2):115–141, 1998.

[FSU19] Antje Fruth, Torsten Schöneborn, and Mikhail Urusov. Optimal trade execution in order books with stochastic liquidity. *Mathematical Finance*, 29(2):507–541, 2019.

[GF00] Israel M. Gelfand and Sergei V. Fomin. *Calculus of Variations*. Dover Books on Mathematics. Dover Publications, 2000.

[GH17] Paulwin Graewe and Ulrich Horst. Optimal trade execution with instantaneous price impact and stochastic resilience. *SIAM J. Control Optim.*, 55(6):3707–3725, 2017.

[GHS16] Paulwin Graewe, Ulrich Horst, and Eric Séré. Smooth solutions to portfolio liquidation problems under price-sensitive market impact. *Stoch. Process. Appl.*, 2016.

[GS13] Jim Gatheral and Alexander Schied. Dynamical models of market impact and algorithms for order execution. In *Handbook on Systemic Risk*, pages 579–602. Cambridge University Press, 2013.

[GSS12] Jim Gatheral, Alexander Schied, and Alla Slynko. Transient linear price impact and Fredholm integral equations. *Math. Finance*, 22(3):445–474, 2012.

[GZ15] Xin Guo and Mihail Zervos. Optimal execution with multiplicative price impact. *SIAM J. Financial Math.*, 6(1):281–306, 2015.

[HH11] Vicky Henderson and David Hobson. Optimal liquidation of derivative portfolios. *Math. Finance*, 21(3):365–382, 2011.

[HHK92] Ayman Hindy, Chi-Fu Huang, and David Kreps. On intertemporal preferences in continuous time: The case of certainty. *J. Math. Econom.*, 21(5):401–440, 1992.

[HM05] Hua He and Harry Mamaysky. Dynamic trading policies with price impact. *J. Econom. Dynam. Control*, 29(5):891–930, 2005.

[Jar94] Robert A. Jarrow. Derivative security markets, market manipulation, and option pricing theory. *J. Financ. Quant. Anal.*, 29(2):241–261, 1994.

[JJZ08] Andrew Jack, Timothy C. Johnson, and Mihail Zervos. A singular control model with application to the goodwill problem. *Stochastic Process. Appl.*, 118(11):2098–2124, November 2008.

[JS03] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, second edition, 2003.

[Kal00] Jan Kallsen. Optimal portfolios for exponential Lévy processes. *Math. Methods Oper. Res.*, 51(3):357–374, 2000.

[Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[Kar85] Ioannis Karatzas. Probabilistic aspects of finite-fuel stochastic control. *P. Natl. Acad. Sci. USA*, 82(17):5579–5581, 1985.

[Kar13] Constantinos Kardaras. On the closure in the Emery topology of semimartingale wealth-process sets. *Ann. Appl. Probab.*, 23(4):1355–1376, 2013.

[Kat14] Takashi Kato. An optimal execution problem with market impact. *Finance Stoch.*, 18(3):695–732, 2014.

[Kle08] Achim Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer London, 2008.

[KMS17] Olga Klein, Ernst Maug, and Christoph Schneider. Trading strategies of corporate insiders. *J. Financial Markets*, 34:48–68, 2017.

[Kob93] T. Ø. Kobila. A class of solvable stochastic investment problems involving singular controls. *Stochastics Stochastics Rep.*, 43(1-2):29–63, 1993.

[Koc19] Torben Koch. Universal bounds and monotonicity properties of ratios of Hermite and parabolic cylinder functions. *P. Am. Math. Soc.*, 2019. Forthcoming.

[KOWZ00] Ioannis Karatzas, Daniel Ocone, Hui Wang, and Mihail Zervos. Finite-fuel singular control with discretionary stopping. *Stochastics*, 71(1-2):1–50, 2000.

[KP91] Thomas G. Kurtz and Philip E. Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.*, 19(3):1035–1070, 1991.

[KP96] Thomas G. Kurtz and Philip E. Protter. Weak convergence of stochastic integrals and differential equations. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 1–41. Springer, Berlin, 1996.

[KP10] Idris Kharroubi and Huyên Pham. Optimal portfolio liquidation with execution cost and risk. *SIAM J. Financial Math.*, 1(1):897–931, 2010.

[KPP95] Thomas G. Kurtz, Étienne Pardoux, and Philip E. Protter. Stratonovich stochastic differential equations driven by general semimartingales. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(2):351–377, 1995.

[KS84] I. Karatzas and S. Shreve. Connections between optimal stopping and singular stochastic control i. monotone follower problems. *SIAM J. Control Optim.*, 22(6):856–877, 1984.

[KS86] Ioannis Karatzas and Steven E. Shreve. Equivalent models for finite-fuel stochastic control. *Stochastics*, 18(3-4):245–276, 1986.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 1991.

[Kum80] Sadatoshi Kumagai. An implicit function theorem: Comment. *J. Optim. Theory Appl.*, 31(2):285–288, 1980.

[KW01] Ioannis Karatzas and Hui Wang. Connections between bounded-variation control and Dynkin games. In *Optimal control and partial differential equations*, pages 363–373. IOS, Amsterdam, 2001.

[Kyl85] Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

[Leb72] Nikolaĭ N. Lebedev. *Special Functions and Their Applications.* Dover Books on Mathematics. Dover Publications, 1972.

[Led16] Sean Ledger. Skorokhod's M1 topology for distribution-valued processes. *Electron. Commun. Probab.*, 21(34):11 pp., 2016.

[LN19] Charles-Albert Lehalle and Eyal Neuman. Incorporating signals into optimal trading. *Finance Stoch.*, 23(2):275–311, 2019.

[Løk12] Arne Løkka. Optimal execution in a multiplicative limit order book. *Preprint, London School of Economics*, 2012.

[Løk14] Arne Løkka. Optimal liquidation in a limit order book for a risk-averse investor. *Math. Finance*, 24:696–727, 2014.

[LS84] Pierre-Louis Lions and Alain-Sol Sznitman. Stochastic differential equations with reflecting boundary conditions. *Comm. Pure Appl. Math.*, 37(4):511–537, 1984.

[LS13] Christopher Lorenz and Alexander Schied. Drift dependence of optimal trade execution strategies under transient price impact. *Finance Stoch.*, 17(4):743–770, 2013.

[Mar81] Steven I. Marcus. Modeling and approximation of stochastic differential equations driven by semimartingales. *Stochastics*, 4(3):223–245, 1980/81.

[NO10] Kaj Nyström and Thomas Önskog. The Skorohod oblique reflection problem in time-dependent domains. *Ann. Probab.*, 38(6):2170–2223, 2010.

[OW13] Anna Obizhaeva and Jiang Wang. Optimal trading strategy and supply/demand dynamics. *J. Financial Markets*, 16:1–32, 2013.

[Pro05] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[PS06] Goran Peskir and Albert Shiryaev. *Optimal stopping and free-boundary problems*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2006.

[PSS11] Silviu Predoiu, Gennady Shaikhet, and Steven Shreve. Optimal execution in a general one-sided limit-order book. *SIAM J. Financial Math.*, 2(1):183–212, 2011.

[PTW07] Guodong Pang, Rishi Talreja, and Ward Whitt. Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surv.*, 4:193–267, 2007.

[PW10] Guodong Pang and Ward Whitt. Continuity of a queueing integral representation in the M_1 topology. *Ann. Appl. Probab.*, 20(1):214–237, 2010.

[PY03] Jim Pitman and Marc Yor. Hitting, occupation and inverse local times of one-dimensional diffusions: Martingale and excursion approaches. *Bernoulli*, 9(1):1–24, 2003.

[Roc11] Alexandre F. Roch. Liquidity risk, price impacts and the replication problem. *Finance Stoch.*, 15(3):399–419, 2011.

[RS13] Alexandre F. Roch and H. Mete Soner. Resilient price impact of trading and the cost of illiquidity. *Int. J. Theor. Appl. Finance*, 16(6):1350037 (27 pages), 2013.

[RW87] L. Chris. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1987. Itô calculus.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften.* Springer, Berlin, 1999.

[Sam65] Paul A. Samuelson. Rational theory of warrant pricing. *Ind. Manag. Rev.*, 6(2):13–32, 1965.

[Sch13] Alexander Schied. Robust strategies for optimal order execution in the Almgren-Chriss framework. *Appl. Math. Finance*, 20(3):264–286, 2013.

[SK85] Ioana Schiopu-Kratina. Tightness of pairs of tight càdlàg processes. *Stochastic Process. Appl.*, 21(1):167–177, 1985.

[Sko56] Anatoli V. Skorokhod. Limit theorems for stochastic processes. *Theory Probab. Appl.*, 1(3):261–290, 1956.

[Sko61] Anatoli V. Skorokhod. Stochastic equations for diffusion processes in a bounded region. *Teoriya Veroyatnostej i Ee Primeneniya*, 6:287–298, 1961.

[SLG84] S. E. Shreve, J. P. Lehoczky, and D. P. Gaver. Optimal consumption for general diffusions with absorbing and reflecting barriers. *SIAM J. Control Optim.*, 22(1):55–75, 1984.

[SZ17] Alexander Schied and Tao Zhang. A state-constrained differential game arising in optimal portfolio liquidation. *Math. Finance*, 27(3):779–802, 2017.

[Tak97] Michael I. Taksar. Infinite-dimensional linear programming approach to singular stochastic control. *SIAM J. Control Optim.*, 35(2):604–625, 1997.

[Tan79] Hiroshi Tanaka. Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math. J.*, 9(1):163–177, 1979.

[Wal98] Wolfgang Walter. *Ordinary differential equations*, volume 182 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1998. Translated from the sixth German (1996) edition by Russell Thompson, Readings in Mathematics.

[WG03] Amy R. Ward and Peter W. Glynn. A diffusion approximation for a Markovian queue with reneging. *Queueing Syst.*, 43(1–2):103–128, 2003.

[Whi02] Ward Whitt. *Stochastic-process limits.* Springer Series in Operations Research. Springer, New York, 2002.

[WZ65] Eugene Wong and Moshe Zakai. On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Statist.*, 36(5):1560–1564, 1965.

[Zhu92] Hang Zhu. Generalized solution in singular stochastic control: the nondegenerate problem. *Appl. Math. Optim.*, 25(3):225–245, 1992.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den   Peter Frentrup