Inverse Optimal Stopping and Optimal Closure of Illiquid Positions

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch–Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Thomas Kruse

aus

Paderborn

Bonn, 2013

Angefertigt mit der Genehmigung der Mathematisch–Naturwissenschaftlichen–Fakultät
der Rheinischen Friedrich–Wilhelms–Universität Bonn.

1. Gutachter: Prof. Dr. Stefan Ankirchner
2. Gutachter: Prof. Dr. Monique Jeanblanc

Tag der Promotion: 31.01.2014
Erscheinungsjahr: 2014

Abstract

Many economic situations are modeled as stopping problems. Examples include job search, pricing of American options, timing of market entry and irreversible investment decisions. The first part of the thesis analyzes optimal stopping in a dynamic mechanism design framework. It deals with a principal-agent problem where the principal and the agent have different preferences over stopping rules. The agent privately observes a one-dimensional Markov process that influences her payoff. Based on her observation the agent decides when to stop. In order to induce the agent to employ a different stopping rule the principal commits to a transfer that depends only on the time the agent stopped. The goal is to characterize the set of stopping rules that can be implemented using such a transfer.

To this end the well-known single crossing condition from static mechanism design is transferred to optimal stopping problems. In a discrete-time framework it is shown that under this condition a stopping rule is implementable if and only if it is of cut-off type. If time is continuous, a cut-off rule is implementable provided that the associated threshold satisfies certain regularity assumptions. The transfer admits a closed form representation based on the reflected version of the underlying Markov process. This allows for a direct verification argument which in discrete-time draws on Bellman's dynamic programming principle. In the continuous-time framework the solution method solely relies on probabilistic techniques such as comparison principles for reflected stochastic differential equations.

A uniqueness result for the transfer is provided. As a consequence one obtains a new nonlinear integral equation characterizing the optimal stopping boundary in one-dimensional stopping problems.

The results are applied in the context of job search with and without recall. The set of time-dependent reservation wage policies of a risk-averse job seeker that can be induced by a transfer of the unemployment agency, is characterized.

The second part of this thesis analyzes the problem of how to close a large asset position in an illiquid market. The first goal is to characterize trading strategies that make very high liquidation costs unlikely. To this end a model that allows for a *price-sensitive* closure of the position is set-up. It provides a simple device for designing and controlling the distribution of the revenues/costs from unwinding the position. The risk inherent in the open position is modeled by a functional that can be interpreted as the time-average of the squared value-at-risk of the open position. Market illiquidity is reflected by a linear, temporary price impact. The stochastic control problem consists of minimizing a weighted sum of the execution costs and the risk functional.

By appealing to dynamic programming, semi-explicit formulas for the optimal execution strategies are derived in a discrete-time framework. Within the continuous-time version of the model the optimal trading rates can be characterized in terms of a partial differential equation (PDE) describing by how much they differ from the optimal risk-neutral trading rate. The PDE possesses a singularity and does not, in general, have a closed-form solution. A uniqueness result for solutions in the viscosity sense is provided, allowing in the following to identify the value function and optimal trading rates. It is

shown that optimal strategies from the discrete model converge to the continuous-time optimal trading rates. A numerical algorithm is presented which approximates optimal execution rates as functions of the price. The convergence of the algorithm is verified by deriving explicit error bounds. Examples for the liquidation of forward positions in illiquid energy markets illustrate the efficiency of the algorithm.

In the next step this model is generalized by incorporating a stochastic price impact. The liquidation constraint is relaxed by introducing a set of scenarios where the position does not have to be closed. A purely probabilistic solution of this not necessarily Markovian control problem is provided by means of a backward stochastic differential equation (BSDE). The BSDE in this problem possesses a *singular terminal condition*. It is shown that a minimal supersolution of the BSDE exists - a result which partly generalizes existence results obtained by Popier in [69] and [70]. The verification step is based on a penalization argument. Special cases for which the control problem has explicit solutions are discussed. The set of model specifications where optimal trading strategies are deterministic, is characterized.

Finally, the impact of a cross-hedging opportunity on liquidation strategies is analyzed. Suppose there is an open position to be closed in an illiquid forward market (e.g. a commodity market) before delivery. The liquidity of the asset increases as the delivery date approaches. Therefore, an early closure eliminates the risk inherent in the open position but also omits the opportunity of reducing execution costs. Assume further that there is a proxy market where forwards of a correlated asset are traded. Liquidity in the proxy market is high and thus performing a cross-hedge reduces execution costs. However, since the prices are not perfectly correlated, this hedging strategy entails basis risk. Using techniques from singular stochastic control theory allows to obtain an optimal trade-off between execution costs and basis risk. The two-dimensional hedging problem is reduced to a family of stopping problems. Explicit optimal hedging strategies for simple liquidity dynamics are derived.

Contents

1. Introduction

This thesis consists of two independent parts. Each part has a separate introduction.

Part I: Inverse Optimal Stopping

In an optimal stopping problem a decision maker chooses a time when to carry out a certain action in order to maximize her reward. The stopping decision is irreversible and future rewards are uncertain. Hence, stopping today implies losing the option to stop later with a potentially larger return. Optimal stopping problems originated in the work of Wald [84] on sequential analysis - a method to determine when to stop sampling new data in a statistical experiment. Subsequently, optimal stopping theory has been highly influential in the economics literature. In labor economics, the seminal contributions of [80] and [55] established the perspective on job search as an optimal stopping problem. In finance the pricing of American options and other financial contracts is a classical optimal stopping problem (cf. [57]). Following [56] the optimal timing of irreversible investments and market entry decisions are modeled as stopping problems in the industrial organization literature (cf. [22]).

The mathematical literature on optimal stopping theory is very rich. The books [67], [78] and [10] present comprehensive introductions, historical notes and references. If the underlying stochastic process is *Markovian*, there is an analytical approach for obtaining optimal stopping rules. It is based on the following reasoning: Under the Markov property the future evolution of the process depends only on its present value but not on its past. Intuitively, it follows that at every point in time the decision of whether to stop or to continue can be based solely on the process' current value. The state space of the process thus decomposes into two regions. There is the stopping region consisting of all points where it is optimal to stop, and there is its complement, the continuation region. An optimal stopping rule is then given as the first time when the process enters the stopping region. Bellman's optimality principle allows for an identification of the stopping region. It can be characterized by the free boundary of a partial differential equation (PDE) in variational form.

The first part of the thesis considers optimal stopping in a mechanism design framework. It addresses the question of how to modify the payoff in an optimal stopping problem such that a *given* stopping rule becomes optimal. We focus on simple modifications that consist of adding a time-dependent function to the original payoff. Formally, let $\mathbb{T}$ be the index set representing the points in time when stopping is allowed (take e.g. $\mathbb{T} = [0, T]$). Let X denote the underlying Markov process and $g : \mathbb{T} \times \mathbb{R} \to \mathbb{R}$ the agent's payoff function depending on time and the value of the process. The goal is to find a time-dependent function $\pi : \mathbb{T} \to \mathbb{R}$ such that a given stopping rule τ^* is optimal in the maximization problem

$$E[g(\tau, X_\tau) + \pi(\tau)] \to \max, \tag{1.1}$$

where the maximum is taken over all stopping rules with values in $\mathbb{T}$. A stopping rule for which such a function exists is called *implementable*. The function π is called a *transfer*. Thus, the aim is to find an objective functional such that a given outcome becomes optimal. Therefore, the problem is called *inverse* optimal stopping problem.

The model can be understood as a dynamic mechanism design or dynamic principal-agent model. For an introduction to principal-agent problems and mechanism design see e.g. Chapters 14 and 23 of [54]. There is an agent who privately observes a stochastic process and chooses a stopping rule. The principal observes the stopping decision of the agent, but not the realization of the process. In order to influence the agent's stopping decision the principal commits to a transfer - a payment which is due at the moment when the agent stops. In this way she aims at inducing the agent to take a particular stopping decision. For example, the agent could be an unemployed worker who receives job offers until she stops the process and accepts an offer. The principal could be the unemployment agency that wants the agent to accept certain offers, but that does not observe the offers that the agent receives. By designing unemployment benefits that only depend on the time the worker has been unemployed the agency aims at influencing the worker's search behavior. Alternatively, the agent could be a firm that has developed a new technology and now has to decide when to introduce it to the market place. The firm observes private signals regarding the demand, and this knowledge changes over time. The principal is a social planner who also takes the consumer surplus of the new technology into account and hence prefers a different stopping decision than the firm. The first part of the thesis analyzes how the planner can align the preferences of the firm by subsidizing the market entry through a transfer.

The approach of this thesis differs from the approach taken in other papers on dynamic mechanism design. Here, the focus is on simple mechanisms where the only information the principal receives is the agent's stopping decision. In particular, there is no communication between the two parties. Enlarging the set of mechanism by allowing for communication leads to more complex optimization problems, since optimal communication strategies are in general not necessarily Markovian. This approach, nevertheless, has been successfully used in a discrete-time setting in [11] for welfare maximization, in [63] for revenue maximization, or in continuous-time to solve principal-agent problems in [74] and [86]. It turns out that for optimal stopping problems the focus on simple mechanisms without communication is not restrictive. In Chapter 2 it is shown that a principal-optimal stopping rule is always implementable by a transfer that only conditions on time.

Both the discrete-time version (Chapter 2) and the continuous-time version (Chapter 3) of the model are considered. In the discrete-time part stopping is only allowed at finitely many points in time (e.g. $\mathbb{T} = \{0, 1, \ldots, T\}$). In this case it is shown how to solve inverse optimal stopping problems for a general class of one-dimensional Markov processes satisfying only mild regularity assumptions. When stopping is allowed at any point of a finite time period ($\mathbb{T} = [0, T]$) the analysis becomes mathematically more challenging. The main properties of reflected stochastic differential equations are worked out in order to transfer the results from discrete to continuous time.

Let us first outline the results in the discrete-time case which is based on [50]. The main theorem shows that under a dynamic single crossing condition all cut-off rules are implementable. A cut-off rule τ_b is a strategy that stops at the first time when the value of the process exceeds a deterministic, time-dependent threshold $b : \mathbb{T} \to \mathbb{R}$,

$$\tau_b = \inf\{t \in \mathbb{T} | X_t \geq b(t)\}.$$

The dynamic single crossing condition is an assumption on the model parameters ensuring that the expected gain from continuing one unit of time is the higher the smaller the value of the process. Formally, assume that the function

$$x \mapsto z(t, x) = E[g(t + 1, X_{t+1}) | X_t = x] - g(t, x) \tag{1.2}$$

is decreasing.

To get an intuition why cut-off rules are implementable, consider a point in time and a value of the process where it is optimal to continue. The dynamic single crossing condition implies that continuing is even better for lower values of the process. Thus, the stopping region lies above the continuation region at every point in time. To implement a cut-off rule, it consequently suffices to ensure that continuing is as good as stopping whenever the process is on the threshold. Taking the future values of the transfer as given, the present value of the transfer could be calculated recursively. But as future values of the transfer are endogenous, this backward iteration requires the calculation of the continuation value at every point in time. It is shown how to circumvent this indirect approach by deriving a closed form representation of the transfer. To this end the constrained version $\tilde{X}$ of the underlying process is introduced. This process is defined as the unique Markov process that has the same dynamics as X below the threshold b but is required to stay at b whenever X exceeds it. In the discrete-time setting of Chapter 2 one can directly construct the transition probabilities $\tilde{P}_{t,t+1}$ of the constrained process by modifying the transition probabilities of the original process. This construction allows to define the candidate transfer

$$\pi(t) = \sum_{s=t}^{T-1} (\tilde{P}_{t,s} z)(t, b(t)).$$

In Chapter 2 it is verified that this transfer indeed implements the cut-off stopping rule τ_b.

Remarkably, the converse implication of this result holds true as well. Under the dynamic single crossing condition only cut-off rules are implementable. Consequently, the set of implementable stopping rules coincides with the set of cut-off strategies.

The applicability of the results is illustrated in the context of job-search with and without recall. Here cut-off stopping rules are called reservation wage policies: A job seeker accepts the first offer whose wage exceeds a certain threshold - the reservation wage. The goal is to characterize the set of reservation wage policies that an unemployment agency can implement through unemployment benefits that only condition on time. An explicit representation of the benefits is derived.

The notion of implementability of stopping rules generalizes the notion of optimality. Clearly, a stopping rule that is optimal for the agent in the first place, can be implemented by the zero transfer $\pi(t) = 0$ for all t. It is shown that a transfer implementing a cut-off rule is unique up to an additive constant. As a consequence one obtains a new probabilistic characterization of stopping boundaries for one-dimensional, discrete-time optimal stopping problems. A cut-off stopping rule τ_b is optimal in the stopping problem $\sup_\tau E[g(\tau, X_\tau)]$ if and only if the associated threshold satisfies

$$\sum_{s=t}^{T-1} (\tilde{P}_{t,s} z)(t, b(t)) = 0 \tag{1.3}$$

for all $t \in \mathbb{T}$. Equation 1.3 can be solved via a backward recursion, and thus provides a new technique to numerically compute optimal stopping rules. This may, for example, be used in the pricing of Bermudan options.

The solution method for inverse optimal stopping problems in continuous-time draws inspiration from the results from the discrete-time framework. Chapter 3, which is a revised version of [49], considers a time-inhomogeneous Brownian diffusion X as underlying stochastic process. Its infinitesimal generator is denoted by $\mathcal{L}$. The dynamic single crossing condition (1.2) translates into the assumption that the function $x \mapsto (\partial_t + \mathcal{L})g(t, x)$ is decreasing. The question whether a cut-off rule τ_b is implementable is closely linked to the existence of a solution $\tilde{X}$ to a stochastic differential equation with reflection at the threshold b. In the continuous-time setting the version $\tilde{X}$ of X that is reflected at b plays the same role as the constrained version of X in the discrete-time framework. In contrast to Chapter 2 the construction of $\tilde{X}$ is not straightforward and requires some regularity assumptions on the threshold b. The main result shows that all cut-off stopping rules τ_b are implementable provided that the threshold b is càdlàg and has summable downward jumps. Furthermore, it is shown that the transfer π implementing τ_b admits the following closed form representation

$$\pi(t) = \mathbb{E}\left[\int_t^T (\partial_t + \mathcal{L})g(s, \tilde{X}_s^{t,b(t)})ds\right]. \tag{1.4}$$

Here $\left(\tilde{X}_s^{t,b(t)}\right)_{s\geq t}$ denotes the unique process starting on the threshold $b(t)$ at time t which results from reflecting the original process X at b. The existence and uniqueness of $\tilde{X}$ for time-dependent thresholds with jumps is based on [72]. The verification argument solely employs probabilistic arguments. Comparison principles for the original process X and its reflected version $\tilde{X}$ are derived. Together with the single crossing condition they allow to verify that the transfer π from Equation (1.4) indeed implements the cut-off rule τ_b. This approach requires only weak regularity assumptions on the model parameters.

In particular, the threshold b is allowed to have jumps. Moreover, there is no ellipticity condition imposed on the diffusion X.

There is a broad literature on reflected diffusions and their connection to solutions of linear partial differential equations (PDE) with Neumann (gradient) boundary conditions (see e.g. [85], [25] or [18]). One can use these Feynman-Kac type formulas to study the Hamilton-Jacobi-Bellman (HJB) equation associated to the stopping problem (1.1). For discontinuous thresholds b, however, the transfer given by Equation (1.4) may have jumps and thus the value function of the stopping problem may be discontinuous as well. Therefore, the value function is only expected to solve the associated HJB equation in a weak (e.g. viscosity) sense. The purely probabilistic approach presented in Chapter 3 circumvents the technical difficulties arising in a PDE characterization of a discontinuous value function.

Furthermore, a uniqueness result for the transfer implementing a cut-off rule is established. Again the proof employs purely probabilistic techniques. As in Chapter 2 this result leads to a new characterization of optimal stopping boundaries in stopping problems of the form $\sup_\tau E[g(\tau, X_\tau)]$. A threshold b which is càdlàg and has summable jumps leads to an optimal stopping rule τ_b if and only if it satisfies

$$\mathbb{E}\left[\int_t^T (\partial_t + \mathcal{L}) g(s, \tilde{X}_s^{t,b(t)}) ds\right] = 0 \tag{1.5}$$

for all $t \in [0, T]$. For future research it is appealing to develop methods to solve Equation (1.5) numerically. The resulting approximation schemes could then for example be used to compute the optimal exercise boundary for American options or optimal irreversible investment strategies. Chapter 3 also discusses the relation to the nonlinear integral equation derived in [45], [41] and [17] (see also [67]).

Part II: Optimal Closure of Illiquid Positions

It is part of the daily business of many companies to close large asset positions in financial markets. For example institutional investors (e.g. hedge funds, investment banks) sell and buy large amounts of stocks when changing their investment strategy. Risk managers of insurance companies need to trade large quantities of their assets to rebalance their portfolio. Energy companies buy large amounts of coal and natural gas in order to run their power plants. In many cases, the size of the position to be closed corresponds to a large proportion of the daily volume traded in the asset. Due to limited market liquidity, large block-trades cause significant price movements, a phenomenon referred to as *price impact*. When closing a position in an illiquid market, trading large amounts of the asset over a short time span, thus moves the price into an unfavorable direction. *Execution costs* are the difference between the revenues if the whole position could be closed at a given benchmark price (e.g. a quoted market price) and the actual revenues from unwinding the position in a market with price impact. To minimize execution costs traders split up their positions into smaller parts and place them successively in the market. Splitting orders over time, however, entails *price risk*: Due to the enlarged holding period the risk of a decline of the asset price increases. Finding an optimal schedule of trades is referred to as optimal liquidation problem (or likewise optimal trade execution, optimal position closure) in the mathematical finance literature.

The importance for institutional investors to take into account the control of execution costs when making investment decisions was pointed out by Pérold [64]. Enhanced by the introduction of electronic trading platforms, research on how to liquidate a large asset position has developed rapidly. Most models are formulated as a *stochastic control problem* where the characteristics of the price impact are exogenously specified. A first model is introduced in Bertsimas and Lo [12], who describe liquidation strategies that minimize execution costs over a fixed time horizon. Almgren and Chriss [4] obtain deterministic liquidation strategies that are optimal for an investor concerned about the mean and the variance of returns. In these papers the authors distinguish between a temporary and a permanent price impact. The first impact is just instantaneous and only affects the current transaction but does not have any influence on subsequent trades. The latter shifts the price permanently. Almgren and Chriss [4] assume that both impact functions depend linearly on the amount traded at every point in time. While being mathematically very tractable this simple framework still captures the most relevant features of illiquid markets. In particular it turns out that spreading orders over time reduces execution costs. Based on the Almgren and Chriss price impact framework a variety of model extensions have emerged. In a follow-up paper Almgren [2] analyzes a liquidation model where the price impact function is of power law type. Schied, Schöneborn and Tehranchi [76] determine optimal trading strategies for investors with constant absolute risk aversion. The solution method in this and several other papers (see e.g. [33], [29], [81]) is based on Bellman's dynamic programming principle. This optimality principle for stochastic control problems leads in continuous-time to a nonlinear partial differential equation (PDE), called the Hamilton-Jacobi-Bellman (HJB) equation, which is satisfied by the value function. A particularity of this approach applied to optimal liquidation problems originates from the liquidation constraint: At the end of the liquidation period the open position has to be closed. This terminal state constraint leads to a singularity in the HJB equation at the terminal time which makes a PDE characterization of the value function challenging from a mathematical perspective.

There is a further strand in the research on optimal trade execution which takes into account a phenomenon called *price resilience*: In practice one often observes that after a large block-trade, the price recovers and returns to its former level. In recent years several papers have studied the influence of resilience on the optimal closure of asset positions. For example Obizhaeva and Wang [60] set-up a model of the supply / demand dynamics within a limit order book. The limit order book is block shaped and the bid-ask-spread recovers exponentially from large trades. Alfonsi, Fruth and Schied [1] consider a generalization of this model by allowing for a general shape of the limit order book.

By now many model variations that aim at describing further aspects of optimal liquidation problems, have been analyzed in the literature. Here are a few examples of recent mathematical contributions: For instance, [37] set up a model with permanent multiplicative price impact and determine optimal execution strategies by solving a singular continuous-time stochastic control problem. The papers [31] and [9] consider models with *time-dependent* deterministic liquidity. In the first paper optimal execution strategies are derived with dynamic programming techniques, whereas the second uses a convex optimization. Løkka [52] analyzes a liquidation problem of an investor with constant absolute risk aversion in a limit order book. Optimal trading strategies turn out to be singular and are characterized in terms of the free boundary of the associated HJB equation. The impact of so-called "dark" pools on liquidation strategies is analyzed in [48].

Frequently traders set a minimum goal for the revenues they intend to generate from selling a large asset position. For example they want to earn at least a certain proportion of the book value of their position. To reach this target they follow a passive-in-the-money strategy: When prices fall, they enlarge the trading speed in order to increase the probability of achieving their goal. When prices increase, the risk of falling below the minimum target gets smaller and they close the position slowly focusing on minimizing execution costs. Accelerating trading when prices move into an unfavorable direction implies that the revenues from closing a long position are right-skewed. The left-hand tail of the distribution is thinner than when selling independently of price moves. Indeed, there is empirical evidence that skewed distributions of proceeds are preferred to unskewed ones (see e.g. [23] and [16]).

The aim of Chapter 4, which is based on [7] and [8], is to provide *price-sensitive* trading strategies that reduce the risk of falling below a target when closing a large position up to some fixed time horizon T. Based on the Almgren and Chriss framework [4] a liquidation problem is set-up that consists of minimizing a weighted sum of expected execution costs and a risk functional. The risk functional can be interpreted as the time average of the squared value-at-risk of the open position. It turns out that the risk associated with the open position depends on the *price evolution* of the asset to unwind. The price of the asset at the beginning of the liquidation period determines a *reference price* to which the average of the *realized* proceed/cost per share is compared once the position is closed. If during the liquidation period prices move in favor of the agent, then the risk exposure is reduced. On the contrary, if prices move into an unfavorable direction, then the risk increases, and one expects that trading speed increases, too.

Chapter 4 analyzes a continuous-time as well as a discrete-time version of the model. In the discrete-time case trading is only allowed at finitely many time steps. Using discrete stochastic dynamic programming allows to derive optimal trading strategies for closing a portfolio consisting of different assets. The strategies are in general not given in closed form, but determined through a backward function recursion.

In the continuous-time case a characterization of optimal trading strategies is established in terms of a solution to a nonlinear PDE. The novelty of the solution approach is that instead of giving a direct PDE characterization of the value function of the control problem, the PDE under consideration describes by how much optimal trading strategies differ from the optimal risk-neutral trading strategy. Due to the liquidation constraint the value function of the control problem possesses a singularity at time T.

By considering the deviation from the optimal risk-neutral strategy one ends up with a nonlinear PDE with finite terminal condition, that possesses a singularity at time T in its generator. This singularity, however, turns out to be benign, allowing to establish a uniqueness result in the class of viscosity solutions. In general the solution to the PDE cannot be expected to be smooth. This is why one has to fall back to weak solutions to nonlinear PDEs. Chapter 4 employs the concept of viscosity solutions. It has been introduced in the 1980s by Lions and Crandall ([19]) and has subsequently been identified as an efficient tool to solve stochastic control problems. The reader is referred to [40] for an introduction. The uniqueness result allows to characterize the value function and optimal strategies of the stochastic control problem with terminal state constraint. Moreover, a stability result for the model is established. As the number of time steps in the discrete-time model goes to infinity, the discrete-time optimal trading strategies converge to their continuous-time counterparts.

Furthermore, an algorithm is presented for computing optimal trading strategies from the discrete-time model numerically. This approximation scheme is based on a linear interpolation method which allows for a quick iteration through the function recursion. The convergence of the algorithm is verified and errour bounds are derived.

To illustrate the efficiency of the algorithm numerical experiments are performed. In a case study from energy finance, liquidation strategies for the closure of forward positions are computed. Statistical properties of the distribution of revenues are analyzed. It is verified that employing price-sensitive trading strategies leads to a right-skewed distribution of revenues.

In financial markets liquidity usually is not constant but fluctuates as time evolves. There are times when trading is cheap and times when trading is expensive. For example many markets exhibit intraday liquidity patterns: Trading activity is high in the morning and before closure but low around noon (see e.g. [59]). In addition to this *deterministic* market behavior there are also *random* variations of liquidity through the day. Research on the influence of stochastic liquidity on how to unwind positions has emerged only very recently. The dissertation of Fruth [30] extends the model of Obizhaeva and Wang [60] by allowing the recovering rate to be stochastic. Almgren [3] models the temporary price impact exogenously as a stochastic process, and derives optimal deterministic strategies minimizing a weighted sum of the mean and the variance of the proceeds. A more abstract perspective is taken by Schied [75], where the temporary price impact process is assumed to be a Markov process. The control problem consists of minimizing a weighted sum of expected execution costs and a risk functional depending on the price paths. The solution is characterized in terms of super-processes. Graewe, Horst and Séré [34] consider a variant of Schied's liquidation problem, where traders are also allowed to submit passive orders to a so-called "dark" pool to close their position. The impact function is assumed to be a function of the price process. Under an ellipticity condition on the price dynamics the authors establish existence and uniqueness of a classical solution to the associated HJB equation with singular terminal condition.

Chapter 5 generalizes the model from Chapter 4 by allowing for a stochastic price impact. In contrast to Schied's setup (cf. [75]) it is based on a Brownian probability space but allows for general non-Markovian model parameters. Moreover, this chapter is a slight extension of [5] as it considers a general liquidation constraint. A set of scenarios is introuduced where the position does not have to be closed. For example a trader might opt against closing the position if market liquidity is low throughout the liquidation period. In all other events the liquidation constraint is still binding. In contrast to Chapter 4, where analytical tools are employed to solve the control problem, Chapter 5 takes a purely probabilistic approach. A stochastic maximum principle of Pontryagin's type is derived and optimal strategies are characterized by means of backward stochastic differential equations (BSDEs). BSDEs have turned out to be a powerful tool for analyzing stochastic control problems, and for providing purely probabilistic solutions. The reader is referred to the survey article [26] and the book by Pham [68] for examples of control problems solved with BSDEs. Due to the liquidation constraint the BSDE considered in Chapter 5 has a singular terminal condition: On the set of scenarios where the position has to be closed, the solution of the BSDE converges to infinity as time approaches the liquidation horizon. BSDEs with singular terminal condition have so far only been studied in Popier [69] and [70]. One of the chapter's goals is to reveal their power for solving stochastic control problems with terminal state constraint.

To this end it is shown that a minimal supersolution of the BSDE exists, thus partly generalizing the existence results obtained in [69] and [70]. Subsequently, it is verified that optimal strategies are determined by this supersolution. The verification is based on a penalization technique. Finally, special cases for which the control problem has explicit solutions are discussed. In addition, the set of model specifications where deterministic strategies are optimal is characterized.

In order to hedge their risk exposure and to protect themselves against price fluctuations many companies trade on *forward* markets. They agree to buy or sell an asset at a particular time in the future for a certain price. For example energy companies approximately know in advance the amount of coal or natural gas they need in order to run their power plants over a given future time span. To protect themselves against rising prices they aim at closing these positions on forward markets long before the delivery starts. Likewise airline companies enter forward contracts on kerosine to hedge against price risk. Often there is no liquid market for these particular contracts and the manager has to find a counterparty that is willing to enter an over-the-counter (OTC) contract. For entering the contract the counterparty will ask for a premium that usually strongly depends on liquidity: The fewer other traders and hence potential risk takers, the higher the liquidity costs. As the delivery period approaches, trading of forward contracts becomes typically more active and thus liquidity increases. On energy exchanges for instance only front periods (e.g. the next three months) are traded actively. Therefore an early closure of the position implies foregoing the option to reduce execution costs.

Frequently there is the opportunity to *cross hedge* the price risk by acquiring a different forward contract that is liquidly traded and follows the price movements of the original contract closely. However, this forward contract does not perfectly replicate the open position and thus entails *basis risk*. Chapter 6 analyzes the impact of cross-hedging opportunities on liquidation strategies. For example consider an energy company running gas power plants in Germany - an illiquid area of the European gas market (cf. [27]). The company might consider to buy the natural gas in the more liquid Dutch market. Once the German market becomes more liquid the operator sells the Dutch gas again and buys the German gas she requires. Following this hedging strategy the operator may reduce execution costs. Since the Dutch and the German market area are physically connected by pipelines the prices in both market areas are highly correlated. Nevertheless the price spread might move into an unfavorable direction and thus this strategy entails basis risk. The aim is to find optimal trade-offs between minimizing execution costs and basis risk.

Chapter 6, which is based on [6], deals with an optimal liquidation problem where an investor has to close a short position in a primary market over a given time period. In addition she has the opportunity to cross hedge the price risk by acquiring a more liquid, positively correlated asset in a proxy market. In contrast to the preceding chapters, the liquidity costs are modeled as proportional transaction costs and there is no volume-dependent price impact. While the transaction costs in the primary asset are exogenously given by a nonincreasing *stochastic* process, they are assumed to be constant in the proxy market. The investor's risk preferences are modeled by means of an increasing function of the variance of the portfolio. In particular a well diversified portfolio reduces the risk exposure of the investor. The aim of the investor is to minimize a weighted sum of execution costs and the risk functional. The focus on proportional transaction costs instead of including a price impact affects the choice of the mathematical tools required to study the two-dimensional control problem. It turns out that optimal strategies are not necessarily absolutely continuous with respect to Lebesgue measure but may possess jumps. Therefore one has to fall back to methods from *singular control theory* instead of relying on continuous control techniques as in Chapter 4 and Chapter 5. A solution method is provided which is based on the well-known connection between singular control and optimal stopping, see e.g. [43],[14],[36] and the references therein. These results show how to reduce a *one-dimensional* singular control problem to a family of optimal stopping problems. Since the aim of Chapter 6 is to find both the optimal strategy in the primary and in the proxy market *simultaneously*, one cannot apply these results directly. A sufficient condition is provided that allows to reduce the two-dimensional problem to one-dimensional problems. In three stylized case studies it is verified that the condition is satisfied and optimal positions in the primary and in the proxy market are determined in closed form. For example, the case where active trading in the primary market kicks in at a random time is analyzed. Furthermore a deterministic concave decay of liquidity costs considered. In both cases one obtains a simple decision rule of the following form: If the liquidity costs in the proxy market exceed $\bar{L}$, then no cross hedge should be performed. Whether the primary position is closed or not, depends on the ratio between the liquidity cost savings and the risk when keeping the position open. If the liquidity costs of the cross hedging instrument are smaller than a given threshold $\bar{L}$, then the position in the primary should be cross hedged.

Acknowledgements

First of all I would like to thank Stefan Ankirchner for supervising this thesis. This work and especially myself benefited a lot from numerous inspiring discussions and his constant support and encouragement. I had a wonderful time working with him. Moreover, I am grateful to Monique Jeanblanc for helpful discussions, valuable advice and for being an examiner for this thesis. I owe many thanks to Philipp Strack. Conducting research with him was a great pleasure. A lot of the results presented in this thesis are a result from intense and joyful discussions with him. A really big thanks is due to Nikolaus Schweizer for numerous helpful advice and for being a great room mate. Many thanks go to Peter Kratz. I highly appreciated working with him. There are many more people I have to thank for various forms of support, help and hospitality. A non-exhaustive list includes Azzouz Dermoune, Sebastian Ebert, Christine Grün, Xin Guo, Christian Pigorsch, Sven Rady, Judith Schneider, Christian Seel and all members of the probability and stochastic analysis group in Bonn. I owe a big thanks to my family and friends. My deepest gratitude goes to Annika and my parents. Finally, I would like to thank the Bonn International Graduate School in Mathematics for its financial and non-financial support.

Part I.

Inverse Optimal Stopping

2. Inverse optimal stopping in discrete-time

Most optimal stopping problems are concerned with a timing decision where all relevant information is directly observable. There are, however, many economic situations modeled as optimal stopping problems where private information is a crucial feature. Important examples include the regulation of a firm that decides when to enter a market, or the design of unemployment benefits where a worker privately observes her job offers. Consider for example a firm that observes the stochastically changing demand for a new product and needs to decide when to enter the market for this product. If the firm starts selling the product, it pays an investment cost and receives the flow returns from selling the product at all future times. When timing its market entry decision, the firm solves an optimal stopping problem. While the resulting timing decision is optimal for the firm, it is typically wasteful from a social perspective as the firm does not take into account the consumer surplus of the new product. A question that naturally arises is how a regulator can influence the incentives of the firm if the demand for the new product is private information of the firm. Given an optimal stopping decision from the regulator's perspective, the challenge is to find a reward mechanism that makes this stopping rule also optimal for the firm.

The aim of this chapter is to analyze this kind of *inverse* optimal stopping problem. Time is discrete and the underlying stochastic process is one-dimensional and satisfies the Markov property. Section 2.1 formulates inverse optimal stopping problems in this framework and introduces the notion of implementable stopping rules. In Section 2.2 it is shown that all cut-off rules are implementable under a single crossing condition. The associated transfer admits a closed form representation involving constrained stochastic processes. The transfer is unique up to an additive constant. This leads to a new characterization of optimal stopping boundaries in Markovian stopping problems. Section 2.3 presents an application of our approach to job search with and without recall. For risk-averse agents the set of time-dependent reservation wage policies that can be implemented through unemployment benefits is characterized.

2.1. The model

2.1.1. Evolution of the private information

Time is discrete and indexed by $t \in \{0, 1, \dots, T\} = \mathbb{T}$, for some fixed time horizon

$T < \infty$. At every point in time $t \in \mathbb{T}$ an agent privately observes a real-valued Markov process $(X_t)_{t\in\mathbb{T}}$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathbb{T}})$. The initial value of the process X_0 is distributed according to the distribution function $F : \mathbb{R} \to [0, 1]$,

$$\mathbb{P}\left[X_0 \leq z\right] = F(z)\,.$$

Our formulation allows for completely general Markov processes which satisfy the following weak regularity assumptions.

Standing Assumption 2.1.1 (Polynomial Growth)**.** X is of polynomial growth, i.e. there exists a number $p > 0$ and a constant $C > 0$ such that

$$\mathbb{E}\left[|X_{t+1}|^p \,|\, X_t = x\right] \leq C(1 + |x|^p) \text{ for all } x \in \mathbb{R}\,.$$

We say that a function $\phi : \mathbb{R} \to \mathbb{R}$ is of polynomial growth if there exists a constant $\tilde{C} > 0$ such that $|\phi(x)| \leq \tilde{C}(1 + |x|^p)$ for all $x \in \mathbb{R}$ with the same p as in Assumption 2.1.1. Assumption 2.1.1 assures that expected values of polynomial growth functions are finite with respect to the conditional probability measure of the process X.

Standing Assumption 2.1.2 (Monotone Transitions)**.** A higher value of the process $x' \geq x$ at time t leads to a higher value of the process at time $t + 1$ in the sense of first order stochastic dominance

$$\mathbb{P}\left[X_{t+1} \leq z \,|\, X_t = x'\right] \leq \mathbb{P}\left[X_{t+1} \leq z \,|\, X_t = x\right] \text{ for all } z \in \mathbb{R}\,. \tag{2.1}$$

Standing Assumption 2.1.3 (Continuous Transitions)**.** For every $t \in \mathbb{T}$ and for every continuous, polynomial growth function $\phi : \mathbb{R} \to \mathbb{R}$ the function $x \mapsto \mathbb{E}\left[\phi(X_{t+1}) \,|\, X_t = x\right]$ is continuous.

Many Markov processes commonly used in the economic literature satisfy these assumptions. The following result proven in the Appendix shows that for processes with independent, multiplicative or additive random shocks the assumptions are satisfied.

Example 2.1.4 (Additive random walk)**.** *Let $X_t = X_0 + \sum_{s\leq t} \epsilon_s$ be the sum of identically distributed, independent shocks $(\epsilon_t)_{t\in\mathbb{T}}$ with finite second moment $\mathbb{E}\left[\epsilon_t^2\right] = \sigma^2 < \infty$.*

Example 2.1.5 (Multiplicative random walk)**.** *Let $X_t = X_0 \prod_{s\leq t} \epsilon_s$ be the product of identically distributed, independent shocks $(\epsilon_t)_{t\in\mathbb{T}}$ with finite second moments $\mathbb{E}\left[\epsilon_t^2\right] = \sigma^2 < \infty$.*

Example 2.1.6 (Search without recall)**.** *Let $(X_t)_{t\in\mathbb{T}}$ be a collection of integrable, independent and identically distributed random variables.*

Example 2.1.7 (Search with recall)**.** *Let $(Y_t)_{t\in\mathbb{T}}$ be a collection of integrable, independent and identically distributed random variables. Set $X_t = \sup_{s\leq t} Y_t$.*

Proposition 2.1.8 (Regularity of random walks and search processes)**.** *Random walks and search processes have continuous and monotone transitions. Random walks are of polynomial growth of order $p = 2$ and search processes of order $p = 1$.*

Assumptions 2.1.1,2.1.2 and 2.1.3 are standing assumptions which hold throughout the chapter. If explicitly stated we impose additional assumptions. For example for the uniqueness result Proposition 2.2.9 we assume that the probability measure governing the transitions of X has full support.

Assumption 2.1.9 (Full Support)**.** For every $x \in \mathbb{R}$, $a < b$ and $t < T$ we have

$$\mathbb{P}\left[X_{t+1} \in [a,b) \,|\, X_t = x\right] > 0\,.$$

Moreover the distribution function F of the initial value X_0 is absolutely continuous.

To simplify notation we define the transition kernel $P_{t,s}$ of X for $t < s \in \mathbb{T}$, which acts on polynomial growth, measurable functions by

$$P_{t,s}\phi(x) = \mathbb{E}\left[\phi(X_s) \,|\, X_t = x\right].$$

If the function ϕ depends also on time we write $P_{t,s}\phi(x) = \mathbb{E}\left[\phi(s, X_s) \,|\, X_t = x\right]$. By slight abuse of notation we sometimes also write

$$P_{t,s}(x, A) = P_{t,s}\mathbf{1}_A(x) = \mathbb{P}\left[X_s \in A | X_t = x\right]$$

for a Borel measurable set $A \subseteq \mathbb{R}$. Assumption 2.1.2 implies that the kernel $P_{t,t+1}$ preserves monotonicity: If $\phi : \mathbb{R} \to \mathbb{R}$ is nonincreasing then $P_{t,t+1}\phi$ is nonincreasing as well. Assumptions 2.1.1 and 2.1.3 ensure that for every continuous, polynomial growth function ϕ the function $P_{t,t+1}\phi$ is continuous and is of polynomial growth as well.

2.1.2. Strategies and payoffs of the agent

Based on her past observations of the process the agent decides when to stop. Denote by $\mathcal{T}$ the set of $(\mathcal{F}_t)$-adapted stopping rules[1].

The payoff consists of three parts. At any time t before stopping the agent receives a flow payoff $f(t, X_t)$. At the time she stops she receives a final payoff $g(\tau, X_\tau)$. At any time t after stopping she obtains a flow payoff $h(t, X_t)$. The agent's expected payoff $V(\tau)$ when using the stopping rule τ equals

$$V(\tau) = \mathbb{E}\left[\left(\sum_{t=0}^{\tau-1} f(t, X_t)\right) + g(\tau, X_\tau) + \left(\sum_{t=\tau+1}^{T} h(t, X_t)\right)\right]. \tag{2.2}$$

The continuous, polynomial growth payoff functions $f, g, h : \mathbb{T} \times \mathbb{R} \to \mathbb{R}$ depend on time t and the value of the process x.

Definition 2.1.10 (Marginal Incentive)**.** We define the marginal incentive of the agent to delay the stopping decision at $t \in \mathbb{T}$ when $X_t = x$ by

$$z(t, x) = f(t, x) + \mathbb{E}\left[g(t+1, X_{t+1}) - g(t, x) - h(t+1, X_{t+1}) \,|\, X_t = x\right]. \tag{2.3}$$

[1] The terms stopping rule and stopping time are used interchangeably in this thesis.

The marginal incentive z of the agent equals the expected change in payoffs when instead of stopping in period t she stops in period $t+1$.

In static mechanism design with real valued private information one often imposes a single crossing condition. The static single crossing condition ensures that the change in utility from getting a higher allocation is decreasing in the private information. In our dynamic setup, allocations differ in the time dimension. We impose the following analogue of the static single crossing condition:

Standing Assumption 2.1.11 (Single Crossing)**.** The marginal incentive $z(t,x)$ is strictly decreasing in x for every $t \in \mathbb{T}$.

To ensure that Assumption 2.1.11 holds one can impose the following monotonicity conditions on f, g and h. If h is increasing and f and $\mathbb{E}\left[g(t, X_{t+1}) - g(t, X_t) \,|\, X_t = x\right]$ are decreasing in x and if there is at least one strict monotonicity, then z is strictly decreasing. The single crossing condition has a natural interpretation in many economic models. In a job search model, it means that it is more costly to reject a good offer than to reject a bad offer. In an irreversible investment problem where the process X represents the demand it means that the loss in profits from entering the market is higher if the demand is higher. A further example is provided by a two-state one-armed bandit model where X is the posterior belief. The single crossing condition is satisfied since the expected payoff is lower the higher the probability that the state of the world is bad. For some results we assume in addition that z is unbounded which will ensure that it is optimal to stop for sufficiently high values of the process and optimal to continue for sufficiently low values of the the process (cf. Proposition 2.2.6).

Assumption 2.1.12 (Unbounded Marginal Incentive)**.** We have $\lim_{x\to\infty} z(t,x) = -\infty$ and $\lim_{x\to-\infty} z(t,x) = \infty$ for every $t \in \mathbb{T}$.

2.1.3. Implementable stopping rules

We want to characterize how the behavior of the agent can be influenced using a transfer $\pi : \mathbb{T} \to \mathbb{R}$ that is only a function of the realized stopping decision, but not of the path of the process X the agent observes. The transfer π is paid to the agent in addition to her payoffs f, g, h and is thereby changing her preferences over stopping times.

Definition 2.1.13 (Implementable Stopping Rule)**.** A stopping rule $\tau^\star$ is implemented by a transfer π if $\tau^\star$ is the minimal[2] stopping rule satisfying

$$\sup_{\tau \in \mathcal{T}} V(\tau) + \mathbb{E}\left[\pi(\tau)\right] = V(\tau^\star) + \mathbb{E}\left[\pi(\tau^\star)\right]. \tag{2.4}$$

We say that $\tau^\star$ is implementable if there exists a transfer π that implements it.

[2] A stopping rule τ is minimal in a set of stopping rules $S \subseteq \mathcal{T}$ if $\tau \leq \tau'$ almost surely for all $\tau' \in S$. Note that minimal stopping times are almost surely unique. If $\tau, \tau' \in S \subseteq \mathcal{T}$ are minimal stopping times in S, then $\tau \leq \tau' \leq \tau$ and hence $\tau = \tau'$ almost surely.

Note that this is a strong notion of implementability since we also demand minimality of the given stopping rule in addition to optimality. This additional requirement allows us to derive a uniqueness result in Subsection 2.2.4.

A transfer rule $\pi : \mathbb{T} \to \mathbb{R}$ has multiple attractive economic features: As the transfer π is independent of the realization of the process X it can be paid even if the realization of X is unobservable. The only required information is the realized stopping decision. Intuitively it suffices to know that the agent stopped, instead of for what reasons she stopped. Furthermore, as the transfer depends only on the stopping decision it requires no communication.

2.2. Characterization of implementable stopping rules and transfers

The expected payoff V defines a preference relation over stopping times. The agent prefers the stopping time τ over τ' if and only if $V(\tau) \geq V(\tau')$. The next result, proven in the Appendix, shows that the preferences of the agent over stopping times depend only on her marginal incentive to delay the allocation.

Proposition 2.2.1. *The expected payoff of the agent when using the stopping rule τ can be represented as the sum of the payoff of stopping in period zero plus her expected marginal incentives*

$$V(\tau) = \mathbb{E}\left[\sum_{t=0}^{\tau-1} z(t, X_t)\right] + V(0).$$

Proposition 2.2.1 shows that the agent's preferences over stopping times are completely determined by the marginal incentive z to delay the allocation to next period. Note that Proposition 2.2.1 is a result about *expected* payoffs. The realized payoffs for a given path of X may differ. As a consequence of Proposition 2.2.1 the definition of implementability can be simplified. A stopping time $\tau^\star$ is implemented by the transfer π if τ is the minimal stopping time satisfying[3]

$$\sup_{\tau \in \mathcal{T}} \mathbb{E}\left[\sum_{t=0}^{\tau-1} z(t, X_t) + \pi(\tau)\right] = \mathbb{E}\left[\sum_{t=0}^{\tau^\star-1} z(t, X_t) + \pi(\tau^\star)\right]. \tag{2.5}$$

It is convenient to generalize the agent's payoff V and to allow for arbitrary initial times t and initial values x. For $t \in \mathbb{T}$ let $\mathcal{T}_{t,T}$ denote the set of all stopping rules with values in $\{t, \ldots, T\}$. For any stopping rule $\tau \in \mathcal{T}_{t,T}$ we define the agent's expected continuation value by

$$V_{t,x}(\tau) = \mathbb{E}\left[\left(\sum_{s=t}^{\tau-1} f(s, X_s)\right) + g(\tau, X_\tau) + \left(\sum_{s=\tau+1}^{T} h(s, X_s)\right) \,\Big|\, X_t = x\right]. \tag{2.6}$$

[3] We use the convention $\sum_{s=t}^{t-1} \cdot = 0$.

For every transfer π we introduce the agent's value function $v_\pi : \mathbb{T} \times \mathbb{R} \to \mathbb{R}$ as

$$v_\pi(t,x) = \sup_{\tau \in \mathcal{T}_{t,T}} \left(V_{t,x}(\tau) + \mathbb{E}\left[\pi(\tau) \,|\, X_t = x\right]\right) - V_{t,x}(t).$$

Hence, $v_\pi(t,x) - \pi(t)$ is the difference between the expected payoff when continuing optimally and stopping immediately. Put differently, the difference $v_\pi(t,x) - \pi(t)$ is the agent's willingness to pay for the option to stop after period t. It follows from the argument in Proposition 2.2.1 that v_π is the supremum of the sum of expected future marginal incentives and the transfer, when the agent follows the optimal continuation strategy

$$v_\pi(t,x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\left(\sum_{s=t}^{\tau-1} z(s, X_s)\right) + \pi(\tau) \,\Big|\, X_t = x\right]. \tag{2.7}$$

Since the agent always has the opportunity to stop immediately, i.e. to choose $\tau = t$, it follows $v_\pi(t,x) - \pi(t) \geq 0$. At each time t the agent faces the binary decision of continuing or stopping. Given a value x of the process at time t she bases her decision on whether the value of the option to continue is positive. If $v_\pi(t,x) - \pi(t) = 0$, then there is no gain from continuing and the agent stops. If $v_\pi(t,x) - \pi(t) > 0$, the agent continues at least one more period. Intuitively, it follows that the minimal optimal stopping rule for the agent is given by

$$\tau^\star = \inf\left\{t \geq 0 \,|\, X_t \in D_\pi(t)\right\} \wedge T, \tag{2.8}$$

where the so-called stopping region $D_\pi(t)$ is defined by

$$D_\pi(t) = \{x \in \mathbb{R} \,|\, v_\pi(t,x) = \pi(t)\}.$$

If the agent decides to stop at time t, she receives the transfer $\pi(t)$. If it is optimal to continue, she obtains the marginal incentive $z(t,x)$ plus the expected value of continuing optimally in the next period $\mathbb{E}\left[v_\pi(t+1, X_{t+1}) \,|\, X_t = x\right]$. This leads to the dynamic programming principle which represents the value function in recursive form for all $t \in \mathbb{T}$ and $x \in \mathbb{R}$

$$v_\pi(t,x) = \max\left\{\pi(t), z(t,x) + \mathbb{E}\left[v_\pi(t+1, X_{t+1}) \,|\, X_t = x\right]\right\}. \tag{2.9}$$

For a rigorous derivation of Equations (2.8) and (2.9) we refer to [67, Chapter 1, Theorem 1.9]. The following Lemma establishes the regularity of the value function for every transfer.

Lemma 2.2.2. *The value function v_π is nonincreasing, continuous and of polynomial growth in x.*

We introduce the notion of cut-off rules. Cut-off rules are stopping rules such that the agent stops the first time the process X exceeds a time-dependent threshold b. A mapping $b : \mathbb{T} \to \overline{\mathbb{R}}$ [4]with $b(T) = -\infty$ is called a cut-off.

[4]Here and in the sequel $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$

Definition 2.2.3 (Cut-Off Rule)**.** A stopping rule τ is a cut-off rule if there exists a cut-off b such that almost surely

$$\tau = \inf \{t \in \mathbb{T} \,|\, X_t \geq b(t)\} .$$

We denote the cut-off rule corresponding to the cut-off b by $\tau_b = \inf \{t \in \mathbb{T} \,|\, X_t \geq b(t)\}$. If $-\infty < b(t) < \infty$ for all $t < T$ we call b a finite cut-off and τ_b a finite cut-off rule. The next Lemma, proven in the Appendix, shows that under the full support assumption we can uniquely recover the cut-off b of a cut-off rule τ.

Lemma 2.2.4. *Suppose that the process X has full support, i.e. Assumption 2.1.9 holds. Then for every cut-off rule τ there exists a unique cut-off b satisfying $\tau = \tau_b$ almost surely.*

In Section 2.2.1 we show that every implementable stopping rule, is a cut-off rule. In Section 2.2.3 we show that all cut-off rules are implementable. The associated transfer admits an explicit representation in terms of the constrained version of X which we introduce in Section 2.2.2.

2.2.1. All implementable stopping rules are cut-off rules

It is known that if one imposes the single crossing condition in a static mechanism design problem an allocation rule can be implemented if and only if it is monotone (see e.g. [35]). In our dynamic setup cut-off rules play the role of monotone allocations. Indeed, Assumption 2.1.2 implies that a cut-off rule τ_b is monotone in the sense of first order stochastic dominance: For any $t < s$ the probability that X exceeds b before time s is increasing in the conditional value of X at time t

$$\mathbb{P}\left[\tau_b \leq s \,|\, X_t = x'\right] \leq \mathbb{P}\left[\tau_b \leq s \,|\, X_t = x\right] \text{ for all } x' < x < b(t) .$$

The next Proposition shows that only cut-off rules are implementable.

Proposition 2.2.5. *If the stopping rule τ is implementable, then τ is a cut-off rule.*

Proof. Denote by π the transfer implementing τ. Let τ_D be the first hitting time of X of the stopping region

$$\tau_D = \inf\{t \geq 0 \,|\, X_t \in D_\pi(t)\} = \inf\{t \geq 0 \,|\, v_\pi(t, X_t) = \pi(t)\} .$$

Then [67, Theorem 1.9] yields that τ_D is a minimal optimal stopping rule for the agent's stopping problem given the transfer π. Fix a point in time t and a value of the process $x \in D_\pi(t)$ such that it is optimal to stop. By Lemma 2.2.2 the value function v_π is nonincreasing and hence for every point $x' \geq x$

$$\pi(t) = v_\pi(t, x) \geq v_\pi(t, x') .$$

By definition the value function v_π is bounded from below by $\pi(t)$ and hence we have $v_\pi(t, x') = \pi(t)$. Thus every value $x' \geq x$ is in the region $x' \in D_\pi(t)$ where it is optimal for the agent to stop. This implies that the stopping region $D_\pi(t)$ is an interval which is unbounded on the right. Again by Lemma 2.2.2 the function $x \mapsto v_\pi(t,x)$ is continuous and hence $D_\pi(t)$ is closed. Therefore there exists some $b(t) \in \overline{\mathbb{R}}$ such that $D_\pi(t) = [b(t), \infty)$. This implies that τ_D is a cut-off rule with cut-off b. For every minimal optimal stopping rule we have $\tau = \tau_D$ almost surely and hence τ is a cut-off rule. □

Under Assumption 2.1.12 the marginal incentive $z(t,x)$ to delay the stopping decision from time t to $t+1$ gets arbitrarily large as x decreases. We show that this assumption suffices to guarantee that for any transfer π there exists some level $\underline{x} \in \mathbb{R}$ where it is strictly optimal for the agent to continue. Moreover there exists a finite threshold $\overline{x}$ where it is optimal to stop. In between these two levels there exists a number $b(t) \in [\underline{x}, \overline{x}]$ where the agent is indifferent between stopping and continuing.

Proposition 2.2.6. *If Assumption 2.1.12 holds and the stopping rule τ is implemented by a transfer π, then τ is a cut-off rule with a <u>finite</u> cut-off b. The agent is indifferent between stopping and continuing at the cut-off*

$$v_\pi(t, b(t)) = \pi(t) = z(t, b(t)) + \mathbb{E}\left[v(t+1, X_{t+1}) | X_t = b(t)\right]. \tag{2.10}$$

Proof. Let b be the cut-off from Proposition 2.2.5 such that $D_\pi(t) = [b(t), \infty)$ and $\tau = \tau_b$ almost surely. Under Assumption 2.1.12, since v_π is nonincreasing in x, there exist $\underline{x}, \overline{x} \in \mathbb{R}$ such that

$$z(t, \overline{x}) + \mathbb{E}\left[v_\pi(t+1, X_{t+1}) \,|\, X_t = \overline{x}\right] \leq \pi(t) < z(t, \underline{x}) + \mathbb{E}\left[v_\pi(t+1, X_{t+1}) \,|\, X_t = \underline{x}\right].$$

Hence, at $\overline{x}$ it is optimal for the agent to stop and we have $\overline{x} \in D_\pi(t)$. In particular it follows that $b(t) \leq \overline{x} < \infty$. The agent strictly prefers continuing to stopping at $\underline{x}$ which means that $\underline{x} \notin D_\pi(t)$ and consequently $b(t) > \underline{x} > -\infty$. We deduce that τ is a finite cut-off rule.

Since $b(t)$ is the minimal element of $D_\pi(t)$ we have that $v_\pi(t, b(t)) = \pi(t)$. On the other hand we have $v_\pi(t,x) = z(t,x) + P_{t,t+1}v_\pi(x)$ for all $x < b(t)$. Now taking the limit $x \nearrow b(t)$ yields $v_\pi(t, b(t)) = z(t, b(t)) + P_{t,t+1}v_\pi(b(t))$, where we used Lemma 2.2.2 and Assumption 2.1.3. The definition of v_π establishes Equation (2.10). □

2.2.2. Constrained processes

In this section we introduce constrained processes. This class of processes plays a crucial role in designing mechanisms for optimal stopping problems. We will show that the transfer implementing a given cut-off rule can always be represented as an expectation over the constrained version of the original process X. For a given cut-off b the constrained version of X is a Markov process $\tilde{X}$ which evolves as X below b, but is constrained to be on the cut-off b whenever X tries to exceed it. For an illustration we first present the construction of a constrained random walk (see also Figure 2.1).

Example 2.2.7 (Constrained random walk)**.** *As in Example 2.1.4 let X be a random walk, i.e. $X_t = X_0 + \sum_{s \leq t} \epsilon_s$ with $\mathbb{P}[\epsilon_s = 1] = \mathbb{P}[\epsilon_s = -1] = \frac{1}{2}$. In this case the* *constrained version can be constructed path-by-path. First, set* $\tilde{X}_0 = X_0 \wedge b(0)$ *and then define* $\tilde{X}$ *recursively by* $\tilde{X}_{t+1} = (\tilde{X}_t + \epsilon_{t+1}) \wedge b(t+1)$ *for all* $t < T$.

For general Markov processes we take a different approach to construct the constrained version. We first modify the transition probabilities of the original process X and then we define $\tilde{X}$ as the Markov process having these new transition probabilities. For every $t < T$ we define a kernel $\tilde{P}_{t,t+1}$ which acts on bounded, measurable functions by

$$\tilde{P}_{t,t+1}\phi(x) = \mathbb{E}\left[\phi(X_{t+1} \wedge b(t+1)) \,|\, X_t = x\right].$$

We also write

$$\tilde{P}_{t,t+1}(x, A) = \tilde{P}_{t,t+1}1_A(x) = \mathbb{P}\left[X_{t+1} \wedge b(t+1) \in A \,|\, X_t = x\right]$$

for $x \in \mathbb{R}$ and a Borel measurable set $A \subset \mathbb{R}$. We extend the family of kernels to a semi-group $(\tilde{P}_{t,s})_{t \le s}$ via the usual composition

$$\tilde{P}_{t,s}\phi = \tilde{P}_{t,t+1}(\tilde{P}_{t+1,t+2} \ldots (\tilde{P}_{s-1,s}\phi)).$$

The operator $\tilde{P}_{t,t}$ is defined to be the identity. Then there exists a Markov process $\tilde{X}$ on some filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}, (\tilde{\mathcal{F}}_t)_{t \in \mathbb{T}})$ satisfying

$$\tilde{\mathbb{P}}\Big[\tilde{X}_{t+1} \in A \,|\, \tilde{X}_t = x\Big] = \tilde{P}_{t,t+1}(x, A) = \mathbb{P}\Big[X_{t+1} \wedge b(t+1) \in A \,|\, X_t = x\Big] \tag{2.11}$$

for all $x \in \mathbb{R}$, $t < T$ and all Borel measurable sets $A \subset \mathbb{R}$. We call $\tilde{X}$ the constrained version of X. Choosing $A \subset (-\infty, b(t+1))$ implies $\tilde{\mathbb{P}}\big[\tilde{X}_{t+1} \in A | \tilde{X}_t = x\big] = \mathbb{P}\big[X_{t+1} \in A | X_t = x\big]$, i.e. the transition probabilities of $\tilde{X}$ and X coincide below the cut-off. For $A = (b(t+1), \infty)$ we obtain $\tilde{\mathbb{P}}\big[\tilde{X}_{t+1} > b(t+1) | \tilde{X}_t = x\big] = 0$, which means that $\tilde{X}$ never exceeds the cut-off.

2.2.3. All (finite) cut-off rules are implementable

As shown in Section 2.2.1 only cut-off rules are implementable. In this section we prove the opposite direction, i.e. every finite cut-off rule is implementable. For a given cut-off rule we define the transfer and explicitly verify that it implements the cut-off rule. In static mechanism design every monotone allocation rule is implemented by a transfer equal to the integral over marginal incentives (cf. [35]). In our dynamic model cut-off rules are the equivalent of monotone allocation rules. The transfer π implementing a finite cut-off rule τ_b equals the expected future marginal incentive z to delay the allocation evaluated at the process $\tilde{X}$ constrained by the cut-off b

$$\pi(t) = \tilde{\mathbb{E}}\left[\sum_{s=t}^{T-1} z(s, \tilde{X}_s) \,|\, \tilde{X}_t = b(t)\right] \text{ for } t < T \text{ and } \pi(T) = 0\,. \tag{2.12}$$

Theorem 2.2.8. *Every finite cut-off rule* τ_b *is implemented by the transfer defined in Equation (2.12).*

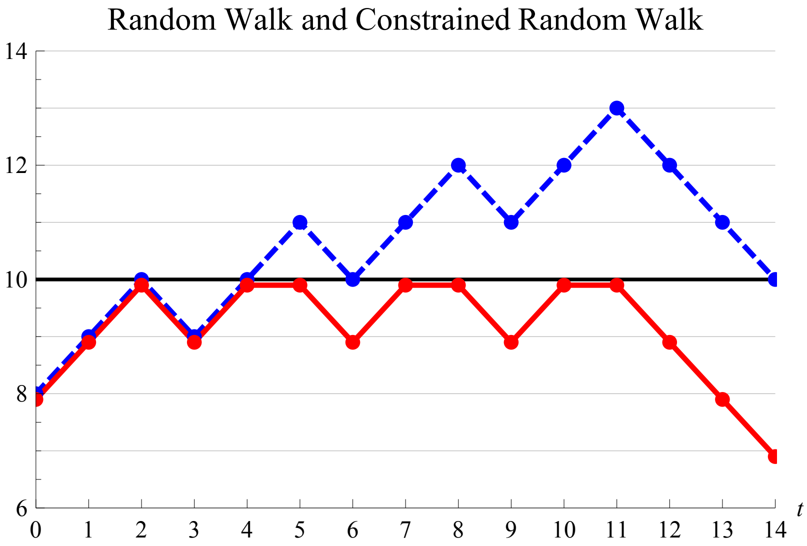

Figure 2.1.: A binomial random walk (dashed line) and its constrained version (solid line) starting in $X_0 = 8$ constrained by the constant cut-off $b(t) = 10$.

Proof. First, we claim that for the transfer defined in Equation (2.12) the value function defined in Equation (2.7) satisfies

$$v_\pi(t,x) = \sum_{s=t}^{T-1} \tilde{P}_{t,s} z(x \wedge b(t)). \tag{2.13}$$

To prove Equation (2.13) we proceed via backward induction. At time T we have by definition $\pi(T) = 0$ and $v_\pi(T,x) = 0$ for all $x \in \mathbb{R}$. At time $t < T$ first observe that we have by induction hypothesis

$$\mathbb{E}\left[v_\pi(t+1, X_{t+1}) \,|\, X_t = x\right] = \sum_{s=t+1}^{T-1} \mathbb{E}\left[\tilde{P}_{t+1,s} z(X_{t+1} \wedge b(t+1)) \,|\, X_t = x\right] = \sum_{s=t+1}^{T-1} \tilde{P}_{t,s} z(x)\,.$$

The dynamic programming principle implies

$$v_\pi(t,x) = \max\left\{\pi(t), z(t,x) + \mathbb{E}\left[v_\pi(t+1, X_{t+1}) \,|\, X_t = x\right]\right\} = \max\left\{\pi(t), \sum_{s=t}^{T-1} \tilde{P}_{t,s} z(x)\right\}.$$

By the single crossing condition and Assumption (2.1) we obtain that the mapping $x \mapsto \tilde{P}_{t,s} z(x)$ is nonincreasing for every $s > t$. As we have by definition $\pi(t) = \sum_{s=t}^{T-1} \tilde{P}_{t,s} z(b(t))$ this yields Equation (2.13).

The fact that $x \mapsto z(t,x)$ is strictly decreasing implies that $x \mapsto \sum_{s=t}^{T-1} \tilde{P}_{t,s} z(x)$ is strictly decreasing as well. From Equation (2.13) we conclude that the stopping region is equal to the interval $D_\pi(t) = [b(t), \infty)$. Then [67, Theorem 1.9] yields that τ_b is a minimal stopping rule implemented by π, i.e. solving Equation (2.5). □

2.2.4. Uniqueness of the transfer

An important result in static mechanism design and auction theory is the revenue equivalence theorem. It was first observed in [82] that many classical auction mechanisms (first-price auctions, Dutch auctions, English auctions, and second-price auctions) lead to the same expected transfers. This result was later generalized to many other auction setups and mechanism design problems. The revenue equivalence theorem states that for every fixed allocation rule the expected transfers implementing it are unique up to a constant. The following proposition shows that revenue equivalence holds in our dynamic model, if one imposes a full support assumption similar to the condition necessary in static mechanism design problems.

Proposition 2.2.9 (Revenue Equivalence)**.** *Suppose that X has full support (Assumption 2.1.9) and let τ be a finite cut-off rule, then the transfer implementing τ is unique up to an additive constant.*

Proof. Let $\pi, \hat{\pi}$ be two payments implementing τ such that $\pi(T) = \hat{\pi}(T)$ and let $v = v_\pi$ and $\hat{v} = v_{\hat{\pi}}$ denote the associated value functions. We show that the two value functions coincide: $v(t, x) = \hat{v}(t, x)$ for all $t \in \mathbb{T}$ and $x \in \mathbb{R}$. This implies uniqueness of the transfer since by Lemma 2.2.4, there exists a unique cut-off b satisfying $\tau = \tau_b$ and hence Equation (2.10) holds for v_π as well as for $v_{\hat{\pi}}$ with the same cut-off b. In particular we have $\pi(t) = v(t, b(t)) = \hat{v}(t, b(t)) = \hat{\pi}(t)$.

At time T we clearly have $v(T, x) = \hat{v}(T, x)$ for all $x \in \mathbb{R}$. Using this as induction basis we obtain by Equation (2.10) for $t < T$

$$\pi(t) = z(t, b(t)) + P_{t,t+1} v(b(t)) = z(t, b(t)) + P_{t,t+1} \hat{v}(b(t)) = \hat{\pi}(t).$$

Therefore the dynamic programming principle (2.9) implies $v(t, x) = \hat{v}(t, x)$ for all $x \in \mathbb{R}$. □

2.2.5. Relation to optimal stopping

In this section we use our results to gain new insights into the structure of optimal stopping problems. Especially, we provide a new closed form characterization of the option value in general optimal stopping problems as an expectation over constrained processes. We consider the standard optimal stopping problem, where the agent optimizes her expected payoff $V(\tau)$ over the set of stopping rules $\tau \in \mathcal{T}$.[5] We say that $\tau^\star$ is the minimal optimal stopping rule, if it is the minimal stopping rule satisfying

$$V(\tau^\star) = \sup_{\tau \in \mathcal{T}} V(\tau). \tag{2.14}$$

[5] Recall that the value V and its generalization $V_{t,x}$ are defined in Equations (2.2) and (2.6), respectively.

The notion of implementability introduced in Definition 2.1.13 generalizes the notion of optimality in the sense that every minimal optimal stopping rule is implementable by a transfer of zero. Hence, we get the following immediate corollary of Proposition 2.2.5, which reproduces the well known result from optimal stopping theory (see e.g. [42] for the result in a continuous-time framework)

Corollary 2.2.10. *The minimal optimal stopping rule is a cut-off rule.*

Proof. As every minimal optimal stopping rule is implementable, the minimal optimal stopping rule is a cut-off rule by Proposition 2.2.5. □

We define the agent's option value by[6]

$$w(t,x) = \sup_{\tau \in \mathcal{T}_{t,T}} V_{t,x}(\tau) - V_{t,x}(t).$$

Similar arguments as in the proof of Proposition 2.2.1 show that w equals the sum of marginal incentives z if the agent uses the optimal stopping rule

$$w(t,x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\sum_{s=t}^{\tau-1} z(s,X_s) \,|\, X_t = x\right] .$$

Note that by definition the option value $w(t,x)$ is nonnegative as the agent can always choose $\tau = t$ and stop immediately. The option value plays an important role in many economic applications of optimal stopping, especially in irreversible investment and search theory.

Example 2.2.11 (Irreversible Investment)**.** *Consider for example the situation of a firm, facing uncertainty over future market conditions. It has to decide if and when to invest in a new factory. Here, $V_{t,x}(t)$ is the net present value (NPV) of investing at time t. A simple investment strategy would be to invest once the NPV is positive. The real options literature (cf. [22]) highlights the fact that this is not an optimal strategy as there is the option value of waiting. Instead it is optimal to invest at the first time when the option value of waiting becomes zero, i.e. the stopping rule*

$$\tau^\star = \inf\{t \in \mathbb{T} \,|\, w(t,X_t) = 0\}\,,$$

is optimal in (2.14) (cf. [67, Chapter 1, Theorem 1.9])

Hence, to characterize optimal behavior of the agent in stopping problems it is important to characterize the option value w. The next result yields a representation of w in terms of constrained processes. Moreover, we characterize the optimal stopping cut-off b.

[6] In the mathematical finance literature w is called the time value of V.

Proposition 2.2.12 (Characterization of the Option Value).
Let b be a finite cut-off. If b satisfies

$$0 = \mathbb{E}\left[\sum_{s=t}^{T-1} z(s, \tilde{X}_s) \mid \tilde{X}_t = b(t)\right] , \tag{2.15}$$

for all $t \in \mathbb{T}$, then τ_b is the minimal optimal stopping rule and the option value of waiting is the expected sum over future marginal incentives evaluated at the constrained process, i.e. for all $x \in \mathbb{R}$ and all $t \in \mathbb{T}$

$$w(t, x) = \mathbb{E}\left[\sum_{s=t}^{T-1} z(s, \tilde{X}_s) \mid \tilde{X}_t = x \wedge b(t)\right] . \tag{2.16}$$

Under Assumption 2.1.9 the converse holds true as well. If τ_b is optimal, then b satisfies Equation (2.15) for all $t \in \mathbb{T}$.

Proof. Assume that Equation (2.15) holds. Then Theorem 2.2.8 implies that τ_b is implemented by the zero transfer which means that τ_b is the minimal optimal stopping time in (2.14). Equation (2.16) follows from Equation (2.13).

Now assume that τ_b is a minimal optimal stopping rule in (2.14). Hence, it is implemented by the zero transfer. By Theorem 2.2.8 it is also implemented by the transfer $t \mapsto \sum_{s=t}^{T-1} \tilde{\mathbb{E}}[z(s, \tilde{X}_s) \mid \tilde{X}_t = b(t)]$. Then Proposition 2.2.9 implies Equation (2.15). □

2.3. Application to search problems

In this section, we use the characterization of the transfer derived in Section 2.2 to analyze search problems. In a search problem, an agent chooses every period t between accepting an offer $X_t \in \mathbb{R}_+$ and waiting for better offers. Here X_t denotes the wage which is payed to the agent in every future period (e.g. week or month) if she accepted the job offer at time t. In periods before accepting an offer the agent receives a nonnegative benefit $\beta(t) \in \mathbb{R}_+$. For example, the agent could be an unemployed worker who sequentially receives job offers. The benefit is provided by an unemployment agency who wants the agent to accept certain offers but that does not observe the offers the agent receives. In the following we analyze how the unemployment agency can influence the agent's search behavior by paying benefits that only depend on the time the worker has spent in unemployment.

As before we denote by $\tau \in \mathcal{T}$ a search strategy (stopping rule)[7]. The offers X arrive according to a Markov process such that Assumptions 2.1.1, 2.1.2, 2.1.3 and 2.1.9 of Section 2.1 are satisfied. The agent discounts the future with a factor $\delta \in (0, 1)$. After accepting an offer the agent receives the value of the offer forever, such that her

[7]Note that by the definition of $\mathcal{T} = \mathcal{T}_{0,T}$ the agent needs to accept an offer before T.

discounted expected average utility equals

$$(1-\delta)\mathbb{E}\left[\sum_{s=0}^{\tau-1}\delta^s u(\beta(s)) + \sum_{s=\tau}^{\infty}\delta^s u(X_\tau)\right] = \mathbb{E}\left[\left(\sum_{s=0}^{\tau-1}(1-\delta)\delta^s u(\beta(s))\right) + \delta^\tau u(X_\tau)\right] . \tag{2.17}$$

The utility function $u : \mathbb{R}_+ \to [\underline{u}, \overline{u}]$ is twice differentiable, surjective, increasing and concave. Note that we do not assume the utility to be bounded, i.e. $\underline{u}, \overline{u} \in \overline{\mathbb{R}}$.

We generalize the definition of implementability when the agent is risk-averse with respect to the benefit in the natural way.

Definition 2.3.1 (Implementability)**.** A search strategy $\tau^\star$ is implemented by a benefit β if $\tau^\star$ is the minimal stopping rule satisfying

$$\sup_{\tau\in\mathcal{T}}\mathbb{E}\left[\sum_{s=0}^{\tau-1}(1-\delta)\delta^s u(\beta(s)) + \delta^\tau u(X_\tau)\right] = \mathbb{E}\left[\sum_{s=0}^{\tau^\star-1}(1-\delta)\delta^s u(\beta(s)) + \delta^{\tau^\star} u(X_{\tau^\star})\right] . \tag{2.18}$$

For every benefit function β we define the transfer $\pi_\beta : \mathbb{T} \to \mathbb{R}$ by

$$\pi_\beta(t) = \sum_{s=0}^{t-1}(1-\delta)\delta^s u(\beta(s)) . \tag{2.19}$$

Using the definition of π_β from Equation (2.19) the search problem of the agent simplifies to

$$\sup_{\tau\in\mathcal{T}}\mathbb{E}\left[\delta^\tau u(X_\tau) + \pi_\beta(\tau)\right] . \tag{2.20}$$

From Proposition 2.2.9 we derive uniqueness of the benefit implementing a cut-off rule.

Corollary 2.3.2. *The benefit implementing a given finite cut-off rule is unique.*

Proof. Assume that a cut-off rule is implemented by a benefit β. Then it is implemented by the transfer $\pi_\beta(t) = \sum_{s=0}^{t-1}(1-\delta)\delta^s u(\beta(s))$. By Proposition 2.2.9 the term $\pi_\beta(t) - \pi_\beta(t-1) = (1-\delta)\delta^t u(\beta(t))$ is independent of β, which implies uniqueness of $\beta(t)$ for any t as u is strictly increasing. ☐

Theorem 2.2.8 shows that there exists a unique (up to a constant) transfer such that a given cut-off rule $\tau_b = \min\{t \in \mathbb{T} \mid X_t \geq b(t)\}$ is optimal for the agent in (2.20). Due to the risk-aversion of the agent and the restriction to nonnegative benefits β, it is, however, not clear that for every transfer π there exists a benefit function β such that $\pi = \pi_\beta$. In the next sections we will explicitly calculate the transfers for the two standard setups of search with and without recall. We then use this transfer to analyze which search strategies can be implemented and explicitly derive the benefits β implementing them.

2.3.1. Search without recall

The process X models the value of the agent's current offer. The agent can not recall offers. If she does not accept the offer X_t at time t, the offer expires and there arrives a new offer at time $t+1$. The offers X_t, $t \in \mathbb{T}$, are identically and independently distributed according to the absolutely continuous, strictly increasing distribution function $G : \mathbb{R}_+ \to [0,1]$. Then Proposition 2.1.8 ensures that all required assumptions are satisfied. Let us denote by $\mu = \int u(x)dG(x) < \infty$ the expected utility of a job offer. The agent's marginal incentive to delay the decision of taking a job by one period equals

$$z(t,x) = \mathbb{E}\left[\delta^{t+1}u(X_{t+1}) - \delta^t u(X_t) \,|\, X_t = x\right] = \delta^t\left(\delta\mu - u(x)\right) .$$

As u is increasing the marginal incentive is decreasing and thus the dynamic single crossing condition (Assumption 2.1.11) is satisfied. Furthermore if u is unbounded the marginal incentive z is unbounded as well (Assumption 2.1.12) and by Proposition 2.2.6 only finite cut-off rules are implementable.

Let b be a finite cut-off. By definition of the constrained process $\tilde{X}$ in Equation (2.11) the values $\tilde{X}_t$ are independently and identically distributed according to the distribution function

$$\begin{cases} 1 & \text{for all } x \geq b(t) \\ G(x) & \text{for all } x < b(t) \end{cases} .$$

It follows that for all $t < s$ the expected marginal incentive evaluated at the constrained process is given by

$$\mathbb{E}\left[z(s,\tilde{X}_s) \,|\, \tilde{X}_t = b(t)\right] \;=\; \delta^s\left(\delta\mu - \int_0^{b(s)} u(x)dG(x) - u(b(s))(1-G(b(s)))\right) .$$

In particular, $\mathbb{E}\left[z(s,\tilde{X}_s) \,|\, \tilde{X}_t = b(t)\right]$ is independent of t. By Theorem 2.2.8 using the cut-off rule τ_b maximizes the agents utility in Equation (2.17) if the benefits β are chosen in such a way that the associated transfer π_β equals up to a constant

$$\delta^t(\delta\mu - u(b(t))) + \sum_{s=t+1}^{T-1} \delta^s\left(\delta\mu - \int_0^{b(s)} u(x)dG(x) - u(b(s))(1-G(b(s)))\right) .$$

This leads to a formula for the benefits paid in every period.

Proposition 2.3.3 (Benefits for Search without Recall)**.** *Let b be a finite cut-off. The search strategy τ_b is implementable if and only if*

$$\frac{u(b(t))}{1-\delta} - \frac{\delta}{1-\delta}\int_{\mathbb{R}_+} u(\max\{b(t+1),x\})dG(x) \in [\underline{u},\overline{u}] \text{ for all } t < T . \tag{2.21}$$

In this case τ_b is implemented by the benefit

$$\beta(t) = u^{-1}\left(\frac{u(b(t))}{1-\delta} - \frac{\delta}{1-\delta}\int_{\mathbb{R}_+} u(\max\{b(t+1),x\})dG(x)\right) . \tag{2.22}$$

While in the risk neutral case without the restriction to nonnegative benefits every cut-off is implementable, Proposition 2.3.3 shows that in the case of bounded utility functions the class of implementable search strategies is further restricted by Equation (2.21).

Example 2.3.4. *Let G be the uniform distribution on $[0,1]$, i.e. $G(x) = \min\{x, 1\}$, and assume that the agent exhibits constant relative risk aversion α, i.e.*

$$u(x) = \begin{cases} \frac{x^{1-\alpha}}{1-\alpha} & \text{for } \alpha \neq 1 \\ \log(x) & \text{for } \alpha = 1 \end{cases} . \tag{2.23}$$

First note that $\mu = (1-\alpha)^{-1} \int_0^1 x^{1-\alpha} dx$ is only finite for $\alpha \in [0, 2)$. Then a finite cut-off rule is implementable if and only if $\alpha = 1$ or for all $t < T$

$$b(t)^{1-\alpha} \geq \frac{\delta}{2-\alpha} \left((1-\alpha) b(t+1)^{2-\alpha} + 1 \right) .$$

The benefit equals

$$\beta(t) = \begin{cases} b(t)^{1-\delta} \exp\left(\frac{\delta}{1-\delta} (1 - b(t+1)) \right) & \text{for } \alpha = 1 \\ \left(\frac{b(t)^{1-\alpha}}{1-\delta} - \frac{\delta}{(2-\alpha)(1-\delta)} \left((1-\alpha) b(t+1)^{2-\alpha} + 1 \right) \right)^{\frac{1}{1-\alpha}} & \text{for } \alpha \neq 1 \end{cases} .$$

2.3.2. Search with recall

In this section we solve the case of search with recall. In every period t the worker receives a job offer with nonnegative wage Y_t. The wages are independent and identically distributed according to the absolutely continuous, strictly increasing distribution $G : \mathbb{R}_+ \to [0,1]$. Again we denote by $\mu = \int u(x) dG(x)$ the expected utility of a job offer. In this section the worker can recall past offers. The process $X_t = \max_{s \leq t} Y_s$ models the highest offer the worker received before time t. Then it follows from Proposition 2.1.8 that all required assumptions are satisfied. The agent's marginal incentive to delay the decision of taking a job equals

$$z(t,x) = \delta^t \mathbb{E}\left[\delta u(X_{t+1}) - u(X_t) | X_t = x \right] = \delta^{t+1} \left(u(x) G(x) + \int_x^\infty u(y) dG(y) \right) - \delta^t u(x).$$

Taking derivatives yields that $z_x(t,x) = -\delta^t u'(x) \left(1 - \delta G(x)\right)$ and thus the dynamic single crossing condition (Assumption 2.1.11) is satisfied. Let b be a finite nonincreasing cut-off, i.e. $b(t) \geq b(t+1)$. Since X is pathwise nondecreasing one can verify, that the process $\tilde{X}_t = b(t) \wedge X_t$ satisfies the definition of the constrained process $\tilde{X}$ in Equation (2.11). It follows that for all $t \leq s$ the expected marginal incentive evaluated at the constrained process is given by

$$\mathbb{E}\left[z(s, \tilde{X}_s) | \tilde{X}_t = b(t) \right] = \delta^t \left(\delta \int_{\mathbb{R}_+} u(\max\{b(s), x\}) dG(x) - u(b(s)) \right)$$

This leads the following proposition.

Proposition 2.3.5 (Benefits for Search with Recall). *Let b be a finite nonincreasing cut-off. The search strategy τ_b is implementable if and only if*

$$\frac{u(b(t))}{1-\delta} - \frac{\delta}{1-\delta}\int_{\mathbb{R}_+} u(\max\{b(t),x\})dG(x) \in [\underline{u},\overline{u}] \textit{ for all } t < T\,. \tag{2.24}$$

In this case τ_b is implemented by the benefit

$$\beta(t) = u^{-1}\left(\frac{u(b(t))}{1-\delta} - \frac{\delta}{1-\delta}\int_{\mathbb{R}_+} u(\max\{b(t),x\})dG(x)\right)\,. \tag{2.25}$$

We obtain the following corollary showing that for the case with recall the monotonicity of the cut-off implies the monotonicity of the benefit.

Corollary 2.3.6. *Let b be a nonincreasing cut-off and β the unique benefit implementing it, then β is nonincreasing.*

Proof. The result follows as the right-hand side of Equation (2.25) is increasing in $b(t)$. □

Finally we obtain a result comparing the benefit implementing a decreasing cut-off with and without recall.

Corollary 2.3.7. *Let b be a nonincreasing cut-off. Then the benefit implementing it if the agent can recall past offers is weakly smaller than the benefit if the agent cannot recall past offers.*

Proof. The result follows as the right-hand side of Equation (2.22) is weakly larger than the right-hand side of Equation (2.25). □

Example 2.3.8. *We take the same framework as in Example 2.3.4, i.e. $G(x) = \min\{x,1\}$ is the uniform distribution on $[0,1]$ and the utility function exhibits constant relative risk aversion α (c.f. Equation (2.23)). Again note that $\mu = (1-\alpha)^{-1}\int_0^1 x^{1-\alpha}dx$ is only finite for $\alpha \in [0,2)$. Then a finite nonincreasing cut-off rule is implementable if and only if $\alpha = 1$ or for all $t < T$*

$$0 \;\leq\; (1-\alpha)\delta b(t)^{2-\alpha} - (2-\alpha)b(t)^{1-\alpha} + \delta\,.$$

The benefit equals

$$\beta(t) \;=\; \begin{cases} b(t)^{1-\delta}\exp\left(\frac{\delta}{1-\delta}(1-b(t))\right) & \textit{for } \alpha = 1 \\ \left(\frac{(1-\alpha)\delta b(t)^{2-\alpha}-(2-\alpha)b(t)^{1-\alpha}+\delta}{(2-\alpha)(1-\delta)}\right)^{\frac{1}{1-\alpha}} & \textit{for } \alpha \neq 1 \end{cases}\,.$$

2.4. Appendix

Proof of Proposition 2.1.8. We first consider the case of multiplicative increments. By definition

$$\mathbb{E}\left[|X_{t+1}^2| \,|\, X_t = x\right] \;=\; x^2\mathbb{E}\left[\epsilon_{t+1}^2\right] = x^2\sigma^2 \le \sigma^2(1+|x|^2)$$

and hence X is of polynomial growth of order $p = 2$. For every $x < x'$ and any increasing function ϕ we have that

$$\mathbb{E}\left[\phi(X_{t+1}) \,|\, X_t = x'\right] - \mathbb{E}\left[\phi(X_{t+1}) \,|\, X_t = x\right] = \mathbb{E}\left[\phi(\epsilon_{t+1}x') - \phi(\epsilon_{t+1}x)\right] \ge 0.$$

Setting $\phi(x) = \mathbf{1}_{\{x \ge z\}}$ yields that the process X has monotone transitions. Next, let ϕ be a continuous, polynomial growth function. Appealing to the dominated convergence theorem yields

$$\begin{aligned} &\lim_{h\to 0} \left|\mathbb{E}\left[\phi(X_{t+1}) \,|\, X_t = x+h\right] - \mathbb{E}\left[\phi(X_{t+1}) \,|\, X_t = x\right]\right| \\ &\le \lim_{h\to 0} \mathbb{E}\left[|\phi(\epsilon_{t+1}(x+h)) - \phi(\epsilon_{t+1}x)|\right] = \mathbb{E}\left[\lim_{h\to 0} |\phi(\epsilon_{t+1}(x+h)) - \phi(\epsilon_{t+1}x)|\right] = 0\,. \end{aligned}$$

Hence X has continuous transitions.

Let us now turn to the additive random walk. We have

$$\mathbb{E}\left[|X_{t+1}^2| \,|\, X_t = x\right] = \mathbb{E}\left[(x+\epsilon_{t+1})^2\right] \le 2(x^2+\sigma^2)$$

and hence X is of polynomial growth of order $p = 2$. For every $x < x'$ and any increasing function ϕ we have that

$$\mathbb{E}\left[\phi(X_{t+1}) \,|\, X_t = x'\right] - \mathbb{E}\left[\phi(X_{t+1}) \,|\, X_t = x\right] = \mathbb{E}\left[\phi(x'+\epsilon_{t+1}) - \phi(x+\epsilon_{t+1})\right] \ge 0.$$

As in the case of a multiplicative random walk setting $\phi(x) = \mathbf{1}_{\{x \ge z\}}$ yields that the process X has monotone transitions. Finally, let ϕ be a continuous, polynomial growth function. Again we employ the dominated convergence theorem to obtain

$$\begin{aligned} &\lim_{h\to 0} \left|\mathbb{E}\left[\phi(X_{t+1}) \,|\, X_t = x+h\right] - \mathbb{E}\left[\phi(X_{t+1}) \,|\, X_t = x\right]\right| \\ &\qquad \le \mathbb{E}\left[\lim_{h\to 0} |\phi(x+h+\epsilon_{t+1}) - \phi(x+\epsilon_{t+1})|\right] = 0\,. \end{aligned}$$

□

In the case of search without recall the conditional expectation $\mathbb{E}[f(X_{t+1})|X_t = x] = \mathbb{E}[f(X_{t+1})]$ does not depend on x and t. Therefore all claims of the proposition follow immediately.

If the agent can recall past offers we have $\mathbb{E}[f(X_{t+1})|X_t = x] = \mathbb{E}[f(X_{t+1} \vee x)]$. The property of monotone transitions follows immediately. Polynomial growth and continuous transition follow from the integrability of X_{t+1}.

Proof of Proposition 2.2.1. Let $\tau \in \mathcal{T}$ be an arbitrary stopping rule. We will show that

$$V(\tau) = \mathbb{E}\left[\sum_{t=0}^{\tau-1} z(t, X_t)\right] + \mathbb{E}\left[g(0, X_0) + \sum_{t=1}^{T} h(t, X_t)\right]$$

which yields the claim. To this end we rewrite $V(\tau)$ as follows

$$V(\tau) = \mathbb{E}\left[\left(\sum_{t=0}^{\tau-1} f(t, X_t) - h(t, X_t)\right) + g(\tau, X_\tau) - h(\tau, X_\tau)\right] + \mathbb{E}\left[\sum_{t=0}^{T} h(t, X_t)\right]. \quad (2.26)$$

Using the the tower property of conditional expectations we can represent the expected payoff $\mathbb{E}\left[g(\tau, X_\tau) - h(\tau, X_\tau)\right]$ as a sum of flow payoffs as follows

$$\begin{aligned}
&\mathbb{E}\left[g(\tau, X_\tau) - h(\tau, X_\tau)\right] \\
&= \mathbb{E}\left[g(0, X_0) - h(0, X_0)\right] + \mathbb{E}\left[\sum_{t=0}^{\tau-1} g(t+1, X_{t+1}) - h(t+1, X_{t+1}) - g(t, X_t) + h(t, X_t)\right] \\
&= \mathbb{E}\left[g(0, X_0) - h(0, X_0)\right] \\
&\qquad + \mathbb{E}\left[\sum_{t=0}^{\tau-1} \mathbb{E}\left[g(t+1, X_{t+1}) - h(t+1, X_{t+1}) - g(t, X_t) + h(t, X_t) \middle| \mathcal{F}_t\right]\right] \\
&= \mathbb{E}\left[g(0, X_0) - h(0, X_0)\right] + \mathbb{E}\left[\sum_{t=0}^{\tau-1} \tilde{z}(t, X_t)\right],
\end{aligned}$$

with

$$\begin{aligned}
\tilde{z}(t, x) &= \mathbb{E}\left[g(t+1, X_{t+1}) - h(t+1, X_{t+1}) \,|\, X_t = x\right] - g(t, x) + h(t, x) \\
&= z(t, x) - f(t, x) + h(t, x).
\end{aligned}$$

Then Equation (2.26) implies

$$V(\tau) = \mathbb{E}\left[\sum_{t=0}^{\tau-1} f(t, X_t) - h(t, X_t) + \tilde{z}(t, X_t)\right] + \mathbb{E}\left[g(0, X_0) - h(0, X_0) + \sum_{t=0}^{T} h(t, X_t)\right],$$

which yields the claim. □

Proof of Lemma 2.2.2. We proceed by backward induction. At time T the value function $v_\pi(T, x) = \pi(T)$ does not depend on x and hence satisfies the assertions of the Lemma. At time $t < T$ the induction hypothesis yields that the function $x \mapsto P_{t,t+1}v(x)$ is continuous, nonincreasing and of polynomial growth of order p. Then the dynamic programming principle (2.9) and the single crossing condition yield the claim for v_π. □

Proof of Lemma 2.2.4. Suppose that Assumption 2.1.9 holds. We show that b is unique. To this end assume that there exist two cut-offs b and $\hat{b}$ such that $\tau = \tau_b = \tau_{\hat{b}}$ and

$b(t) < \hat{b}(t)$ for some $t \in \{0, \dots, T-1\}$. By conditioning on $\mathcal{F}_{t-1}$ and using the Markov property of X we obtain

$$\mathbb{P}[\tau_b = t] = \mathbb{P}[X_t \geq b(t), \tau_b > t-1] = \mathbb{E}[\mathbb{E}[\mathbf{1}_{\{X_t \geq b(t)\}} | X_{t-1}] \mathbf{1}_{\{\tau_b > t-1\}}].$$

Similar considerations for $\tau_{\hat{b}}$ yield

$$\begin{aligned} 0 &= \mathbb{P}[\tau_b = t] - \mathbb{P}[\tau_{\hat{b}} = t] = \mathbb{E}[\mathbb{E}[\mathbf{1}_{\{X_t \geq b(t)\}} - \mathbf{1}_{\{X_t \geq \hat{b}(t)\}} | X_{t-1}] \mathbf{1}_{\{\tau_b > t-1\}}] \\ &= \mathbb{E}[\mathbb{E}[\mathbf{1}_{\{b(t) \leq X_t < \hat{b}(t)\}} | X_{t-1}] \mathbf{1}_{\{\tau_b > t-1\}}]. \end{aligned}$$

By the full support assumption the random variable $\mathbb{E}[\mathbf{1}_{\{b(t) \leq X_t < \hat{b}(t)\}} | X_{t-1}]$ is strictly positive. Moreover, the full support assumption implies that the event $\{\tau_b > t-1\}$ happens with positive probability. This leads to the contradiction

$$\mathbb{E}\left[\mathbb{E}[\mathbf{1}_{\{b(t) \leq X_t < \hat{b}(t)\}} \,|\, X_{t-1}] \mathbf{1}_{\{\tau_b > t-1\}}\right] > 0\,. \qquad \square$$

Proof of Proposition 2.3.3. First assume that Condition (2.21) is satisfied and let β be given by Equation (2.22). Then we have by Equation (2.19)

$$\begin{aligned} \pi_\beta(t) &= \sum_{s=0}^{t-1} \delta^s \left(u(b(s)) - \delta \int_{\mathbb{R}_+} u(\max\{b(s+1), x\}) dG(x) \right) \\ &= \delta^t \int_0^{b(t)} u(x) dG(x) - u(b(t))G(b(t)) - \int_0^{b(0)} u(x) dG(x) + u(b(0))G(b(0)) \\ &\quad - \sum_{s=0}^{t-1} \delta^s \left(\delta\mu - \int_0^{b(s)} u(x) dG(x) - u(b(s))(1 - G(b(s)) \right) \\ &= C + \sum_{s=t}^{T-1} \mathbb{E}\left[z(s, \tilde{X}_s) \,|\, \tilde{X}_t = b(t) \right], \end{aligned}$$

where the constant C is given by

$$\begin{aligned} u(b(0))G(b(0)) &- \int_0^{b(0)} u(x) dG(x) \\ &- \sum_{s=0}^{T} \delta^s \left(\delta\mu - \int_0^{b(s)} u(x) dG(x) - u(b(s))(1 - G(b(s))) \right). \end{aligned}$$

Then Theorem (2.2.8) implies that τ_b is implemented by β.

For the other direction assume that τ_b is implemented by the benefit β. Since β is unique by Corollary 2.3.2 similar calculations as above imply that β satisfies for all $t < T$

$$(1-\delta)\delta^t u(\beta(t)) = u(b(t)) - \delta \int_{\mathbb{R}_+} u(\max\{b(t+1), x\}) dG(x)\,.$$

This implies that Condition (2.21) is satisfied. $\square$

Proof of Proposition 2.3.5. If Condition (2.24) holds true, then the benefit β from Equation (2.25) is well-defined and satisfies

$$\begin{aligned}\pi_\beta(t) &= \sum_{s=0}^{t-1} u(b(s)) - \delta \int_{\mathbb{R}_+} u(\max\{b(s), x\}) dG(x) \\ &= C + \sum_{s=t}^{T-1} \mathbb{E}\left[z(s, \tilde{X}_s) \mid \tilde{X}_t = b(t)\right] ,\end{aligned}$$

with

$$C = \sum_{s=0}^{T-1} u(b(s)) - \delta \int_{\mathbb{R}_+} u(\max\{b(s), x\}) dG(x) \,.$$

The converse direction follows by the same argument as in the proof of Proposition 2.3.3. □

3. Inverse optimal stopping in continous-time

This chapter deals with the analogue of Chapter 2 in a continuous-time framework. Some remarks concerning the differences between the two chapters are in order. First, the model in this chapter is set-up in a Brownian framework and the underlying stochastic process follows time-inhomogeneous diffusion dynamics. Second, the definition of implementability is modified. The notion of implementability is introduced for *subsets* of the state space $[0, T] \times \mathbb{R}$. A set is called implementable if its first hitting time of the diffusion is implemented by some transfer. The notion is strong since it requires optimality of first hitting times for arbitrary initial conditions of the diffusion. This approach allows to derive a uniqueness result without imposing a full support assumption on the diffusion. Moreover, in contrast to Chapter 2, the requirement of minimality of the stopping times in the definition of implementability is dropped.

Section 3.1 introduces the model in detail and in particular provides the notion of implementability. The subsequent sections proceed in the spirit of their discrete-time counterparts. However, the mathematical techniques employed differ considerably. For example, the verification arguments presented in this chapter do not rely on Bellman's optimality principle. In Section 3.2 it is shown that only cut-off regions are strictly implementable. Section 3.3 is devoted to the converse implication. First, the explicit representation of the transfer is formally derived. To this end reflected processes are introduced (Subsection 3.3.1). Moreover, the crucial properties of reflected processes which are required in the following sections, are established. Subsection 3.3.2 employs these results to derive the main result about implementability of cut-off regions. Subsection 3.3.3 presents the some properties of the transfer and Subsection 3.3.4 provides a uniqueness result. In Section 3.4 an integral equation characterizing optimal stopping boundaries is presented.

3.1. Problem formulation

3.1.1. Dynamics

In this chapter we consider optimal stopping problems with finite time horizon $T < \infty$. The underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supports a one-dimensional Brownian motion W. Let $\mathbb{F} = (\mathcal{F}_t)_{t \in [0,T]}$ be the filtration generated by W satisfying the usual assumptions. We denote the set of $\mathbb{F}$-stopping times with values in $[0, T]$ by $\mathcal{T}$. For $t < T$ we refer to $\mathcal{T}_{t,T}$ as the subset of stopping times which take values in $[t, T]$. The process X follows the time-inhomogeneous diffusion dynamics

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t. \tag{3.1}$$

We denote by $\mathcal{L} = \mu\partial_x + \frac{1}{2}\sigma^2\partial_{xx}$ the infinitesimal generator of X. The coefficients $\mu, \sigma : [0, T] \times \mathbb{R} \to \mathbb{R}$ are continuous and Lipschitz continuous in the x variable uniformly in t, i.e. there exists a positive constant L such that

$$|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \quad \leq \quad L|x - y|$$

for all $t \in [0, T]$ and $x, y \in \mathbb{R}$. Under this assumption there exists a unique solution $(X_s^{t,x})_{s \geq t}$ to (3.1) for every initial condition $X_t^{t,x} = x$. Moreover, it follows that the comparison principle holds true (see e.g. [44, Proposition 2.18]): The path of a signal starting at a lower level $x \leq x'$ at time t is smaller than the path of a signal starting in x' at all later times $s > t$

$$X_s^{t,x} \leq X_s^{t,x'} \qquad \mathbb{P} - a.s. \tag{3.2}$$

3.1.2. Payoffs and transfers

As long as the process X is not stopped there is a flow payoff f and at the time of stopping there is a terminal payoff g. The payoffs $f, g : [0, T] \times \mathbb{R} \to \mathbb{R}$ depend on time and the value of the signal. Thus the expected payoff for using a stopping time $\tau \in \mathcal{T}_{t,T}$ equals

$$W(t, x, \tau) = \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x})ds + g(\tau, X_\tau^{t,x})\right],$$

given that the signal starts in $x \in \mathbb{R}$ at time $t \in [0, T]$. We assume that the payoff function f is continous and Lipschitz continuous in the x variable uniformly in t. Moreover, we suppose that $g \in C^{1,2}([0, T] \times \mathbb{R})$ with bounded derivatives.

We will analyze how preferences over stopping times change if there is an additional payoff which only depends on time.

Definition 3.1.1. A measurable, bounded function $\pi : [0, T] \to \mathbb{R}$ is called a transfer.

We define the value function of the stopping problem with payoffs f and g and an additional transfer π by

$$v^\pi(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}} \left(W(t, x, \tau) + \mathbb{E}\left[\pi(\tau)\right]\right). \tag{3.3}$$

Moreover we introduce for every $t \in [0, T]$ the stopping region

$$D_t^\pi = \{x \in \mathbb{R} \,|\, v^\pi(t, x) = g(t, x) + \pi(t)\}.$$

3.1.3. Implementability

A measurable set $A \subset [0,T] \times \mathbb{R}$ is called *time-closed* if for each time $t \in [0,T]$ the slice $A_t = \{x \in \mathbb{R} \,|\, (t,x) \in A\}$ is a closed subset of $\mathbb{R}$. Let X start in $x \in \mathbb{R}$ at time $t \in [0,T]$. For a time-closed set A we introduce the first time when X hits A by

$$\tau_A^{t,x} = \inf \left\{ s \geq t \,|\, X_s^{t,x} \in A_s \right\} \wedge T.$$

We now come to the definition of implementability.

Definition 3.1.2 (Implementability)**.** A time-closed set A is implemented by a transfer π if the stopping time $\tau_A^{t,x}$ is optimal in (3.3), i.e. for every $t \in [0,T]$ and $x \in \mathbb{R}$

$$v^\pi(t,x) = W(t,x,\tau_A^{t,x}) + \mathbb{E}\left[\pi(\tau_A^{t,x})\right] .$$

For a time-closed set A a necessary condition for implementability is that each slice A_t is included in the stopping region D_t^π. Indeed, let A be implemented by π and let $t \in [0,T]$ and $x \in A_t$. Then we have $\tau_A^{t,x} = t$. Since $\tau_A^{t,x}$ is optimal, this implies $v^\pi(t,x) = g(t,x) + \pi(t)$ and hence $x \in D_t^\pi$. Consequently, we have $A_t \subseteq D_t^\pi$.

Observe that the converse inclusion $D_t^\pi \subseteq A_t$ does not necessarily hold true, since optimal stopping times are in general not unique. At some point $(t,x) \in [0,T] \times \mathbb{R}$ it might be optimal to stop immediately ($x \in D_t^\pi$) as well as to wait a positive amount of time until X hits A ($x \notin A_t$). A particularly simple example is the case where X is a martingale and $f(t,x) = 0$ and $g(t,x) = x$. The optional stopping theorem implies that all stopping times $\tau \in \mathcal{T}_{t,T}$ generate the same expected payoff $W(t,x,\tau) = x$. Therefore, every set A is implemented by the zero transfer. The stopping region consists of the whole state space $D_t^0 = \mathbb{R}$.

We introduce the notion of strict implementability, where ambiguity in optimal strategies is ruled out: whenever it is optimal to continue a positive amount of time it is not optimal to stop.

Definition 3.1.3 (Strict implementability)**.** A time-closed set A is strictly implemented by a transfer π if A is implemented by π and $v^\pi(t,x) > g(t,x) + \pi(t)$ for all $x \notin A_t$ and $t \in [0,T]$.

In particular, every strictly implementable set A satisfies $A_t = D_t^\pi$ for the transfer π. Since the stopping regions D_t^π are closed (see Lemma 3.2.1 below) the restriction to time-closed sets is no loss of generality. Any set which is not time-closed can not be strictly implemented.

Note that the notion of implementability generalizes the notion of optimal stopping times. If $\tau_A^{t,x}$ is an optimal stopping time in a stopping problem of the form

$$\sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\tau f(s,X_s^{t,x})ds + g(\tau, X_\tau^{t,x})\right]$$

for all $(t,x) \in [0,T] \times \mathbb{R}$, then it is implemented by the zero transfer.

3.1.4. Single crossing and cut-off regions

Next we introduce the main structural condition on the payoff functions.

Assumption 3.1.4 (Single Crossing)**.** We say that the single crossing condition is satisfied if the mapping $x \mapsto f(t,x) + (\partial_t + \mathcal{L})g(t,x)$ is nonincreasing. If this monotonicity is strict, then we say that the strict single crossing condition holds.

Moreover, we define a special subclass of time-closed sets.

Definition 3.1.5 (Cut-off regions)**.** A time-closed set A is called a cut-off region if there exists a function $b : [0,T] \to \overline{\mathbb{R}}$ such that $A_t = [b(t), \infty)$. In this case we call b the associated cut-off and we write

$$\tau_A^{t,x} = \tau_b^{t,x} = \inf\{s \geq t \mid X_s^{t,x} \geq b(s)\} \wedge T$$

for $(t,x) \in [0,T] \times \mathbb{R}$. We call τ_b a cut-off rule. We say that a cut-off region A is regular, if the associated cut-off $b : [0,T] \to \mathbb{R}$ is càdlàg (i.e. is right continuous and has left limits in $\mathbb{R}$) and has summable downward jumps, i.e.

$$\sum_{0 \leq s \leq t} (\Delta b_s)^- < \infty.$$

3.2. Strictly implementable regions are cut-off regions

For optimal stopping problems it is well-known that under the single crossing condition (or a weaker version of it) only cut-off rules are optimal (see e.g. [46], [42] or [83]). In this section we show that this result holds more generally for implementable stopping times: Only cut-off regions can be strictly implemented.

We first state the following regularity result about v^π.

Lemma 3.2.1. *For every transfer π and every $t \in [0,T]$ the mapping $x \mapsto v^\pi(t,x)$ is Lipschitz continuous. In particular, the stopping region D_t^π is closed.*

Proof. Fix $t \in [0,T]$ and $x, y \in \mathbb{R}$. By Lipschitz continuity of f and g we have

$$\begin{aligned} |v^\pi(t,x) - v^\pi(t,y)| &\leq \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\tau \left|f(s, X_s^{t,x}) - f(s, X_s^{t,y})\right| ds + \left|g(\tau, X_\tau^{t,x}) - g(\tau, X_\tau^{t,y})\right|\right] \\ &\leq C\mathbb{E}\left[\sup_{s \in [t,T]} \left|X_s^{t,x} - X_s^{t,y}\right|\right]. \end{aligned}$$

By the well-known moment estimate for solutions of stochastic differential equations (see e.g. [51, Theorem 3.2]) there exists a constant $\tilde{C}$ such that

$$\mathbb{E}\left[\sup_{s \in [t,T]} \left|X_s^{t,x} - X_s^{t,y}\right|\right] \leq \tilde{C}|x - y|.$$

This yields the claim. ☐

The next result states that under the single crossing condition only cut-off regions are strictly implementable.

Proposition 3.2.2. *Assume that the single crossing condition holds true and let A be strictly implemented by some transfer π. Then A is a cut-off region.*

Proof. Fix $t \in [0,T]$. First observe that the single crossing condition implies that $x \mapsto v^\pi(t,x) - g(t,x)$ is nonincreasing. Indeed, Itô's formula applied to $g(\cdot, X)$ yields

$$\begin{aligned} W(t,x,\tau) &= g(t,x) + \mathbb{E}\left[\int_t^\tau \left(f(s,X_s^{t,x}) + (\partial_t + \mathcal{L})g(s,X_s^{t,x})\right) ds\right] \\ &\quad + \mathbb{E}\left[\int_t^\tau g_x(s,X_s^{t,x})\sigma(s,X_s^{t,x}) dW_s\right] \end{aligned}$$

for every $x \in \mathbb{R}$ and $\tau \in \mathcal{T}_{t,T}$. Since g_x is bounded and σ has linear growth the process $\int_t^\cdot g_x(s,X_s^{t,x})\sigma(s,X_s^{t,x})dW_s$ is a martingale. It follows from the comparison principle (3.2) and the single crossing condition that for $x \leq y$

$$\begin{aligned} v^\pi(t,x) - g(t,x) &= \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\tau \left(f(s,X_s^{t,x}) + (\partial_t + \mathcal{L})g(s,X_s^{t,x})\right) ds + \pi(\tau)\right] \\ &\geq \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\tau \left(f(s,X_s^{t,y}) + (\partial_t + \mathcal{L})g(s,X_s^{t,y})\right) ds + \pi(\tau)\right] \\ &= v^\pi(t,y) - g(t,y). \end{aligned}$$

This implies that $y \in D_t^\pi$ if $y \geq x$ and $x \in D_t^\pi$. Hence D_t^π is an interval which is unbounded on the right. By Lemma 3.2.1 the set D_t^π is closed. Hence there exists some $b(t) \in \overline{\mathbb{R}}$ such that $D_t^\pi = [b(t), \infty)$. This implies that A is a cut-off region since $A_t = D_t^\pi$ by the definition of strict implementability. □

3.3. Implementability of cut-off regions

In this section we prove that the converse implication of Proposition 3.2.2 holds true as well: Every regular cut-off region is implementable. We derive a closed form representation for the transfer in terms of the reflected version of X in Subsection 3.3.1. In Subsection 3.3.2 we verify that this candidate solution to the inverse optimal stopping problem indeed implements cut-off regions. The main properties of the transfer are presented in Subsection 3.3.3. In Subsection 3.3.4 we provide a uniqueness result for transfers implementing a cut-off region.

3.3.1. Reflected SDEs and a formal derivation of the candidate transfer

A solution to a reflected stochastic differential equation (RSDE) is a pair of processes $(\tilde{X}, l)$, where the process $\tilde{X}$ evolves according to the dynamics of the associated SDE (3.1) below a given barrier $b : [0,T] \to \mathbb{R}$ and is pushed below the barrier by the process l whenever it tries to exceed b. Next we give a formal definition.

Definition 3.3.1. Let $b : [0,T] \to \mathbb{R}$ be càdlàg, $t \in [0,T]$ a fixed point in time and $\tilde{\xi} \leq b(t)$ a $\mathcal{F}_t-$measurable square-integrable random variable. A pair $(\tilde{X}, l)$ of adapted processes (with càdlàg trajectories) is called a solution to the stochastic differential equation (3.1) reflected at b with initial condition $(t, \tilde{\xi})$ if it satisfies the following properties.

1. $\tilde{X}$ is constrained to stay below the barrier, i.e. $\tilde{X}_s \leq b(s)$ almost surely for every $s \in [t,T]$.

2. For every $s \in [t,T]$ the following integral equation holds almost surely

$$\tilde{X}_s = \tilde{\xi} + \int_t^s \mu(r, \tilde{X}_r)dr + \int_t^s \sigma(r, \tilde{X}_r)dW_r - l_s\,. \tag{3.4}$$

3. The process l is nondecreasing and only increases when $\tilde{X}_t = b(t)$, i.e.

$$\int_t^T (b(s) - \tilde{X}_s)dl_s = 0\,. \tag{3.5}$$

To stress the dependence of $\tilde{X}$ on the initial value we sometimes write $\tilde{X}^{t,\tilde{\xi}}$.

Remark 3.3.2. Consider the situation where b has a downward jump at time t and $\tilde{X}$ is above $b(t)$ shortly before time t, i.e. $\tilde{X}_{t-}(\omega) \in (b(t), b(t-)]$ for some $\omega \in \Omega$. Since $\tilde{X}_t \leq b(t)$ the reflected process $\tilde{X}$ has a downward jump at time t as well. Equation (3.4) implies that l has an upward jump at time t. Then Equation (3.5) yields that $\tilde{X}$ is on the barrier at time t, i.e. $\tilde{X}_t = b(t)$. Hence, the jump of b is rather absorbed by $\tilde{X}$ than truly reflected (which would mean $\tilde{X}_t = 2b(t) - \tilde{X}_{t-}$). In this sense our definition of $\tilde{X}$ coincides with the maximal version of X which stays below b. This property is crucial in the proof of Theorem 3.3.5. Existence and uniqueness of $\tilde{X}$ are established in [72]. We also refer to [79] who allow for general modes of reflection. For results about RSDEs with "true" jump reflections we refer to [18].

A formal derivation

Here we establish the link between inverse optimal stopping problems and RSDEs and derive the representation of a transfer implementing a cut-off region. To this end assume that the cut-off region $A = [b(t), \infty)$ is implemented by a transfer π. Without loss of generality we assume that $\pi(T) = 0$ (else take $\tilde{\pi}(t) = \pi(t) - \pi(T)$). Since we are only interested in a formal derivation here, we make some regularity assumptions. We assume that the value function of the stopping problem (3.3) is smooth ($v^\pi \in C^{1,2}([0,T] \times \mathbb{R})$) and that b is continuous such that $\tilde{X}$ has continuous paths as well. Then v^π satisfies (see e.g. [67, Chapter IV])

$$\begin{aligned} \min\{-(\partial_t + \mathcal{L})v^\pi - f, v^\pi - (g+\pi)\} &= 0 \\ v^\pi(T,\cdot) &= g(T,\cdot) \end{aligned}$$

and b is the free boundary of this variational partial differential equation. In particular, below the cut-off b the value function v^π satisfies the continuation equation

$$(\partial_t + \mathcal{L})v^\pi(t,x) = -f(t,x)$$

for all $x \le b(t)$. On the cut-off, v^π satisfies the Dirichlet boundary condition

$$v^\pi(t,b(t)) = g(t,b(t)) + \pi(t)$$

for all $t \in [0,T]$. Moreover, if b is sufficiently regular the smooth fit principle

$$v_x(t,b(t)) = g_x(t,b(t))$$

holds for all $t \in [0,T]$ (see e.g. [67, Section 9.1]) Then Itô's formula implies

$$\begin{aligned}
\mathbb{E}\left[g(T,\tilde{X}_T^{t,b(t)})\right] &= \mathbb{E}\left[v^\pi(T,\tilde{X}_T^{t,b(t)})\right] \\
&= v^\pi(t,b(t)) + \mathbb{E}\left[\int_t^T (\partial_t+\mathcal{L})v^\pi(s,\tilde{X}_s^{t,b(t)})ds - \int_t^T v_x(s,\tilde{X}_s^{t,b(t)})dl_s\right] \\
&= g(t,b(t)) + \pi(t) - \mathbb{E}\left[\int_t^T f(s,\tilde{X}_s^{t,b(t)})ds + \int_t^T g_x(s,\tilde{X}_s^{t,b(t)})dl_s\right].
\end{aligned}$$

A further application of Itô's formula yields the following representation of π

$$\begin{aligned}
\pi(t) &= \mathbb{E}\left[g(T,\tilde{X}_T^{t,b(t)}) + \int_t^T f(s,\tilde{X}_s^{t,b(t)})ds + \int_t^T g_x(s,\tilde{X}_s^{t,b(t)})dl_s\right] - g(t,b(t)) \\
&= \mathbb{E}\left[\int_t^T f(s,\tilde{X}_s^{t,b(t)}) + (\partial_t+\mathcal{L})g(s,\tilde{X}_s^{t,b(t)})ds\right]. \qquad (3.6)
\end{aligned}$$

In Theorem 3.3.5 below we verify that Equation (3.6) indeed leads to a transfer π implementing A. The proof does neither rely on any analytic methods nor on results from the theory of partial differential equations. Instead we employ purely probabilistic arguments based on the single crossing condition and comparison results for SDEs and RSDEs. This methodology requires weak regularity assumptions on the model parameters. In particular there is no ellipticity condition on σ and b is allowed to have jumps.

Properties of RSDEs

The next proposition proves auxiliary results about RSDEs which we will use in the proof of Theorem 3.3.5. There is a broad literature on RSDEs including comparison results (see e.g. [13]). To the best of our knowledge the comparison principles for RSDE with càdlàg barriers and summable downward jumps as needed for our result have not been shown before. While all results follow by standard arguments we give a proof in the Appendix for the convenience of the reader. For the existence and uniqueness result we refer to [72].

Proposition 3.3.3. *For every regular[1] cut-off b there exists a unique solution $\tilde{X}$ to the RSDE (3.4). The process l is given by*

$$l_s = \sup_{t \le r \le s} (\tilde{\xi} + \int_t^r \mu(u, \tilde{X}_u) du + \int_t^r \sigma(u, \tilde{X}_u) dW_u - b(r))^+. \tag{3.7}$$

Moreover, $\tilde{X}$ satisfies

1. *(Square Integrability)* $\mathbb{E}\left[\sup_{t \le s \le T} (\tilde{X}_s^{t,\xi})^2\right] < \infty$ *for all* $t \in [0, T]$.
2. *(Minimality)* $\tilde{X}_s^{t,\xi} 1_{\{s < \tau_b\}} = X_s^{t,\xi} 1_{\{s < \tau_b\}}$ *a.s. for all* $s \in [t, T]$.
3. *(Comparison Principle for the Reflected Process) If* $\xi_1 \le \xi_2$ *a.s., then for* $s \in [t, T]$ *we have* $\tilde{X}_s^{t,\xi_1} \le \tilde{X}_s^{t,\xi_2}$ *a.s.*
4. *(Moment Estimate) For* $\xi_1, \xi_2 \in L^2(\mathcal{F}_t)$ *there exists a constant* $K > 0$ *such that* $\mathbb{E}\left[\sup_{t \le r \le s} |\tilde{X}_r^{t,\xi_1} - \tilde{X}_r^{t,\xi_2}|^p | \mathcal{F}_t\right] \le K |\xi_1 - \xi_2|^p$ *a.s. for all* $s \in [t, T]$ *and* $p = 1, 2$.
5. *(Comparison Principle for the Original Process)* $\tilde{X}_s^{t,\xi} \le X_s^{t,\xi}$ *a.s. for all* $s \in [t, T]$.
6. *(Left continuity) Let* $t \in [0, T]$ *and* $x \le b(t) \wedge b(t-)$. *Then* $\tilde{X}_t^{s, y \wedge b(s)} \to x$ *in* L^2 *for* $s \nearrow t$ *and* $y \to x$.

Using similar arguments as in [71, Chapter V, Section 6] one can show that $\tilde{X}$ satisfies the strong Markov property.

Definition and Lemma 3.3.4. For $s \ge t$ we define the transition kernel $\tilde{P}_{t,s}$ of $\tilde{X}$ by

$$\tilde{P}_{t,s} \varphi(t, x) = \mathbb{E}\left[\varphi(s, \tilde{X}_s^{t,x})\right]$$

for any Borel measurable, bounded function $\varphi : [0, T] \times \mathbb{R} \to \mathbb{R}$. Then $\tilde{X}$ satisfies for any stopping time $\tau \in \mathcal{T}$ and $u \ge 0$

$$\mathbb{E}\left[\varphi((\tau + u) \wedge T, \tilde{X}_{(\tau+u)\wedge T}) \,|\, \mathcal{F}_\tau\right] = \tilde{P}_{\tau, (\tau+u)\wedge T} \varphi(\tau, \tilde{X}_\tau). \tag{3.8}$$

Moreover, uniqueness of solutions of RSDEs implies the following flow property of $\tilde{X}$. For $t \le r \le s$ and $x \in \mathbb{R}$ we have a.s.

$$\tilde{X}_s^{t,x} = \tilde{X}_s^{r, \tilde{X}_r^{t,x}}. \tag{3.9}$$

[1] see Definition 3.1.5

3.3.2. Regular cut-off regions are implementable

In this section we prove our main theorem stating that every regular cut-off region is implemented by the transfer derived in Subsection 3.3.1.

Theorem 3.3.5. *Assume that the single crossing condition is satisfied. Let A be a regular cut-off region with cut-off b. Then it is implemented by the transfer*

$$\pi(t) = \mathbb{E}\left[\int_t^T f(s, \tilde{X}_s^{t,b(t)}) + (\partial_t + \mathcal{L})g(s, \tilde{X}_s^{t,b(t)})ds\right]. \tag{3.10}$$

Proof. First observe that the cut-off rule $\tau_b^{t,x}$ is a stopping time for all $(t,x) \in [0,T] \times \mathbb{R}$. Indeed, since X has continous paths and b is right-continuous, the Début-theorem (see e.g. [20, Chapter IV, Section 50]) implies $\tau_b^{t,x} \in \mathcal{T}_{t,T}$.

Let π be given by Equation (3.10). For the boundedness and measurability of π we refer to Proposition 3.3.9. We set $h = f + (\partial_t + \mathcal{L})g$. As in the proof of Proposition 3.2.2 we have

$$W(t,x,\tau) = g(t,x) + \mathbb{E}\left[\int_t^\tau h(s, X_s^{t,x})ds\right].$$

Note that we can write π in terms of the transition function $\tilde{P}$ of $\tilde{X}$ as follows

$$\pi(t) = \int_t^T \tilde{P}_{t,s}h(t, b(t))ds.$$

The strong Markov property (Equation (3.8)) of $\tilde{X}$ implies

$$\tilde{P}_{\tau,\tau+u}h(\tau, b(\tau)) = \mathbb{E}\left[h(\tau+u, \tilde{X}_{\tau+u}^{\tau,b(\tau)})|\mathcal{F}_\tau\right]$$

for any stopping time $\tau \in \mathcal{T}$ and $u \geq 0$. Hence we have

$$\pi(\tau) = \mathbb{E}\left[\int_\tau^T h(s, \tilde{X}_s^{\tau,b(\tau)})ds|\mathcal{F}_\tau\right]. \tag{3.11}$$

Fix $t \in [0,T]$ and $x \geq b(t)$. Let $\tau \in \mathcal{T}_{t,T}$ be an arbitrary stopping time. The comparison principle between the original and the reflected process (Property 5.) implies $X_s^{t,x} \geq X_s^{t,b(t)} \geq \tilde{X}_s^{t,b(t)}$ a.s. for every $s \in [t,T]$. From the flow property (Equation (3.9)) and the comparison principle for reflected processes (Property 3.) follows that $\tilde{X}_s^{t,b(t)} = \tilde{X}_s^{\tau,\tilde{X}_\tau^{t,b(t)}} \leq \tilde{X}_s^{\tau,b(\tau)}$ a.s. for every $s \in [\tau,T]$. Therefore the single crossing condition implies

$$\begin{aligned}
\mathbb{E}\left[\int_t^\tau h(s, X_s^{t,x})ds + \pi(\tau)\right] &= \mathbb{E}\left[\int_t^\tau h(s, X_s^{t,x})ds + \int_\tau^T h(s, \tilde{X}_s^{\tau,b(\tau)})ds\right] \\
&\leq \mathbb{E}\left[\int_t^\tau h(s, \tilde{X}_s^{t,b(t)})ds + \int_\tau^T h(s, \tilde{X}_s^{t,b(t)})ds\right] \\
&= \pi(t).
\end{aligned}$$

This implies $W(t,x,\tau) + \mathbb{E}\left[\pi(\tau)\right] \leq W(t,x,t) + \pi(t)$. Hence $\tau_b^{t,x} = t$ is optimal in (3.3) as claimed.

In the second step fix $x < b(t)$ and let $\tau \in \mathcal{T}_{t,T}$ be an arbitrary stopping time. To shorten notation we write $\tau_b = \tau_b^{t,x}$. First, we prove that the stopping $\min\{\tau, \tau_b\}$ performs at least as well as τ. By (3.11) we have

$$\begin{aligned} \mathbb{E}\left[1_{\{\tau_b<\tau\}}\pi(\tau)\right] &= \mathbb{E}\left[1_{\{\tau_b<\tau\}}\mathbb{E}\left[\int_\tau^T h(s, \tilde{X}_s^{\tau,b(\tau)})ds \,|\, \mathcal{F}_\tau\right]\right] \\ &= \mathbb{E}\left[1_{\{\tau_b<\tau\}}\int_\tau^T h(s, \tilde{X}_s^{\tau,b(\tau)})ds\right]. \end{aligned}$$

This leads to

$$\begin{aligned} &\mathbb{E}\left[1_{\{\tau_b<\tau\}}\left(\int_t^\tau h(s, X_s^{t,x})ds + \pi(\tau)\right)\right] \\ &= \mathbb{E}\left[1_{\{\tau_b<\tau\}}\left(\int_t^{\tau_b} h(s, X_s^{t,x})ds + \int_{\tau_b}^\tau h(s, X_s^{t,x})ds + \int_\tau^T h(s, \tilde{X}_s^{\tau,b(\tau)})ds\right)\right]. \end{aligned}$$

By construction of the reflected process $\tilde{X}$ we have $\tilde{X}_{\tau_b}^{t,x} = b(\tau_b)$ if $\tau_b < T$. The comparison principle between the original and the reflected process (Property 5.) and the flow property of reflected processes (Equation (3.9)) imply almost surely

$$\tilde{X}_s^{\tau_b,b(\tau_b)} = \tilde{X}_s^{\tau_b,\tilde{X}_{\tau_b}^{t,x}} = \tilde{X}_s^{t,x} \leq X_s^{t,x}$$

for $T > s \geq \tau_b$. Since $\tilde{X}_\tau^{\tau_b,b(\tau_b)} \leq b(\tau)$ we have on the set $\{\tau > \tau_b\}$

$$\tilde{X}_s^{\tau,b(\tau)} \geq \tilde{X}_s^{\tau,\tilde{X}_\tau^{\tau_b,b(\tau_b)}} = \tilde{X}_s^{\tau_b,b(\tau_b)}$$

for all $s \geq \tau$. These two inequalities combined with the monotonicity of h yield that

$$\begin{aligned} &\mathbb{E}\left[1_{\{\tau_b<\tau\}}\left(\int_t^\tau h(s, X_s^{t,x})ds + \pi(\tau)\right)\right] \\ &\leq \mathbb{E}\left[1_{\{\tau_b<\tau\}}\left(\int_t^{\tau_b} h(s, X_s^{t,x})ds + \int_{\tau_b}^\tau h(s, \tilde{X}_s^{\tau_b,b(\tau_b)})ds\right) + \int_\tau^T h(s, \tilde{X}_s^{\tau_b,b(\tau_b)})ds\right] \\ &= \mathbb{E}\left[1_{\{\tau_b<\tau\}}\left(\int_t^{\tau_b} h(s, X_s^{t,x})ds + \pi(\tau_b)\right)\right]. \end{aligned}$$

Consequently using the stopping time $\min\{\tau, \tau_b\}$ is at least as good as using τ

$$\begin{aligned} W(t,x,\tau) + \mathbb{E}\left[\pi(\tau)\right] &= g(t,x) + \mathbb{E}\left[\int_t^\tau h(s, X_s^{t,x})ds + \pi(\tau)\right] \\ &\leq g(t,x) + \mathbb{E}\left[\int_t^{\tau\wedge\tau_b} h(s, X_s^{t,x})ds + \pi(\min\{\tau,\tau_b\})\right] \\ &= W(t,x,\min\{\tau,\tau_b\}) + \mathbb{E}\left[\pi(\min\{\tau,\tau_b\})\right]. \end{aligned}$$

Thus it suffices to consider stopping rules $\tau \leq \tau_b$. In this case we have

$$\begin{aligned}&\mathbb{E}\left[\int_t^\tau h(s, X_s^{t,x})ds + \pi(\tau)\right]\\&= \mathbb{E}\left[\int_t^\tau h(s, X_s^{t,x})ds + \int_\tau^{\tau_b} h(s, \tilde{X}_s^{\tau,b(\tau)})ds + \int_{\tau_b}^T h(s, \tilde{X}_s^{\tau,b(\tau)})ds\right].\end{aligned}$$

From the comparison principle for reflected processes (Property 3.) and the flow property Equation (3.9) follows $\tilde{X}_s^{t,x} = \tilde{X}_s^{\tau,\tilde{X}_\tau^{t,x}} \leq \tilde{X}_s^{\tau,b(\tau)}$ for all $s \geq \tau$. By the minimality property of reflected processes (Property 2.) we have that $X_s^{t,x} = \tilde{X}_s^{t,x}$ for all $s < \tau_b$. Similar considerations as above yield

$$\tilde{X}_s^{\tau_b,b(\tau_b)} = \tilde{X}_s^{\tau_b,\tilde{X}_{\tau_b}^{t,x}} = \tilde{X}_s^{t,x} = \tilde{X}_s^{\tau,\tilde{X}_\tau^{t,x}} \leq \tilde{X}_s^{\tau,b(\tau)}$$

a.s. for $s \geq \tau_b$. The monotonicity of h implies

$$\begin{aligned}&\mathbb{E}\left[\int_t^\tau h(s, X_s^{t,x})ds + \pi(\tau)\right]\\ \leq\ &\mathbb{E}\left[\int_t^\tau h(s, X_s^{t,x})ds + \int_\tau^{\tau_b} h(s, X_s^{t,x})ds + \int_{\tau_b}^T h(s, \tilde{X}_s^{\tau_b,b(\tau_b)})ds\right]\\ =\ &\mathbb{E}\left[\int_t^{\tau_b} h(s, X_s^{t,x})ds + \pi(\tau_b)\right]\end{aligned}$$

and hence $W(t, x, \tau) + \mathbb{E}[\pi(\tau)] \leq W(t, x, \tau_b) + \mathbb{E}[\pi(\tau_b)]$. This completes the proof of implementability. □

Example 3.3.6. *Assume that $X = \sigma W$ is a Brownian motion with volatility $\sigma > 0$. Further suppose that there is no flow payoff $f = 0$ but only a final payoff of the form $g(t, x) = x^2$. Then the single crossing condition is satisfied and the transfer from Theorem 3.3.5 does not depend on the cut-off b:*

$$\pi(t) = \sigma^2(T - t).$$

Indeed, $-\pi$ is (up to a constant) the increasing part of the Doob-Meyer decomposition of the submartingale $g(\cdot, X)$. Hence, the process $g(\cdot, X) + \pi$ is a martingale and therefore every region A is implemented by π. The stopping region (and thus the only strictly implementable region) of the stopping problem with payoff $g + \pi$ is the whole state space $D_t^\pi = \mathbb{R}$.

In Proposition 3.2.2 we showed that strictly implementable regions are necessarily of cut-off type. The next result establishes the converse direction. Under the strict single crossing condition cut-off regions are strictly implementable.

Theorem 3.3.7. *If the strict single crossing condition holds true, then a regular cut-off region with cut-off b is strictly implemented by the transfer from Equation (3.10).*

Proof. We use the same notation as in the proof of Theorem 3.3.5. Let $t \in [0,T]$ and $x < b(t)$. Then the right-continuity of b and $\tilde{X}$ and the strict monotonicity of h imply that

$$\mathbb{E}\left[\int_t^{\tau_b} h(s, \tilde{X}_s^{t,x})ds\right] > \mathbb{E}\left[\int_t^{\tau_b} h(s, \tilde{X}_s^{t,b(t)})ds\right]$$

Consequently we have

$$\begin{aligned} \mathbb{E}\left[\int_t^{\tau_b} h(s, X_s^{t,x})ds + \pi(\tau_b)\right] &= \mathbb{E}\left[\int_t^{\tau_b} h(s, \tilde{X}_s^{t,x})ds + \int_{\tau_b}^T h(s, \tilde{X}_s^{\tau_b,b(\tau_b)})ds\right] \\ &> \mathbb{E}\left[\int_t^{\tau_b} h(s, \tilde{X}_s^{t,b(t)})ds + \int_{\tau_b}^T h(s, \tilde{X}_s^{t,b(t)})ds\right] \\ &= \pi(t). \end{aligned}$$

This implies $v^\pi(t,x) > \pi(t) + g(t,x)$ and hence A is strictly implemented by π. □

In general the distribution of the reflected process $\tilde{X}$ is not explicitly known. Hence, one has to fall back to numerical methods to approximate the transfer from Theorem 3.3.5. For example one could use discretization schemes for the RSDE (3.4) and Monte Carlo simulations to evaluate the expectation in Equation (3.10) (see e.g. [73], [15] or [61]). If X evolves according to a Brownian motion, then the distribution of $\tilde{X}$ is available in closed form.

Example 3.3.8. *First assume that X evolves according to a Brownian motion with volatility $\sigma > 0$ and drift $\mu \in \mathbb{R}$*

$$dX_t = \mu dt + \sigma dW_t.$$

For regular cut-offs b the reflected version $(\tilde{X}, l)$ of X is given by

$$\begin{aligned} \tilde{X}_s^{t,x} &= X_s^{t,x} - \sup_{t \le r \le s} \left(X_r^{t,x} - b(r)\right)^+ \\ l_s &= \sup_{t \le r \le s} \left(X_r^{t,x} - b(r)\right)^+ \end{aligned}$$

If $b(t) = b$ is constant and X has vanishing drift ($\mu = 0$) we have

$$\tilde{X}_s^{t,b(t)} = b + \sigma(W_s - W_t) - \sigma \sup_{t \le r \le s} (W_r - W_t).$$

It follows from the reflection principle for the Brownian motion (see e.g. [44, Chapter 2 Section 8A]) that

$$\tilde{X}_s^{t,b(t)} \sim b - \sigma |W_{s-t}|.$$

This leads to the following representation of the transfer from Theorem 3.3.5

$$\pi(t) = \int_t^T \int_0^\infty \sqrt{\frac{2}{\pi(s-t)}} e^{-\frac{x^2}{2(s-t)}} h(s, b - \sigma x) dx ds$$

where $h = f + (\partial_t + \mathcal{L})g$.

3.3.3. Properties of the transfer

The next proposition summarizes properties of transfer implementing a cut-off region.

Proposition 3.3.9. *Let $b : [0,T] \to \mathbb{R}$ be a regular cut-off. The transfer π from Equation (3.10) satisfies the following properties*

1. *π is càdlàg. In particular π is bounded and measurable.*
2. *π is continuous at $t \in [0,T]$ if b is continuous at t or if b has a downward jump at t.*
3. *π has no upward jumps.*
4. *If π has a downward jump at $t \in [0,T]$, then b has an upward jump at t.*
5. *π converges to 0 at time T: $\lim_{t\nearrow T} \pi(t) = 0$.*

Proof. As in the proof of Theorem 3.3.5 we introduce the function $h(t,x) = f(t,x) + (\partial_t + \mathcal{L})\, g(t,x)$. By assumption h is Lipschitz continuous and has linear growth in x. The transfer π is given by

$$\pi(t) = \mathbb{E}\left[\int_t^T h(s, \tilde{X}_s^{t,b(t)})ds\right].$$

We first show that π is right-continuous. For $t \in [0,T]$ and $\epsilon > 0$ we have

$$\begin{aligned} |\pi(t) - \pi(t+\epsilon)| \;\leq\; & \mathbb{E}\left[\int_t^{t+\epsilon} \left|h(s, \tilde{X}_s^{t,b(t)})\right| ds\right] \\ & + \mathbb{E}\left[\int_{t+\epsilon}^T \left|h(s, \tilde{X}_s^{t,b(t)}) - h(s, \tilde{X}_s^{t+\epsilon,b(t+\epsilon)})\right| ds\right]. \end{aligned}$$

It follows from the linear growth of h and Property 1. of $\tilde{X}$ from Proposition 3.3.3 that $\mathbb{E}\left[\int_t^{t+\epsilon} \left|h(s, \tilde{X}_s^{t,b(t)})\right| ds\right] \to 0$ as $\epsilon \to 0$. Moreover, the Lipschitz continuity of h implies

$$\mathbb{E}\left[\int_{t+\epsilon}^T \left|h(s, \tilde{X}_s^{t,b(t)}) - h(s, \tilde{X}_s^{t+\epsilon,b(t+\epsilon)})\right| ds\right] \leq C\mathbb{E}\left[\sup_{s\in[t+\epsilon,T]} \left|\tilde{X}_s^{t,b(t)} - \tilde{X}_s^{t+\epsilon,b(t+\epsilon)}\right|\right]$$

for some constant $C > 0$. By the flow property (Equation (3.9)) we have $\tilde{X}_s^{t,b(t)} = \tilde{X}_s^{t+\epsilon, \tilde{X}_{t+\epsilon}^{t,b(t)}}$. Property 4. from Proposition 3.3.3 yields

$$\mathbb{E}\left[\sup_{s\in[t+\epsilon,T]} \left|\tilde{X}_s^{t,b(t)} - \tilde{X}_s^{t+\epsilon,b(t+\epsilon)}\right|\right] \leq \tilde{C}\mathbb{E}\left[\left|\tilde{X}_{t+\epsilon}^{t,b(t)} - b(t+\epsilon)\right|\right].$$

Right continuity of $\tilde{X}$ and b then implies $\pi(t+) = \pi(t)$.[2]

[2]Here and in the sequel we use the notation $\pi(t+) = \lim_{\epsilon\searrow 0} \pi(t+\epsilon)$ and $\pi(t-) = \lim_{\epsilon\searrow 0} \pi(t-\epsilon)$ for the one-sided limits.

Concerning the left-hand limits of π we show that

$$\pi(t-) = \mathbb{E}\left[\int_t^T h(s, \tilde{X}_s^{t,b(t)\wedge b(t-)})ds\right]. \tag{3.12}$$

for all $t \in (0, T]$. Equation (3.12) implies all remaining claims of Proposition 3.3.9. If b is continuous at t or has a downward jump ($b(t) \leq b(t-)$), then Equation (3.12) yields continuity of π at t: $\pi(t-) = \pi(t)$. Monotonicity of h and the comparison principle for the reflected process imply $\pi(t-) \geq \pi(t)$, i.e. π has no upward jumps. If π has a downward jump at time t ($\pi(t-) > \pi(t)$), then Equation (3.12) yields that b has necessarily an upward jump ($b(t) > b(t-)$). Moreover, it follows from Equation (3.12) that $\pi(T-) = 0$. To prove Equation (3.12) let $t \in (0, T]$ and $\epsilon > 0$. Then consider

$$\begin{aligned}
&\left|\pi(t-\epsilon) - \mathbb{E}\left[\int_t^T h(s, \tilde{X}_s^{t,b(t)\wedge b(t-)})ds\right]\right| \\
&\leq \mathbb{E}\left[\int_{t-\epsilon}^t \left|h(s, \tilde{X}_s^{t-\epsilon,b(t-\epsilon)})\right| ds\right] + \mathbb{E}\left[\int_t^T \left|h(s, \tilde{X}_s^{t-\epsilon,b(t-\epsilon)}) - h(s, \tilde{X}_s^{t,b(t)\wedge b(t-)})\right| ds\right].
\end{aligned}$$

Property 6. from Proposition 3.3.3 yields $\tilde{X}_s^{t-\epsilon,b(t-\epsilon)} \to \tilde{X}_s^{t,b(t)\wedge b(t-)}$ in L^2 if $\epsilon \searrow 0$. Lipschitz continuity and linear growth of h then imply that

$$\mathbb{E}\left[\int_{t-\epsilon}^t \left|h(s, \tilde{X}_s^{t-\epsilon,b(t-\epsilon)})\right| ds\right] \to 0$$

and

$$\mathbb{E}\left[\int_t^T \left|h(s, \tilde{X}_s^{t-\epsilon,b(t-\epsilon)}) - h(s, \tilde{X}_s^{t,b(t)\wedge b(t-)})\right| ds\right] \to 0$$

for $\epsilon \searrow 0$. This yields the claim. □

3.3.4. Uniqueness of the transfer

To prove a uniqueness result for the transfer from Theorem 3.3.5 we need the following auxiliary result about cut-off stopping times.

Lemma 3.3.10. *Let $b : [0, T] \to \mathbb{R}$ be bounded from below. Then we have $\tau_b^{t,x} \nearrow T$ a.s. for $x \searrow -\infty$ and for every $t \in [0, T]$.*

Proof. Fix $t \in [0, T]$. By [51, Lemma 3.7] there exists a constant $C > 0$ such that

$$\mathbb{E}\left[\sup_{t\leq s\leq T}\left(\frac{1}{1 + (X_s^{t,x})^2}\right)^2\right] \leq C\left(\frac{1}{1 + x^2}\right)^2.$$

Then Fatou's lemma implies

$$\begin{aligned}\mathbb{E}\left[\liminf_{x\to-\infty}\sup_{t\le s\le T}\left(\frac{1}{1+(X_s^{t,x})^2}\right)^2\right] &\le \liminf_{x\to-\infty}\mathbb{E}\left[\sup_{t\le s\le T}\left(\frac{1}{1+(X_s^{t,x})^2}\right)^2\right]\\ &\le \liminf_{x\to-\infty} C\left(\frac{1}{1+x^2}\right)^2\\ &= 0.\end{aligned}$$

Consequently we have $\limsup_{x\to-\infty}\inf_{t\le s\le T}|X_s^{t,x}| = \infty$ a.s. Together with the comparison principle for X this yields $\limsup_{x\to-\infty}\sup_{t\le s\le T} X_s^{t,x} = -\infty$ a.s. It follows that $\tau_b^{t,x} \nearrow T$ for $x \searrow -\infty$. □

Theorem 3.3.11. *Let A be a regular cut-off region with cut-off b. Assume that A is implemented by two transfers π and $\hat\pi$ satisfying $\lim_{t\nearrow T}\pi(t) = \lim_{t\nearrow T}\hat\pi(t)$. Then $\pi(t) = \hat\pi(t)$ for all $t \in [0,T)$.*

Proof. Fix $t \in [0,T)$. To shorten notation we set $v = v^\pi$ and $\hat v = v^{\hat\pi}$. By Lemma 3.2.1 the functions v and $\hat v$ are Lipschitz continuous in the x variable. Similar considerations yield that the function $x \mapsto W(t,x,\tau)$ is Lipschitz continuous for every $\tau \in \mathcal{T}_{t,T}$. In particular, these functions are absolutely continuous. Appealing to the envelope theorem from [58, Theorem 1] yields that

$$v_x(t,x) = W_x(t,x,\tau_b^{t,x}) = \hat v_x(t,x)$$

for Lebesgue almost every $x \in \mathbb{R}$. Integrating from $x < b(t)$ to $b(t)$ gives

$$v(t,b(t)) - v(t,x) = \hat v(t,b(t)) - \hat v(t,x)$$

or equivalently

$$\pi(t) - \hat\pi(t) = \mathbb{E}\left[\pi(\tau_b^{t,x}) - \hat\pi(\tau_b^{t,x})\right].$$

Since π and $\hat\pi$ are bounded we can appeal to Lemma 3.3.10 to obtain

$$\pi(t) - \hat\pi(t) = \lim_{x\to-\infty}\mathbb{E}\left[\pi(\tau_b^{t,x}) - \hat\pi(\tau_b^{t,x})\right] = 0,$$

where we used the dominated convergence theorem. □

3.4. Application to optimal stopping

From Theorem 3.3.11 we derive a probabilistic characterization of optimal stopping times for stopping problems of the form

$$v(t,x) = \sup_{\tau\in\mathcal{T}_{t,T}}\mathbb{E}\left[\int_t^\tau f(s,X_s^{t,x})ds + g(\tau,X_\tau^{t,x})\right], \tag{3.13}$$

where f, g and X satisfy the single crossing condition. We say that a stopping time $\tau \in \mathcal{T}_{t,T}$ is optimal in (3.13) for $(t,x) \in [0,T] \times \mathbb{R}$ if

$$v(t,x) = \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x})ds + g(\tau, X_\tau^{t,x})\right].$$

Corollary 3.4.1. *Assume that the single crossing condition is satisfied and let $b : [0,T] \to \mathbb{R}$ be a regular cut-off. The stopping time $\tau_b^{t,x}$ is optimal in (3.13) for all $(t,x) \in [0,T] \times \mathbb{R}$, if and only if b satisfies the nonlinear integral equation*

$$\mathbb{E}\left[\int_t^T f(s, \tilde{X}_s^{t,b(t)}) + (\partial_t + \mathcal{L})g(s, \tilde{X}_s^{t,b(t)})ds\right] = 0 \tag{3.14}$$

for all $t \in [0,T]$.

Proof. First assume that (3.14) holds true for every $t \in [0,T]$. Then Theorem 3.3.5 implies that the cut-off region with cut-off b is implemented by the zero transfer. This means that $\tau_b^{t,x}$ is optimal in (3.13) for every $(t,x) \in [0,T] \times \mathbb{R}$.

For the converse direction assume that $\tau_b^{t,x}$ is optimal in (3.13) for every $(t,x) \in [0,T] \times \mathbb{R}$. Then the cut-off region with cut-off b is implemented by the zero transfer $\hat{\pi} = 0$. By Theorem 3.3.5 it is also implemented by the transfer

$$\pi(t) = \mathbb{E}\left[\int_t^T f(s, \tilde{X}_s^{t,b(t)}) + (\partial_t + \mathcal{L})g(s, \tilde{X}_s^{t,b(t)})ds\right].$$

By Proposition 3.3.9 the transfer π satisfies $\lim_{t \nearrow T} \pi(t) = 0$. Then Theorem 3.3.11 implies that $\pi(t) = \hat{\pi}(t) = 0$ for all $t \in [0,T]$. ☐

In the literature on optimal stopping there is a well known link between optimal stopping boundaries and a nonlinear integral equation differing from Equation (3.14). It was established by [45], [41] and [17] who considered the optimal exercise of an American option. For a more general framework and an overview we refer the reader to [67]. In [67, Chapter IV, Section 14] the authors derive the integral equation

$$\mathbb{E}\left[\int_t^T \left(f(s, X_s^{t,b(t)}) + (\partial_t + \mathcal{L})g(s, X_s^{t,b(t)})\right) 1_{\{X_s^{t,b(t)} \leq b(s)\}} ds\right] = 0 \tag{3.15}$$

as a *necessary* optimality condition for a stopping boundary b. To verify that a solution to Equation (3.15) indeed yields an optimal stopping time the authors proceed as follows. First, it is shown that an optimal stopping time τ_b exists and that the associated boundary b necessarily satisfies Equation (3.15). In a second step the authors verify that Equation (3.15) has at most one solution which consequently has to coincide with b. This verification step, however, is carried out only for specific examples. These examples include the cases of American ([65]) or Russian ([66]) option payoffs where the price process X evolves according to a geometric Brownian motion. For general diffusion processes and payoffs Equation (3.15) does not provide a *sufficient* condition for optimality. This is illustrated by the next simple example.

Example 3.4.2. *Consider the static case, i.e. assume that* $\mu = \sigma = 0$ *and hence* $X_s^{t,x} = x$ *for all* $s \geq t$ *and* $x \in \mathbb{R}$. *Then every strictly decreasing function satisfies Equation (3.15) since* $1_{\{X_s^{t,b(t)} \leq b(s)\}} = 0$ *for all* $s > t$. *But clearly not every strictly decreasing boundary is optimal in (3.13) for an arbitrary choice of* f *and* g. *Choose for example* $f(t,x) = -x$ *and* $g(t,x) = 0$, *where the unique optimal stopping boundary is given by* $b(t) = 0$.

By contrast Equation (3.14) leads to the following characterization of the optimal stopping boundary under the single crossing condition. Since $\tilde{X}_s^{t,b(t)} = b(s)$ *for all* $s \in [t,T]$, *Equation (3.14) is satisfied by a function* b *if and only if* $f(t,b(t)) + g_t(t,b(t)) = 0$ *for all* $t \in [0,T]$. *In the case* $f(t,x) = -x$ *and* $g(t,x) = 0$, *this condition indeed yields the stopping boundary* $b(t) = 0$.

3.5. Appendix

Proof of Proposition 3.3.3. Existence and uniqueness of $(\tilde{X}, l)$ follow from [72]. See also [79, Theorem 3.4] for the time-homogeneous case. By construction of $(\tilde{X}, l)$ we also have 1.

We next show 2. Note that the solution to the unreflected SDE (3.1) solves the reflected SDE for $s < \tau_b$. As the solution to the reflected SDE is unique 2. follows.

To prove 3. and 4. we consider without loss of generality only the case $t = 0$. For $\xi_1, \xi_2 \in \mathbb{R}$ we write $(\tilde{X}^i, l^i) = (\tilde{X}^{0,\xi_i}, l^{0,\xi_i})$, $(i = 1,2)$ and introduce the processes $D_t = \tilde{X}_t^1 - \tilde{X}_t^2$ and $\Gamma_t = \sup_{s \leq t} \max(0, D_s)^2$. Applying the Meyer-Itô formula [71, Theorem 71, Chapter 4] to the function $x \mapsto \max(0,x)^2$ yields

$$\begin{aligned}\max(0,D_s)^2 =& \max(0,D_0)^2 + 2\int_0^s 1_{\{D_{r-}>0\}} D_{r-} dD_r + \int_0^s 1_{\{D_{r-}>0\}} d\,[D]_r^c \\ &+ \sum_{0<r\leq s} \left(\max(0,D_r)^2 - \max(0,D_{r-})^2 - 1_{\{D_{r-}>0\}} D_{r-}\Delta D_r\right).\end{aligned} \tag{3.16}$$

Since D only jumps when b has a downward jump and since $\tilde{X}^i$ jumps to the barrier we have $-(\Delta b(r))^- \leq \Delta D_r \leq 0$ on the set $\{D_{r-} > 0\}$. Moreover, D has bounded paths. Since b has summable downward jumps this implies $\sum_{0<r\leq s} 1_{\{D_{r-}>0\}} |D_{r-}\Delta D_r| < \infty$ a.s. Hence, we can rewrite Equation (3.16) as follows

$$\begin{aligned}\max(0,D_s)^2 =& \max(0,D_0)^2 + 2\int_0^s 1_{\{D_r>0\}} D_r dD_r^c + \int_0^s 1_{\{D_{r-}>0\}} d\,[D]_r^c \\ &+ \sum_{0<r\leq s} \left(\max(0,D_r)^2 - \max(0,D_{r-})^2\right).\end{aligned} \tag{3.17}$$

Regarding the jump terms in Equation (3.17), assume that there exists $r \in (0,s]$ such that $\max(0,D_r)^2 > \max(0,D_{r-})^2$. This implies $D_r > 0$ and $D_r > D_{r-}$. Since $\tilde{X}^i$ jumps if and only if l^i jumps $(i = 1,2)$ we obtain $\tilde{X}_r^1 > \tilde{X}_r^2$ and $l_r^2 - l_{r-}^2 > l_r^1 - l_{r-}^1$. It follows that $l_r^2 - l_{r-}^2 > 0$, since l^1 is nondecreasing. Hence, l^2 jumps at r, which implies that

$\tilde{X}^2_r = b(r)$. Thus, we obtain the contradiction $\tilde{X}^1_r > b(r)$. Therefore we have

$$\sum_{0<r\leq s} \left(\max(0, D_r)^2 - \max(0, D_{r-})^2\right) \leq 0.$$

For the last integral in Equation (3.17) the Lipschitz continuity of σ implies

$$\begin{aligned}\int_0^s 1_{\{D_r>0\}} d\langle D\rangle^c_r = \int_0^s 1_{\{D_r>0\}}(\sigma(r,\tilde{X}^1_r) - \sigma(r,\tilde{X}^2_r))^2 dr &\leq L^2 \int_0^s \max(0,D_r)^2 dr \\ &\leq L^2 \int_0^s \Gamma_r dr.\end{aligned}$$

The first integral of Equation (3.17) decomposes into the following terms, which we will consider successively. By the Lipschitz continuity of μ we have

$$2\int_0^s 1_{\{D_r>0\}} D_r(\mu(r,\tilde{X}^1_r) - \mu(r,\tilde{X}^2_r))dr \leq 2L\int_0^s \max(0,D_r)^2 dr \leq L\int_0^s \Gamma_r dr.$$

Next, we have

$$-2\int_0^s 1_{\{D_r>0\}} D_r dl^{1,c}_r \leq 0$$

and

$$2\int_0^s 1_{\{D_r>0\}} D_r dl^{2,c}_r = 2\int_0^s 1_{\{\tilde{X}^1_r > b(r)\}} D_r dl^{2,c}_r = 0.$$

Moreover, it follows from the Burkholder-Davis-Gundy inequality, the Lipschitz continuity of σ and Young's inequality that

$$\begin{aligned}\mathbb{E}\left[\sup_{s\leq t}\int_0^s 1_{\{D_r>0\}} D_r(\sigma(r,\tilde{X}^1_r) - \sigma(r,\tilde{X}^2_r))dW_r\right] &\leq C\mathbb{E}\left[\sqrt{\int_0^t 1_{\{D_r>0\}} D^4_r dr}\right] \\ &\leq C\mathbb{E}\left[\sqrt{\Gamma_t \int_0^t \Gamma_r dr}\right] \\ &\leq \frac{1}{2}\mathbb{E}\left[\Gamma_t\right] + \frac{1}{2}C^2\mathbb{E}\left[\int_0^t \Gamma_r dr\right].\end{aligned}$$

Putting everything together, we obtain

$$\mathbb{E}\left[\Gamma_t\right] \leq \Gamma_0 + K\int_0^t \mathbb{E}\left[\Gamma_r\right] dr$$

for some constant $K > 0$. Then Gronwall's lemma yields

$$\mathbb{E}\left[\sup_{s\leq t}\max(0,\tilde{X}^1_s - \tilde{X}^2_s)^2\right] = \mathbb{E}[\Gamma_t] \leq C\Gamma_0 = C\max(0,\xi_1 - \xi_2)^2 \tag{3.18}$$

for some constant $C > 0$. If $\xi_1 \le \xi_2$ this directly yields 3. For 4. observe that we have

$$\mathbb{E}\left[\sup_{s\le t}(\tilde X^1_s - \tilde X^2_s)^2\right] \le \mathbb{E}\left[\sup_{s\le t}\max(0, \tilde X^1_s - \tilde X^2_s)^2\right] + \mathbb{E}\left[\sup_{s\le t}\max(0, \tilde X^2_s - \tilde X^1_s)^2\right].$$

Then Inequality (3.18) yields $\mathbb{E}\left[\sup_{s\le t}(\tilde X^1_s - \tilde X^2_s)^2\right] \le \tilde C(\xi_1 - \xi_2)^2$. The case $p = 1$ follows from Jensen's inequality. Claim 5. follows by performing similar arguments with $D = \tilde X^{t,\xi} - X^{t,\xi}$.

In order to prove Equation (3.7), we set

$$Y_s = Y^{t,\xi}_s = \int_t^s \mu(u, \tilde X^{t,\xi}_u)du + \int_t^r \sigma(u, \tilde X^{t,\xi}_{u-})dW_u, \quad \hat l_s = \sup_{t\le r\le s}(\xi + Y_r - b(r))^+ \tag{3.19}$$

and $\hat X = \xi + Y_s - \hat l_s$. Then it is straightforward to show that $(\hat X, \hat l)$ is a solution to the Skorokhod problem associated with Y and barrier b (cf. [79, Definition 2.5]). Since $(\tilde X, l)$ is also a solution, we obtain Equation (3.7) by uniqueness of solutions to the Skorokhod problem (cf. [79, Proposition 2.4]).

Finally we prove Claim 6. To this end let $x \le b(t) \wedge b(t-)$ and $t_n \nearrow t$ and $x_n \to x$ as $n \to \infty$. We write $\tilde X^n = \tilde X^{t_n, x_n \wedge b(t_n)}$ and $Y^n = Y^{t_n, x_n\wedge b(t_n)}$ (see Equation (3.19) for the definition of Y). Then we have

$$\begin{aligned}|\tilde X^n_t - x| &= |x_n \wedge b(t_n) - x + Y^n_t - \sup_{t_n\le r\le t}(x_n\wedge b(t_n) + Y^n_r - b(r))^+| \\ &\le |x_n\wedge b(t_n) - x| + \sup_{t_n\le r\le t}(x_n \wedge b(t_n) - b(r))^+ + 2\sup_{t_n\le r\le t}|Y^n_r|.\end{aligned}$$

Squaring this inequality and taking expectations yields

$$\mathbb{E}\left[|\tilde X^n_t - x|^2\right] \le 3|x_n\wedge b(t_n) - x|^2 + 3\sup_{t_n\le r\le t}\left((x_n\wedge b(t_n) - b(r))^+\right)^2 + 6\mathbb{E}\left[\sup_{t_n\le r\le t}|Y^n_r|^2\right].$$

The first two terms converge to 0 for $n \to \infty$ since b is càdlàg and $x \le b(t)\wedge b(t-)$. Regarding the last term, observe that Jensen's and the Burkholder-Davis-Gundy inequality yields

$$\mathbb{E}\left[\sup_{t_n\le r\le t}|Y^n_r|^2\right] \le C\int_{t_n}^t \mathbb{E}\left[\mu(s, \tilde X^n_s)^2 + \sigma(s, \tilde X^n_s)^2\right]ds$$

for some constant $C > 0$ (not depending on n). It remains to prove that the sequence $\mathbb{E}\left[\mu(s, \tilde X^n_s)^2 + \sigma(s, \tilde X^n_s)^2\right]$ is bounded. To this end assume without loss of generality that $\tilde X^0 = \tilde X^{0,b(0)}$, then the linear growth of μ and σ, the Markov property of $\tilde X^0$ and Claim 4 imply

$$\begin{aligned}\mathbb{E}\left[\mu(s, \tilde X^n_s)^2 + \sigma(s, \tilde X^n_s)^2\right] &\le C_1\left(1 + \mathbb{E}\left[(\tilde X^n_s - \tilde X^0_s)^2 + (\tilde X^0_s)^2\right]\right) \\ &\le C_2\left(1 + \mathbb{E}\left[(\tilde X^0_{t_n} - x_n\wedge b(t_n))^2 + (\tilde X^0_s)^2\right]\right)\end{aligned}$$

for some $C_1, C_2 > 0$. This is a bounded sequence by Claim 1. which yields

$$\mathbb{E}[\sup_{t_n \le r \le t} |Y_r^n|^2] \to 0$$

as $n \to \infty$. □

Proof of Lemma 3.3.4. Fix a stopping time $\tau \in \mathcal{T}$ and introduce the σ-field $\mathcal{G}^\tau = \sigma\{W_{(\tau+u)\wedge T} - W_\tau | u \ge 0\}$, which is independent of $\mathcal{F}_\tau$. Denote by $\tilde{X}^{\tau,x}$ the unique solution to the RSDE (3.4) starting at time τ in $x \in \mathbb{R}$. It follows from the representations (3.4) and (3.7) that $\tilde{X}^{\tau,x}_{(\tau+u)\wedge T}$ is $\mathcal{G}^\tau$-measurable for every $u \ge 0$. In [72] the solution $\tilde{X}^{t,x}$ of the RSDE (3.4) is constructed as the limit of the usual Picard iteration on the space of adapted, càdlàg processes Y with norm $\mathbb{E}[\sup_{0\le s\le T} Y_s^2]$. As in the proof of [71, Theorem 32, Chapter V] (see also [71, Theorem 62, Chapter IV]) it follows that the mapping $(x, u, \omega) \mapsto \tilde{X}^{\tau,x}_{(\tau+u)\wedge T}$ is $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}([0,T]) \otimes \mathcal{F}$-measurable and that by uniqueness of solutions to RSDEs the flow property $\tilde{X}^{0,x}_{(\tau+u)\wedge T} = \tilde{X}^{\tau,\tilde{X}^{0,x}_\tau}_{(\tau+u)\wedge T}$ holds true.

Let $\varphi : [0,T] \times \mathbb{R} \to \mathbb{R}$ be a bounded function. Using the independence of $\mathcal{F}_\tau$ and $\mathcal{G}^\tau$ we have

$$\mathbb{E}[\varphi((\tau+u)\wedge T, \tilde{X}^{0,x}_{(\tau+u)\wedge T})|\mathcal{F}_\tau] = \mathbb{E}[\varphi((\tau+u)\wedge T, \tilde{X}^{\tau,\tilde{X}^{0,x}_\tau}_{(\tau+u)\wedge T})|\mathcal{F}_\tau] = \psi(\tau, \tilde{X}^{0,x}_\tau)$$

with $\psi(t,y) = \mathbb{E}[\varphi((t+u)\wedge T, \tilde{X}^{t,y}_{(t+u)\wedge T})]$. This establishes the claim. □

Part II.

Optimal Closure of Illiquid Positions

4. Price-sensitive liquidation

When closing a large asset position agents often set a minimum goal for the average price they want to get for any unit sold. The aim of this chapter is to provide trading strategies that make it very unlikely to fall extremely below such a target. In Section 4.1 we set up a discrete-time liquidation model that allows for price-sensitive risk preferences of agents. An additive risk functional is introduced that can be interpreted as the time average of the squared value-at-risk of the open position. The value-at-risk of the open position is assumed to depend on the current asset price level. The value function of the control problem is a quadratic form in the remaining position size. The coefficient functions are determined by a backward function recursion. Optimal trading strategies are characterized by means of these coefficient functions and it is shown that at any time, the optimal amount to trade is proportional to the remaining position size.

The aim of Section 4.2 is to derive optimal liquidation strategies in the continuous-time version of the model presented in Section 4.1. The linear price impact entails that the optimal change in the position size is absolutely continuous with respect to the Lebesgue measure. In particular, there are no jumps in the position size as time evolves. Thus, liquidation strategies are determined by trading rates, i.e. the amount of shares sold in an infinitesimal amount of time. It turns out that optimal trading rates can be characterized in terms of a PDE describing how much they differ from the optimal trading rates of a risk-neutral agent. The PDE is nonstandard as it possesses a singularity at the end of the liquidation period T. The singularity arises from the terminal state constraint that the remaining position has to be zero at time T. A PDE characterization of optimal trading strategies is provided in the viscosity sense. Moreover it is shown that the optimal strategies from the discrete-time model from Section 4.1 converge to the continuous-time optimal trading rates.

In general the optimal trading strategies from Sections 4.1 and 4.2 do not admit a closed form representation. Section 4.3 provides an algorithm of how to numerically calculate the function coefficients derived in Section 4.1. As a case study the liquidation of forward positions in illiquid energy markets is considered (Section 4.4). In numerical experiments, the algorithm proves to be very fast, allowing to calculate optimal trading strategies and statistical properties of the trading performance within a few seconds. This simulation analysis shows that the model can incorporate skewness preferences of agents. The more sensitive the risk measure responds to price changes, the more skewed the realized proceeds are.

4.1. Price-sensitive liquidation in discrete-time

In this section we set up a discrete-time liquidation model that allows for price-sensitive risk preferences of agents. The portfolio to be liquidated consists of long and short positions in d different assets. We allow for an absolute as well as for a relative linear, temporary price impact. In what concerns the asset price dynamics, the martingale and the Markov property are assumed, but there is no explicit assumption on the price distribution.

The optimization problem of how to optimally liquidate the portfolio is then solved by means of discrete dynamic programming. The formulas for the value function and the optimal trade execution are semi-analytic. The value function is a quadratic form of the remaining position. The coefficients of the quadratic form are functions of the price, and can be characterized in terms of a *function recursion*, but in general not in closed form.

4.1.1. Model description

Consider an agent wanting to unwind a position $x = (x^1, \dots, x^d)$ of $d \in \mathbb{N}$ assets up to some time horizon T. Assume that the agent can split the asset position into several pieces and close them consecutively in $N \in \mathbb{N}$ trading periods, e.g. trading hours. We denote by $0 = t_0 < \ldots < t_N = T$, the beginning of each trading period. We quote prices as forward prices and denote the no-impact price of the assets at time t_k by the vector $S_k = (S_k^1, \dots, S_k^d)^T \in \mathbb{R}^d_{>0}$.

Asset price dynamics

We assume that every asset price (S_k^i) is a martingale and follows geometric time-homogeneous Markovian dynamics. More precisely, let $(I_k)_{1 \le k \le N}$ be an iid collection of square-integrable d-dimensional vectors of positive random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For all $1 \le k \le N$ and $1 \le i \le d$ we assume that $\mathbb{E}[I_k^i] = 1$. The price process (S_k) satisfies

$$S_k^i = S_{k-1}^i I_k^i.$$

Note that $\mathbb{E}[|S_k|^2] < \infty$, where $|\cdot|$ denotes the Euclidian norm in $\mathbb{R}^d$. Throughout let $(\mathcal{F}_k)$ be the filtration generated by (S_k).

Execution strategies

We denote the remaining position at time t_k by $x_k \in \mathbb{R}^d$. Besides we refer to $z_k \in \mathbb{R}^d$ as the amount sold resp. bought between t_k and t_{k+1}. In other words, z_k is the change in the portfolio's position between t_k and t_{k+1}, and it must hold

$$x_{k+1} = x_k - z_k.$$

In the following we will sometimes refer to (x_k) as the *position trajectory.*

By an *execution strategy* of a position $x \in \mathbb{R}^d$ until time t_N we mean an $(\mathcal{F}_k)$-adapted stochastic process $(z_k)_{k=0,\dots N-1}$ with values in $\mathbb{R}^d$ such that

$$\sum_{k=0}^{N-1} z_k = x,$$

or, equivalently, $x_N = 0$. For integrability reasons we always require that $\mathbb{E}[|z_k|^p] < \infty$ for all $p \geq 1$ and $0 \leq k \leq N-1$.

Price impact

Prices at which trades are executed depend on the trade size. Given an execution strategy (z_l), the realized price per unit $\tilde{S}_l$ is assumed to be given by

$$\tilde{S}_l = S_l - \eta(S_l) z_l,$$

where η takes values in the set of diagonal matrices in $\mathbb{R}^{d \times d}$. The term $\eta(S_l)z_l$ describes the *temporary price impact* of the current order. In the sequel we will consider two choices for η. Note that $\mathrm{diag}(v)$ with $v \in \mathbb{R}^d$ denotes the $d \times d$ diagonal matrix with the entries of v on the diagonal.

- *Absolute price impact:* η does not depend on the price $s \in \mathbb{R}^d_{>0}$, i.e. $\eta(s) = \mathrm{diag}(\eta)$ for a vector $\eta \in \mathbb{R}^d_{>0}$.
- *Relative price impact:* For every $s \in \mathbb{R}^d_{>0}$ it holds that $\eta(s) = \mathrm{diag}(\eta_1 s_1, \dots, \eta_d s_d)$ with $\eta \in \mathbb{R}^d_{>0}$.

The relative price impact may be more accurate in the long run. In the short run both impact types essentially coincide since absolute price levels do not change much. When closing position over short time intervals one may therefore use the more convenient absolute impact. In the following we assume either an absolute or a relative impact.

The revenues of following an execution strategy (z_l) amount to

$$R((z_l)) \;=\; \sum_{l=0}^{N-1} \tilde{S}_l^T z_l = \sum_{l=0}^{N-1} S_l^T z_l - \sum_{l=0}^{N-1} z_l^T \eta(S_l) z_l,$$

where we interpret the sum $\sum_{l=0}^{N-1} z_l^T \eta(S_l) z_l$ as the *liquidation costs*.

Risk

We assume that the risk associated to a position trajectory (x_l) is of the form

$$\sum_{l=0}^{N} x_l^T \lambda(S_l) x_l, \tag{4.1}$$

where λ takes values in the set of positive semi-definite matrices in $\mathbb{R}^{d\times d}$. We keep λ general at this stage, but assume that it incorporates the position's correlation structure and a weight function rating the risk in dependence of the price development. We will refer to λ as the *risk function* and give some examples below.

Notice that the positive semi-definiteness of λ implies that the risk is always nonnegative. Moreover, reducing the positions by factor $\frac{1}{2}$ decreases the risk by factor $\frac{1}{4}$.

Objective functional

We suppose that the agent's objective is to maximize, over all execution strategies (z_l) and associated position trajectories (x_l), the expected revenues minus the risk

$$\text{Liquidation problem:} \qquad \mathbb{E}\left[R((z_l)) - \sum_{l=0}^{N} x_l^T \lambda(S_l) x_l\right] \to \max! \tag{4.2}$$

The additive form of the risk functional (4.1) is very convenient for solving the optimization problem of finding the execution strategy maximizing (4.2). Nevertheless it is flexible enough to allow for price-sensitive risk preferences.

Examples for the risk function

Suppose that the agent needs to liquidate a long position of a *single* asset. Possible choices for the risk function λ are

$$\lambda_1(s) = \left(\max(0, c(a-s))\right)^2, \tag{4.3}$$

$$\lambda_2(s) = \max(0, c(a-s)), \tag{4.4}$$

where $a > 0$ is a reference price level and c a price sensitivity. We next give an interpretation of the risk functions λ_1 and λ_2 in terms of a value-at-risk.

Suppose that the agent sets a threshold level $\bar{s} \in \mathbb{R}_{>0}$ as a minimum target price. Given a price $s \in \mathbb{R}_{>0}$ at time t_k and assuming an open long position $x \geq 0$ until time t_{k+1}, the agent interprets $Y_k = x(\bar{s} - sI_{k+1})$ as loss at time t_{k+1}. Suppose the price risk is quantified in terms of the value-at-risk of Y_k at level $\alpha \in (0,1)$. Denote by Q_α, $\alpha \in (0,1)$, the α-quantile. Then

$$Q_\alpha(Y_k) = x\left(sQ_\alpha(-I_{k+1}) + \bar{s}\right) = xc(a-s),$$

with $a = -\frac{\bar{s}}{Q_\alpha(-I_{k+1})}$ and $c = -Q_\alpha(-I_{k+1})$. Considering the positive part only, the squared value-at-risk satisfies

$$\max(0, Q_\alpha(-Y_k))^2 = \lambda_1(s)x^2.$$

Squaring only the position size corresponds to choosing a risk function of the form λ_2.

When closing a *short* position $x < 0$, one can choose risk functions as in (4.3) and (4.4) with a and s interchanged.

We next give an example for λ in the multi-dimensional case. Recall the variance-covariance method for estimating the value-at-risk of a portfolio of $d > 1$ assets (see e.g. [21, Chapter 20]). The value-at-risk of the position x is approximately equal to

$$a\sqrt{(x \cdot s)^T C (x \cdot s)} = a\sqrt{\sum_{i,j} x_i s_i C_{ij} x_j s_j},$$

where C is the covariance matrix of the logreturns, a a quantile of the standard normal distribution and $x \cdot s$ the pointwise product of the vectors x and s.

We draw inspiration from the variance-covariance method and suggest as a possible choice for the risk function $\lambda(s) = \xi(s)C$, where $\xi : \mathbb{R}^d_{>0} \to \mathbb{R}_+$ is scalar function reflecting the price sensitivity. For numerical purposes it turns out to be very convenient to work with a weighted geometric mean G of the price vector s, and to set $\xi(s) = \max(0, G)^2$. The advantage of choosing ξ this way will be explained in more detail in Subsections 4.1.2 and 4.4.2.

Possible extensions

If wanted, one can extend the model by a *permanent price impact* by assuming that the realized price, given an execution strategy (z_l), satisfies

$$\tilde{S}_l = S_l - \eta(S_l)z_l - \sum_{i=0}^{l-1} \gamma(S_i)z_i,$$

where γ is a matrix-valued function of the price. The term $\sum_{i=0}^{l-1} \gamma(S_i)z_i$ represents the permanent price impact accumulated by all transactions up to time t_{l-1}. If either both impacts are absolute or relative, then the liquidation problem can be simplified to a problem with a temporary impact *only*. Indeed, observe that in this case the revenues of following an execution strategy (z_l) amount to

$$\begin{aligned}
R((z_l)) &= \sum_{l=0}^{N-1} \tilde{S}_l^T z_l = \sum_{l=0}^{N-1} S_l^T z_l - \sum_{l=0}^{N-1} z_l^T \eta(S_l) z_l - \sum_{l=0}^{N-1} z_l^T \sum_{i=0}^{l-1} \gamma(S_i) z_i \\
&= \sum_{l=0}^{N-1} S_l^T z_l - \sum_{l=0}^{N-1} z_l^T \eta(S_l) z_l - \sum_{l=0}^{N-1} x_{l+1}^T \gamma(S_l) z_l \\
&= \sum_{l=0}^{N-1} S_l^T z_l - \sum_{l=0}^{N-1} z_l^T \left(\eta(S_l) - \frac{1}{2}\gamma(S_l) \right) z_l - \frac{1}{2} \sum_{l=0}^{N-1} (x_l + x_{l+1})^T \gamma(S_l) z_l.
\end{aligned}$$

The martingale property of (S_l) implies

$$\mathbb{E}\left[\sum_{l=0}^{N-1} (x_l + x_{l+1})^T \gamma(S_l) z_l\right] = \mathbb{E}\left[\sum_{l=0}^{N-1} x_l^T \gamma(S_l) x_l - x_{l+1}^T \gamma(S_l) x_{l+1}\right] = x^T \gamma(S_0) x.$$

Hence, the expected revenues of following (z_l) are given by

$$\mathbb{E}\left[R((z_l))\right] = x^T S_0 - \frac{1}{2}x^T\gamma(S_0)x - \mathbb{E}\left[\sum_{l=0}^{N-1} z_l^T \left(\eta(S_l) - \frac{1}{2}\gamma(S_l)\right) z_l\right]. \tag{4.5}$$

Consequently, an execution strategy maximizing the revenues under a temporary impact η and permanent impact γ also maximizes the revenues if there is only a temporary impact of $\eta - 1/2\gamma$.

4.1.2. Semi-explicit representation of the value function

In this section we calculate the execution strategy maximizing (4.2). By making the objective functional dynamic, we can appeal to discrete dynamic programming, and show that the value function is a quadratic form in the remaining position. The coefficient of the quadratic form is a matrix-valued function of the price vector, and can be derived via a backward recursion. In every step of the recursion one has to calculate a conditional expectation of the subsequent function with respect to the price transition probability.

For any $x \in \mathbb{R}^d$ and $0 \leq k \leq N$ we denote by $\mathcal{A}_k(x)$ the set of execution strategies (z_l), in the sense of Section 4.1.1, such that $\sum_{l=k}^{N-1} z_l = x$. The value function of the liquidation problem, given an open position $x \in \mathbb{R}^d$ and a price $s \in \mathbb{R}^d$ at time t_k, is defined by

$$V_k(s,x) = \sup_{z \in \mathcal{A}_k(x)} \mathbb{E}\left[\sum_{l=k}^{N-1} S_l^T z_l - z_l^T \eta(S_l) z_l - x_l^T \lambda(S_l) x_l \middle| S_k = s, x_k = x\right]. \tag{4.6}$$

To apply the Dynamic Programming Principle we make integrability assumptions that are collected in the next paragraph.

Assumption 4.1.1. For $s \in \mathbb{R}^d_{>0}$ and $k \geq 0$ let μ_s^k denote the distribution of S_k conditional to $S_0 = s$. Moreover, define $\mathcal{L}^1 = \bigcap_{s \in \mathbb{R}^d_{>0}} L^1(\mu_s^1)$, with $L^1(\mu_s^1) = \{f : \mathbb{R}^d_{>0} \to \mathbb{R}^{d \times d} | \int_{\mathbb{R}^d_{>0}} |f_{i,j}| d\mu_s^1 < \infty, \quad \forall i,j \leq d\}$. We assume that

(A1) $\eta, \lambda \in \mathcal{L}^1$;

(A2) the maximum eigenvalue of $\lambda(S_k)$, denoted by $\bar{\lambda}(S_k)$, is square integrable for all $0 \leq k \leq N$;

(A3) the spectral condition of $\eta(S_k)$ satisfies $\mathbb{E}|\kappa(\eta(S_k))|^p < \infty$, for all $p \geq 1$ and $0 \leq k \leq N$. Recall that the spectral condition $\kappa(M)$ of a positive definite matrix M is equal to the ratio of its largest and smallest eigenvalue.

Remark 4.1.2. Note that for $d = 1$ we have $\kappa(\eta(S_k)) = 1$ and hence the integrability assumption (A3) on the spectral condition is trivially satisfied. The same holds true for an absolute price impact, where the spectral condition of the price impact matrix is constant.

Optimal execution strategies are linear in the position

Due to its linear quadratic nature in the x and z variables the value function admits an explicit representation in the x variable. In the price variable it is determined by a backward function recursion.

Proposition 4.1.3. *Assume that (A1)-(A3) hold true. Then the value function is quadratic in the open position. More precisely, there exist functions* $a_k : \mathbb{R}^d_{>0} \to \mathbb{R}^{d\times d}$ *such that the value function is given by*

$$V_k(s,x) = x^T s - x^T a_k(s) x. \tag{4.7}$$

The matrices a_k *belong to* $\mathcal{L}^1$ *and are positive definite for every* $s \in \mathbb{R}^d_{>0}$. *They are determined by the following recursion:*

$$\begin{aligned} a_{N-1}(s) &= \eta(s) + \lambda(s) \\ a_k(s) &= \eta(s)(\eta(s) + Ta_{k+1}(s))^{-1} Ta_{k+1}(s) + \lambda(s) \end{aligned} \tag{4.8}$$

for $0 \le k < N-1$, *where* T *is the operator on* $\mathcal{L}^1$ *defined by* $Tf(s) = \mathbb{E}[f(S_{k+1})|S_k = s]$. *The optimal execution strategy is given by*

$$z_k(s,x) = (\eta(s) + Ta_{k+1}(s))^{-1} Ta_{k+1}(s)x, \qquad 0 \le k \le N-2. \tag{4.9}$$

Observe that the value function (4.7) is the difference of the initial *book value* $x^T s$ and a quadratic form comprising expected liquidation costs and risk. Moreover, the optimal execution strategy (4.9) is linear in x. This means that at any time the amount traded is proportional to the remaining position size. One can interpret the matrix $M_k(s) = (\eta(s) + Ta_{k+1}(s))^{-1} Ta_{k+1}(s)$ as the *selling rate*. In the one-dimensional case the selling rate is equal to the percentage of the open position that is sold resp. bought.

We next consider the limiting cases where either λ or η vanish. Suppose first that $\lambda = 0$. In this case we have $M_k(s) = \operatorname{diag}(\frac{1}{N-k}, \ldots, \frac{1}{N-k})$, which implies that the position is closed linearly. Spreading orders evenly over time minimizes liquidation costs and hence maximizes the objective functional of a risk-neutral agent.

Next suppose that $\lambda(s)$ is positive definite for all $s \in \mathbb{R}^d_{>0}$. If the price impact matrix $\eta(s)$ vanishes, then $M_k(s)$ converges to the identity matrix in $\mathbb{R}^d$. In the limit there are no liquidation costs and hence the position is closed immediately.

In the one-dimensional case and under absolute price impact the selling rate inherits the monotonicity of the risk function. Let us assume that λ is nonincreasing, which is a suitable assumption in the case of closing a long position (c.f. Subsection 4.1.1). Then a straightforward proof by induction shows that the coefficient functions a_k are nonincreasing for all $k \le N-1$. This implies that the selling rate M_k is nonincreasing as well.

The strategy (4.9) can also be characterized as the strategy minimizing the sum of expected liquidation costs and risk. Indeed the martingale property of (S_l) implies that for any strategy $(z_l) \in \mathcal{A}_k(x)$ we have

$$\mathbb{E}\left[\sum_{l=k}^{N-1} S_l^T z_l | S_k = s\right] = x^T s,$$

and hence the value function satisfies

$$V_k(s,x) = x^T s - \inf_{z \in \mathcal{A}_k(x)} \mathbb{E}\left[\sum_{l=0}^{N-1} z_l^T \eta(S_l) z_l + x_l^T \lambda(S_l) x_l | S_k = s\right].$$

Proof of Proposition 4.1.3. Notice first that for any $(z_l) \in \mathcal{A}_k(x)$ the revenues

$$\sum_{l=k}^{N-1} S_l^T z_l - z_l^T \eta(S_l) z_l - x_l^T \lambda(S_l) x_l$$

are integrable. The scalar product $S_l^T z_l$ is integrable since S and z are square-integrable. Moreover, recall that η is a diagonal matrix with entries depending linearly on the price, and hence $z_l^T \eta(S_l) z_l$ is integrable. The risk part $x_l^T \lambda(S_l) x_l$ is integrable, since the maximum eigenvalue of $\lambda(S_k)$, denoted by $\bar{\lambda}(S_k)$, is square integrable for all $0 \leq k \leq N$. Indeed, by the Cauchy-Schwarz Inequality and the fact that the spectral matrix norm $\|\cdot\|$ on the set of $d \times d-$matrices is induced by the Euclidian norm on $\mathbb{R}^d$ (see e.g. Section 5.6.6 in [38]), we have

$$|x_k^T \lambda(S_k) x_k| \leq |x_k||\lambda(S_k) x_k| \leq |x_k|^2 \|\lambda(S_k)\|.$$

Since $\lambda(S_k)$ is positive semidefinite, we get $\|\lambda(S_k)\| = \bar{\lambda}(S_k)$, and hence Hölder's Inequality yields integrability of $|x_k^T \lambda(S_k) x_k|$.

We proceed by showing (4.7) and (4.9) via backward induction. For $k = N-1$ the only execution strategy is $z_{k-1} = x$. Hence, we get

$$V_{N-1} = x^T s - x^T(\eta(s) + \lambda(s))x$$

Consequently, $a_{N-1}(s) = \eta(s) + \lambda(s)$, which is positive definite and belongs to $\mathcal{L}^1$ by assumption. Let us now assume that $V_{k+1}(s,x) = x^T s - x^T a_{k+1}(s) x$ for some positive definite matrix $a_{k+1}(s)$ for every $s \in \mathbb{R}^d_{>0}$ and $a_{k+1} \in \mathcal{L}^1$. We set $d_k(s) = T a_{k+1}(s)$. The Dynamic Programming Principle implies

$$\begin{aligned} V_k(s,x) &= \sup_{z \in \mathbb{R}} \mathbb{E}[s^T z - z^T \eta(s) z - x^T \lambda(s) x + V_{k+1}(S_{k+1}, x - z)] \\ &= s^T x - x^T(\lambda(s) + d_k(s))x + \sup_{z \in \mathbb{R}}(-z^T(\eta(s) + d_k(s))z + 2x^T d_k(s) z) \quad (4.10) \end{aligned}$$

By induction hypothesis $a_{k+1}(s)$ is positive definite for every $s \in \mathbb{R}^d_{>0}$ and so is $d_k(s)$. Since $\eta(s)$ is positive definite as well, there exists a unique maximizer of (4.10) which is given by

$$z_k(s,x) = (\eta(s) + d_k(s))^{-1} d_k(s) x.$$

This implies

$$\begin{aligned} V_k(s,x) &= s^T x - x^T\left[\lambda(s) + d_k(s) - d_k(s)(\eta(s) + d_k(s))^{-1} d_k(s)\right] x \\ &= s^T x - x^T\left[\lambda(s) + \eta(s)(\eta(s) + d_k(s))^{-1} d_k(s)\right] x. \end{aligned}$$

Thus, we have

$$\begin{aligned} a_k(s) &= \eta(s)(\eta(s)+d_k(s))^{-1}d_k(s)+\lambda(s) \\ &= (d_k(s)^{-1}+\eta(s)^{-1})^{-1}+\lambda(s), \end{aligned}$$

which is positive definite for every $s \in \mathbb{R}^d_{>0}$ by assumption and induction hypothesis. Moreover, note that

$$\begin{aligned} a_k(s) &= \eta(s)-\eta(s)(\eta(s)+d_k(s))^{-1}\eta(s)+\lambda(s) \\ &= \eta(s)+\lambda(s)-(\eta^{-1}(s)+\eta^{-1}(s)d_k(s)\eta^{-1}(s))^{-1}. \end{aligned}$$

Since $(\eta^{-1}(s)+\eta^{-1}(s)d_k(s)\eta^{-1}(s))^{-1}$ is positive definite we have $a_k(s) \leq \eta(s)+\lambda(s)$. This implies $v^T a_k v \in L^1(\mathbb{R}^d_{>0}, \mu^k_s)$ for every $s \in \mathbb{R}^d_{>0}$ and $v \in \mathbb{R}^d$. Using the factorization

$$(a_k)_{i,j} = \frac{1}{2}(e_j^T a_k e_j + e_i^T a_k e_i - (e_j-e_i)^T a_k (e_j-e_i)),$$

where e_i denotes the i-th unit vector, we see that $(a_k)_{i,j} \in L^1(\mathbb{R}^d_{>0}, \mu^k_s)$. This implies $a_k \in \mathcal{L}^1$.

It remains to show that the strategy $(z_l(S_l,x_l))_{0\leq l\leq N-1}$ is L^p-integrable for all $p \geq 1$. To this end it is enough to prove that $\mathbb{E}[|x_l|^p] < \infty$ for all $p \geq 1$ and $0 \leq l \leq N-1$. We proceed via forward induction. Since x_0 is deterministic, the induction basis holds true. For the induction step, first note that

$$\begin{aligned} x_{k+1} &= x_k - z_k = (\mathrm{Id}-(\eta(S_k)+d_k(S_k))^{-1}d_k(S_k))x_k \\ &= (\eta(S_k)+d_k(S_k))^{-1}\eta(S_k)x_k, \end{aligned}$$

and hence

$$|x_{k+1}| \leq \|(\eta(S_k)+d_k(S_k))^{-1}\|\|\eta(S_k)\||x_k|.$$

The maximal eigenvalue of $(\eta(S_k)+d_k(S_k))^{-1}$ is smaller than the smallest eigenvalue of $\eta(S_k)$, and therefore

$$|x_{k+1}| \leq \kappa(\eta(S_k))|x_k|.$$

By the induction hypothesis, the integrability assumption on the condition of $\eta(S_k)$ and Hölder's Inequality, we obtain $\mathbb{E}[|x_{k+1}|^p] < \infty$ for all $p \geq 1$. □

Remark 4.1.4. Observe that in the case of a price independent risk function λ and absolute price impact Proposition 4.1.3 follows from [47, Theorem 1.3.4], where liquidation paths even allow for the presence of dark pools.

Scalar price sensitivity

If the matrix-valued risk function $\lambda(s)$ is given by a price independent covariance matrix multiplied with a nonnegative real function, then, under *absolute* price impact, a change of coordinates reduces the matrix-valued recursion (4.8) to d *independent* real function recursions.

To illustrate the dimension reduction we assume in this subsection that the price impact is absolute, and that the trader's risk preferences depend on the price *only* through the variable $r = r(s) = \prod_{j=1}^{d}(s^j)^{\beta_j}$ with $\beta_j \in \mathbb{R}$; i.e. $\lambda(s) = \xi(r)C$, where C is a covariance matrix and $\xi : \mathbb{R} \to \mathbb{R}_+$. The geometric Markovian nature of the price process implies that the process $R_k = \prod_{j=1}^{d}(S_k^j)^{\beta_j}$ satisfies the Markov property as well (notice that the process R in general is *not* a martingale). Therefore,

$$T(\xi \circ r)(s) = \mathbb{E}[\xi(R_1)|S_0 = s] = \mathbb{E}[\xi(R_1)|R_0 = r].$$

As in Proposition 4.1.3 we suppose that (A1)-(A3) hold true.

Proposition 4.1.5. *There exists a regular matrix A such that $A^TCA = \operatorname{diag}(c)$ for some $c \in \mathbb{R}_+^d$ and the solutions $b_i^k : \mathbb{R}_{>0} \to \mathbb{R}$ for $0 \le k \le N-1$ and $1 \le i \le d$ of the decoupled system of function recursions*

$$\begin{aligned} b_{N-1}^i(r) &= 1 + \xi(r)c_i \\ b_k^i(r) &= \frac{Tb_{k+1}^i(r)}{1 + Tb_{k+1}^i(r)} + \xi(r)c_i \end{aligned} \tag{4.11}$$

yield the following representation of the value function:

$$V_k(s, Ax) = s^T Ax - \sum_{i=1}^{d} b_k^i(r)x_i^2.$$

Moreover, the optimal execution strategy is given by

$$z_k(s, Ax) = A \operatorname{diag}\left(\frac{Tb_{k+1}^1(r)}{1 + Tb_{k+1}^1(r)}, \ldots, \frac{Tb_{k+1}^d(r)}{1 + Tb_{k+1}^d(r)}\right) x.$$

Proof. Let $\sqrt{\eta} = \operatorname{diag}(\sqrt{\eta_1}, \ldots, \sqrt{\eta_d})$ be the positive definite square root of η. The matrix $\sqrt{\eta}^{-1}C\sqrt{\eta}^{-1}$ is positive semidefinite by assumption. Hence, there exists an orthogonal matrix O such that $O^T\sqrt{\eta}^{-1}C\sqrt{\eta}^{-1}O = \operatorname{diag}(c)$ for some $c \in \mathbb{R}_+^d$. We set $A = \sqrt{\eta}^{-1}O$. Then we have $A^T\eta A = \operatorname{Id}$ and $A^TCA = \operatorname{diag}(c)$. Let a_k denote the matrix-valued functions introduced in Proposition 4.1.3. Then we have $V_k(s, Ax) = s^TAx - x^TA^Ta_k(s)Ax$. Hence, we have to show that

$$A^Ta_k(s)A = \operatorname{diag}(b_k^1(r), \ldots, b_k^d(r)). \tag{4.12}$$

We proceed via backward induction. At time t_{N-1} we have

$$A^Ta_{N-1}(s)A = A^T(\eta + \xi(r)C)A = \operatorname{Id} + \xi(r)\operatorname{diag}(c) = \operatorname{diag}(b_{N-1}^1(r), \ldots, b_{N-1}^d(r)).$$

For $k \leq N-2$ note that

$$a_k(s) \;=\; \eta(\eta + Ta_{k+1}(s))^{-1}Ta_{k+1}(s) + \xi(r)C = ((Ta_{k+1}(s))^{-1} + \eta^{-1})^{-1} + \xi(r)C.$$

Applying the induction hypothesis yields

$$\begin{aligned} A^T a_k(s)A &= \left[(T(A^T a_{k+1}(s)A))^{-1} + (A^T\eta A)^{-1}\right]^{-1} + \xi(r)\operatorname{diag}(c) \\ &= \operatorname{diag}(b_k^1(r), \dots, b_k^d(r)). \end{aligned}$$

Concerning the optimal execution strategy, note that we have

$$A(\operatorname{Id} + T(A^T a_{k+1}(s)A))^{-1}A^T = (\eta + Ta_{k+1}(s))^{-1}.$$

Then Proposition 4.1.3 and Equation (4.12) imply the desired result

$$\begin{aligned} z_k(s, Ax) &= (\eta + Ta_{k+1}(s))^{-1}Ta_{k+1}(s)Ax \\ &= A(\operatorname{Id} + T(A^T a_{k+1}(s)A))^{-1}A^T Ta_{k+1}(s)Ax \\ &= A\operatorname{diag}\left(\frac{Tb_{k+1}^1(r)}{1+Tb_{k+1}^1(r)}, \dots, \frac{Tb_{k+1}^d(r)}{1+Tb_{k+1}^d(r)}\right)x. \end{aligned}$$

□

4.2. Price-sensitive liquidation in continuous-time

In this section we consider the continuous-time counterpart of the stochastic control problem from Section 4.1. We restrict attention to a single asset liquidation under a linear, absolute and temporary price impact. This implies that optimal position paths are absolutely continuous and therefore uniquely determined by their derivative which is called the trading rate. The asset's forward price is assumed to be a nonnegative Brownian martingale. The requirement of closing the position up to the fixed time horizon T leads to a terminal state constraint in the control problem. Due to the linear price impact execution costs grow quadratically in the trading rate. Hence, it becomes arbitrarily expensive to close a nonzero position as time runs out. Therefore, the terminal state constraint entails that the value function of the control problem tends to infinity as time approaches maturity T. We identify the speed of the explosion and show that the value function grows inversely proportional to time to maturity. Moreover, we prove that the value function is a quadratic form in the position size. Instead of giving a direct PDE characterization of the value function we analyze how the optimal *relative* trading rates differ from the *linear* relative rate which minimizes expected execution costs. Due to the risk functional it is optimal to *enlarge* the linear relative trading rate by a factor which depends on time and the asset price. We refer to this factor as the trading rate *inflator*. As time tends to maturity T the price risk associated to an open position gets negligible and it is optimal to close the position nearly linearly. Thus the inflator converges to one as $t \nearrow T$. Hence, in contrast to the value function, the inflator converges to a finite terminal value. This allows to characterize the inflator as the unique viscosity solution of a PDE. As a corollary we also obtain a PDE characterization of the value function. Finally, we prove that the inflator of the associated time-discrete model approximation converges to the continuous-time inflator as the number of time steps converges to infinity.

4.2.1. Model description

We fix a time horizon $T > 0$. The forward price process $(S_r)_{r\in[0,T]}$ is a nonnegative Brownian martingale driven by the time-homogeneous dynamics $dS_r = \sigma(S_r)dW_r$, where (W_r) is a Brownian motion on the stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]})$ and σ is Lipschitz continuous and has linear growth, i.e. there exists a constant $L > 0$ such that for all $s_1, s_2 \in (0, \infty)$ we have

$$|\sigma(s_1) - \sigma(s_2)| \;\; \leq \;\; L|s_1 - s_2|.$$

The price process conditional to $S_t = s$ will be denoted by $(S_r^{t,s})$, $r \in [t, T]$.

An *execution strategy* of an open position x at time t is a progressively measurable process (z_r) satisfying $\int_t^T z_r dr = x$. We write $\mathcal{A}_t(x)$ for the set of all execution strategies. The open position trajectory (x_r) associated to an execution strategy (z_r) is given by $x_r = x - \int_t^r z_u du$, $r \in [t, T]$. Selling at a rate of z_t at time t is only possible at a temporarily modified price

$$\tilde{S}_t = S_t - \eta z_t,$$

where $\eta > 0$ represents the price impact parameter. The expected revenues following a strategy (z_r) satisfy

$$\mathbb{E}\left[\int_0^T z_r \tilde{S}_r dr\right] = \mathbb{E}\left[\int_0^T z_r S_r dr - \int_0^T \eta z_r^2 dr\right] = xS_0 - \mathbb{E}\int_0^T \eta z_r^2 dr.$$

The expectation, thus, consists of the so-called *initial book value* xS_0 and the *expected implementation shortfall* $\mathbb{E}\int_0^T \eta z_r^2 dr$. We assume that an agent aims at minimizing the expected implementation shortfall *and* the risk of the open position, more precisely an objective functional J defined as

$$J(t, s, x; (z_r)) = \mathbb{E}\left[\int_t^T \eta z_r^2 + \lambda(S_r)x_r^2 dr \middle| S_t = s, x_t = x\right]. \tag{4.13}$$

Here, as in Section 4.1 the continuous function $\lambda : [0, \infty) \to \mathbb{R}_+$ measures the price sensitivity of the agent's risk preferences. We distinguish two cases: If the agent intends to close a long position we assume that $\lambda(0) = \max_{s\geq 0} \lambda(s)$. If a short position has to be liquidated we assume that $\lambda(0) = \min_{s\geq 0} \lambda(s)$. We suppose that λ satisfies a polynomial growth condition, i.e. there exist $C > 0$ and $p \in \mathbb{N}$ such that

$$\lambda(s) \leq C(1 + s^p)$$

for all $s \in [0, \infty)$. Moreover, we require that for every $(t, s) \in [0, T) \times (0, \infty)$ the random variable $\lambda(S_\cdot^{t,s}) : [t, T] \times \Omega \to \mathbb{R}$ satisfies

$$\lambda(S_\cdot^{t,s}) \to \lambda(0) \quad \text{in } L^1(Leb[t, T] \times \mathbb{P}) \tag{4.14}$$

as $s \searrow 0$. Note that a geometric Brownian motion meets this requirement. In the sequel we will refer to λ as *risk function*. See Subsection 4.1.1 for possible choices of λ.

For $(t, s, x) \in [0, T) \times (0, \infty) \times \mathbb{R}$ the value function of the liquidation problem is given by

$$V(t, s, x) = \inf_{(z_r) \in \mathcal{A}_t(x)} J(t, s, x; (z_r)). \tag{4.15}$$

Remark 4.2.1. The results presented in the following can be shown to hold true also if the price process is a *time-inhomogeneous* diffusion. For ease of notation, in particular in Section 4.2.4, we stick to the homogeneous case only.

4.2.2. Formal derivation of optimal execution strategies

The Hamilton-Jacobi-Bellman Equation (HJB) associated to the stochastic control problem (4.15) is given by

$$-V_t - \frac{1}{2}\sigma(s)^2 V_{ss} - \lambda(s)x^2 - \inf_{z \in \mathbb{R}}(\eta z^2 - V_x z) = 0 \tag{4.16}$$

in $[0, T) \times (0, \infty) \times \mathbb{R}$. Since the infimum is attained at

$$z = \frac{V_x}{2\eta} \tag{4.17}$$

Equation (4.16) is equivalent to the semilinear PDE

$$-V_t - \frac{1}{2}\sigma(s)^2 V_{ss} + \frac{1}{4\eta} V_x^2 - \lambda(s)x^2 = 0. \tag{4.18}$$

The terminal state constraint $x_T = 0$ leads to the following singular terminal condition

$$\lim_{t \nearrow T} V(t, s, x) = \begin{cases} \infty \text{ for } x \neq 0 \\ 0 \text{ for } x = 0. \end{cases}$$

In Proposition 4.2.6 we derive the following two properties of V. First, we identify the growth behavior of the value function as $t \nearrow T$. Rescaling V by the factor $T - t$ leads to the finite terminal condition $\lim_{t \nearrow T}(T - t)V(t, s, x) = \eta x^2$. Moreover, one can reduce the dimension of the state space, since in the x variable the value function has the explicit representation $V(t, s, x) = V(t, s, 1)x^2$. It is convenient, therefore, to introduce the function

$$I(t, s) := \frac{T - t}{\eta} V(t, s, 1) \tag{4.19}$$

in $[0,T)\times(0,\infty)$ and study the PDE

$$-I_t - \frac{1}{2}\sigma(s)^2 I_{ss} - \frac{I}{T-t} + \frac{I^2}{(T-t)} - (T-t)\frac{\lambda(s)}{\eta} = 0 \tag{4.20}$$

with finite terminal condition $I(T,s)=1$. Observe that by the transformation (4.19) the singularity at time T moved from the terminal condition of the PDE (4.18) to the generator of the PDE (4.20). In the following paragraph we present an economic interpretation of the function I.

Interpretation of I as trading rate inflator

As Equation (4.17) suggests the optimal trading rate z is given in feedback form by

$$z(t,s,x) = \frac{V_x(t,s,x)}{2\eta} = I(t,s)\frac{x}{T-t}.$$

Hence, the optimal *relative* trading rate $r(t,s,x) = \frac{z(t,s,x)}{x}$ is independent of the remaining open position x and satisfies $r(t,s) = \frac{I(t,s)}{T-t}$. In the risk-neutral case where the agent is aiming at minimizing the expected implementation shortfall $\mathbb{E}\int_0^T \eta z_r^2 dr$ it follows from Jensen's inequality that it is optimal to close the position linearly $x_t = \frac{T-t}{T}x_0$. In this case the optimal relative trading rate is given by $r^{\text{lin}}(t) = \frac{1}{T-t}$. Consequently, the factor I describes the deviation from a linear closure in the case with nonvanishing risk function

$$r(t,s) = I(t,s)r^{\text{lin}}(t)$$

In Proposition 4.2.6 we show that I is bounded from below by one. Therefore, agents with price-sensitive risk preferences increase the trading speed compared to agents with risk-neutral preferences. The factor $I(t,s)$ specifies the enlargement of the relative trading rate in dependence on the asset price s at time $t\in[0,T]$. Therefore, we refer to I as the trading rate *inflator* in the sequel.

4.2.3. Characterization of optimal execution strategies

In this Subsection we characterize the value function and the optimal execution strategy by appealing to viscosity solutions of PDE. To this end we employ the explicit representation of the value function for risk functions λ that are price independent, in which case the control problem becomes deterministic. If λ is constant equal to $c\in\mathbb{R}_+$ we define the value function

$$V_c(t,x) = \inf_{(z_r)\in\mathcal{A}_t(x),\ \text{det}} \int_t^T \eta z_r^2 + c x_r^2 dr,$$

for all $(t,x)\in[0,T)\times\mathbb{R}$. Moreover we introduce the function $I_c:[0,T]\to[1,\infty)$ by

$$I_c(t) = \begin{cases} \sqrt{\frac{c}{\eta}}(T-t)\coth\left(\sqrt{\frac{c}{\eta}}(T-t)\right) & \text{if } c>0 \\ 1 & \text{if } c=0. \end{cases}$$

The next result stems from [47, Section 2.5.1], where optimal liquidation in presence of dark pools is considered. Note that $\coth(x) = \frac{e^{2x}+1}{e^{2x}-1}$ denotes the hyperbolic cotangent function.

Lemma 4.2.2. *In the case where the risk function is constant equal to $c \geq 0$, the optimal strategy of closing a postion $x \in \mathbb{R}$ at time $t \in [0, T)$ is given by $z_r = I_c(r)\frac{x_r}{T-r}$ for $r \in [t, T]$. In particular it is deterministic and the associated position trajectory is strictly decreasing (increasing) if $x > 0$ ($x < 0$). The value function is given by $V_c(t, x) = \frac{\eta}{T-t} I_c(t) x^2$.*

Before stating the main theorem of this section, we remark that the control problem (4.15) has a finite value function.

Remark 4.2.3. The assumptions made in Section 4.2.1 imply that the value function is finite for every $(t, s, x) \in [0, T) \times (0, \infty) \times \mathbb{R}$. The Lipschitz and linear growth conditions imposed on σ imply the following moment estimate (cf. [51, Theorem 3.2]). For any $p \in \mathbb{N}$ there exists a constant $K > 0$ such that

$$\mathbb{E}[\sup_{t \leq r \leq T} (S_r^{t,s})^p] \leq K(1 + s^p). \tag{4.21}$$

Choosing the execution strategy with constant trading rate $z_r = x/(T-t)$ for all $r \in [t, T)$ implies

$$\begin{aligned} V(t, s, x) &\leq J(t, s, x; (z_r)) \\ &= \eta \frac{x^2}{T-t} + x^2 \mathbb{E}\left[\int_t^T \lambda(S_r^{t,s}) \left(\frac{T-r}{T-t}\right)^2 dr\right] \\ &\leq \eta \frac{x^2}{T-t} + C x^2 \mathbb{E}\left[\int_t^T 1 + (S_r^{t,s})^p dr\right] \\ &\leq \eta \frac{x^2}{T-t} + C x^2 (T-t)(1 + K(1 + s^p)). \end{aligned}$$

In particular, the value function is locally bounded in $[0, T) \times (0, \infty) \times \mathbb{R}$.

The following theorem gives a rigorous justification of the results in Subsection 4.2.2. The optimal trading rate at time $t \in [0, T)$, given an open position $x \in \mathbb{R}$ and a price $s \in (0, \infty)$, is equal to $z(t, s, x) = I(t, s)\frac{x}{T-t}$ and the inflator I is a the unique viscosity solution of Equation (4.20).

Notice that the PDE (4.20) has a singularity in the time variable at time T. Therefore we can not refer to standard results guaranteeing that the PDE possesses a unique solution. We provide a self-contained proof of existence and uniqueness. In a first step we need to show that the inflator is bounded from below by 1, and bounded from above by a function that converges to 1 as $t \uparrow T$.

Theorem 4.2.4. *The inflator I is the unique viscosity solution of* (4.20) *in* $[0,T) \times (0,\infty)$ *that is bounded from below by one, has polynomial growth in s and satisfies the boundary conditions*

$$\begin{aligned} &\lim_{\substack{t\nearrow T\\ s\to s_0}} I(t,s) = 1 && \textit{for all } s_0 \in (0,\infty),\\ &\lim_{\substack{t\to t_0\\ s\searrow 0}} I(t,s) = I_{\lambda(0)}(t_0) && \textit{for all } t_0 \in [0,T). \end{aligned} \tag{4.22}$$

Moreover, I is continuous.

The optimal execution strategy is Markovian: For an open position $x \in \mathbb{R}$ and a price $s \in (0,\infty)$ at time $t \in [0,T)$ the optimal trading speed is given by $z(t,s,x) = I(t,s)\frac{x}{T-t}$. This leads to the associated position trajectory

$$x_r = x \exp\left(- \int_t^r \frac{I(u, S_u^{t,s})}{T-u} du \right). \tag{4.23}$$

The proof of Theorem 4.2.4 is split into several parts. First we show that the inflator is a solution of the PDE (4.20) (Prop. 4.2.6); then we show that it is the unique one. Finally, for the proof that $z(t,s,x) = I(t,s)\frac{x}{T-t}$ is the optimal trading rate we use the discrete model approximations provided by Section 4.1. We show that the optimal trading strategies from scaled discrete models converge to the optimal continuous-time trading rate (Section 4.2.4).

Remark 4.2.5. Under additional assumptions on the volatility function σ, e.g. a uniform ellipticity condition, one can probably show that $I(t,s)$ is a *classical solution* of the PDE (4.20). In general, however, if σ is degenerate, no classical solution of (4.20) may exist. For example if σ is equal to zero, and λ is not differentiable, then $I_{\lambda(s)}(t)$ is a solution in the viscosity, but not in the classical sense.

Let us introduce some notation: For a locally bounded function $f : \mathcal{O} \subset \mathbb{R}^d \to \mathbb{R}$ we define its upper semicontinuous (u.s.c.) envelope f^* and its lower semicontinuous (l.s.c.) envelope f_* by

$$f^*(y) := \limsup_{y'\to y} f(y'), \quad f_*(y) := \liminf_{y'\to y} f(y'), \quad y \in \mathcal{O}.$$

For the definition of viscosity solutions we refer to [68, Definition 4.2.1]. We will also make use of the equivalent characterization of viscosity solutions via super- and subjets [68, Lemma 4.4.5] or [28, Chapter V, Lemma 4.1].

Proposition 4.2.6.

i) For all $(t,s,x) \in [0,T) \times (0,\infty) \times \mathbb{R}$ the value function is given by $V(t,s,x) = I(t,s)\frac{\eta x^2}{T-t}$.

ii) The inflator I is a viscosity solution of (4.20) *in* $[0,T) \times (0,\infty)$ *with boundary conditions* (4.22)*. Moreover, I has polynomial growth of order p in s and is bounded from below by one.*

Proof. i) Fix $(t,s) \in [0,T) \times (0,\infty)$. Let $y_1 \neq 0$ and let (z_r^n) be a sequence of execution strategies in $\mathcal{A}_t(y_1)$ such that $V(t,s,y_1) = \lim_{n\to\infty} J(t,s,y_1;(z_r^n))$. Then for $y_2 \in \mathbb{R}$ we define a sequence $(\xi_r^n) \in \mathcal{A}_t(y_2)$ by $\xi_r^n = z_r^n y_2/y_1$. Then we get

$$V(t,s,y_2) \leq \lim_{n\to\infty} J(t,s,y_2;(\xi_r^n)) = \frac{y_2^2}{y_1^2} \lim_{n\to\infty} J(t,s,y_1;(z_r^n)) = \frac{y_2^2}{y_1^2} V(t,s,y_1).$$

For $x \neq 0$ choosing $y_1 = 1$ and $y_2 = x$ yields $V(t,s,x) \leq V(t,s,1)x^2$, while choosing $y_1 = x$ and $y_2 = 1$ yields the opposite inequality. The case $x = 0$ is trivial.

ii) We will only show that I is a viscosity supersolution, since the arguments for the subsolution property are identical. Fix a point $(t_0,s_0) \in [0,T) \times (0,\infty)$ and let $\varphi \in C^{1,2}([0,T) \times (0,\infty))$ such that $\varphi \leq I_*$ and $\varphi(t_0,s_0) = I_*(t_0,s_0)$. Then *i)* implies that we have

$$\begin{aligned} \tilde{\varphi}(t,s,x) &:= \frac{\eta x^2}{T-t}\varphi(t,s) \leq V_*(t,s,x) \\ \tilde{\varphi}(t_0,s_0,x) &= V_*(t_0,s_0,x) \end{aligned}$$

for all $(t,s,x) \in [0,T) \times (0,\infty) \times \mathbb{R}$. It is straightforward to show that V is a viscosity solution of (4.18) in $[0,T) \times (0,\infty) \times \mathbb{R}$ (see e.g. [68, Section 4.3] for verification results). This implies at (t_0,s_0) for all $x \in \mathbb{R}$

$$\begin{aligned} 0 &\leq -\tilde{\varphi}_t - \frac{1}{2}\sigma^2\tilde{\varphi}_{ss} + \frac{1}{4\eta}\tilde{\varphi}_x^2 - \lambda(s_0)x^2 \\ &= \frac{\eta x^2}{T-t_0}\left(-\varphi_t - \frac{1}{2}\sigma^2\varphi_{ss} - \frac{\varphi}{T-t_0} + \frac{\varphi^2}{T-t_0} - (T-t_0)\frac{\lambda(s_0)}{\eta}\right), \end{aligned}$$

which yields the viscosity supersolution property.

Let us now show that I is bounded from below by one. In the case of a vanishing risk function (i.e. $c = 0$) Lemma 4.2.2 implies that it is optimal to close an open position linearly and $V_0(t,x) = \frac{\eta}{T-t}x^2$. Moreover, we have $V_0(t,x) \leq V(t,s,x)$ for every $(t,s,x) \in [0,T) \times (0,\infty) \times \mathbb{R}$. This implies $I \geq 1$. Now, let us turn to the terminal condition $\lim_{\substack{t\nearrow T \\ s\to s_0}} I(t,s) = 1$. Closing the postion $x = 1$ at time t with constant trading rate $z_r = \frac{1}{T-t}$ is suboptimal and hence we have

$$\begin{aligned} I(t,s) &\leq 1 + \frac{T-t}{\eta}\mathbb{E}\left[\int_t^T \lambda(S_r^{t,s})\left(\frac{T-r}{T-t}\right)^2 dr\right] && (4.24) \\ &\leq 1 + \frac{T-t}{\eta}\mathbb{E}\left[\int_t^T \lambda(S_r^{t,s})\right] && (4.25) \\ &\to 1 && (4.26) \end{aligned}$$

as $t \nearrow T$. Since I is bounded from below by one this yields the claim.

For the second boundary condition, note that we have to show that

$$\lim_{\substack{t\to t_0\\ s\searrow 0}} V(t,s,1) = V_{\lambda(0)}(t_0,1)$$

for any $t_0 < T$. Given a point $(t,s)\in[0,T)\times(0,\infty)$ let $(\tilde z_r)\in\mathcal{A}_t(1)$ denote an ϵ-optimal strategy of $V(t,s,1)$. We define $z_r := \tilde z_r \mathbf{1}_{\tilde x_r\in[0,1]}$, where $(\tilde x_r)$ denotes the position trajectory associated to $(\tilde z_r)$. Then, (z_r) belongs to $\mathcal{A}_t(1)$ and is ϵ-optimal as well and the position trajectory (x_r) satisfies $0\le x_r\le 1$ a.s. for every $r\in[t,T]$. This implies

$$\begin{aligned} V_{\lambda(0)}(t,1) - V(t,s,1) &\le \mathbb{E}\left[\int_t^T (\lambda(0)-\lambda(S_r^{t,s}))x_r^2 dr\right] + \epsilon \\ &\le \mathbb{E}\left[\int_t^T |\lambda(0)-\lambda(S_r^{t,s})| dr\right] + \epsilon. \end{aligned}$$

Letting ϵ tend to 0 implies

$$V_{\lambda(0)}(t,1) - V(t,s,1) \le \mathbb{E}\left[\int_t^T |\lambda(0)-\lambda(S_r^{t,s})| dr\right] \to 0,$$

as $t\to t_0$ and $s\searrow 0$ by assumption (4.14). For the opposite inequality, let (z_r) denote the optimal strategy from Lemma 4.2.2 for closing the position $x=1$ at time $t<T$ with constant risk function $c=\lambda(0)$. Then we have

$$V(t,s,1) - V_{\lambda(0)}(t,1) \le \mathbb{E}\left[\int_t^T (\lambda(S_r^{t,s})-\lambda(0))x_r^2 dr\right].$$

Note that (x_r) is deterministic and strictly decreasing. Then the same considerations as above yield the claim.

The polynomial growth of I is a direct consequence of Remark 4.2.3, which completes the proof.

□

Remark 4.2.7. In the case where the agent intends to close a long position, it is convenient to assume that $\lambda(s)\to 0$ as $s\to\infty$. If, moreover, $\lambda(S_\cdot^{t,s})$ converges to 0 in $L^1(Leb[t,T]\times\mathbb{P})$ as $s\to\infty$, then not only a polynomial growth, but even the finite boundary condition

$$\lim_{\substack{t\to t_0\\ s\to\infty}} I(t,s) = 1 \quad \text{for} \quad t_0\in[0,T),$$

can be derived. This is a direct consequence of (4.25) and the fact that I is bounded from below by one.

To complete the characterization of the inflator we derive a uniqueness result for (4.20) via the following comparison principle.

Proposition 4.2.8. *Let I_1 (respectively I_2) be a u.s.c. viscosity subsolution (respectively l.s.c. viscosity supersolution) of* (4.20) *in* $[0, T) \times (0, \infty)$*, satisfying the boundary conditions*

$$\begin{aligned} \lim_{\substack{t \nearrow T \\ s \to s_0}} I_1(t,s) &\le \lim_{\substack{t \nearrow T \\ s \to s_0}} I_2(t,s) \quad \textit{for all } s_0 \in (0,\infty), \\ \lim_{\substack{t \to t_0 \\ s \searrow 0}} I_1(t,s) &\le \lim_{\substack{t \to t_0 \\ s \searrow 0}} I_2(t,s) \quad \textit{for all } t_0 \in [0,T). \end{aligned} \tag{4.27}$$

Moreover, assume that I_1 and I_2 are bounded from below by one and have polynomial growth in s, i.e. there exist $K > 0, q \in \mathbb{N}$ such that for all $(t,s) \in [0,T) \times (0,\infty)$ we have

$$1 \le I_1(t,s), I_2(t,s) \le K(1 + s^q). \tag{4.28}$$

Then, we have $I_1 \le I_2$ in $[0,T) \times (0,\infty)$.

Proof. Let $M := \sup_{[0,T)\times(0,\infty)} I_1 - I_2$. We divide the proof into two parts. In the first part we consider the special case where M is attained in $[0,T) \times (0,\infty)$ and show that this implies $M \le 0$. In the second part we illustrate how to employ conditions (4.27) and (4.28) to reduce the general case to this special case.

Part i) Assume that M is attained in some bounded set $\mathcal{O} \subset [0,T) \times (0,\infty)$, i.e. $M = \max_{\mathcal{O}} I_1 - I_2$. We argue by contradiction and assume that $M > 0$. For $\epsilon > 0$ we introduce the functions $\varphi_\epsilon, \Phi_\epsilon : [0,T)^2 \times (0,\infty)^2 \to \mathbb{R}$ defined by

$$\begin{aligned} \varphi_\epsilon(t_1,t_2,s_1,s_2) &:= \frac{1}{2\epsilon}\left((t_1-t_2)^2 + (s_1-s_2)^2\right) \\ \Phi_\epsilon(t_1,t_2,s_1,s_2) &:= I_1(t_1,s_1) - I_2(t_2,s_2) - \varphi_\epsilon(t_1,t_2,s_1,s_2). \end{aligned}$$

Let $M_\epsilon := \max_{\bar{\mathcal{O}}^2} \Phi_\epsilon$ denote the maximum of Φ_ϵ in $\bar{\mathcal{O}}^2$, which is attained by some point $(t_1^\epsilon, t_2^\epsilon, s_1^\epsilon, s_2^\epsilon) \in \bar{\mathcal{O}}^2$ since Φ_ϵ is u.s.c.. Note that we have

$$\begin{aligned} M \le M_\epsilon &= I_1(t_1^\epsilon, s_1^\epsilon) - I_2(t_2^\epsilon, s_2^\epsilon) - \varphi_\epsilon(t_1^\epsilon, t_2^\epsilon, s_1^\epsilon, s_2^\epsilon) \\ &\le I_1(t_1^\epsilon, s_1^\epsilon) - I_2(t_2^\epsilon, s_2^\epsilon), \end{aligned} \tag{4.29}$$

which is equivalent to

$$0 \le \varphi_\epsilon(t_1^\epsilon, t_2^\epsilon, s_1^\epsilon, s_2^\epsilon) \le I_1(t_1^\epsilon, s_1^\epsilon) - I_2(t_2^\epsilon, s_2^\epsilon) - M. \tag{4.30}$$

Since $\bar{\mathcal{O}}^2$ is compact, we have $\lim_{\epsilon \to 0}(t_1^\epsilon, s_1^\epsilon, t_2^\epsilon, s_2^\epsilon) = (\bar{t}_1, \bar{s}_1, \bar{t}_2, \bar{s}_2)$ for some subsequence. We will only work with this subsequence in the sequel. Since $I_1 - I_2$ is u.s.c., the sequence $(I_1(t_1^\epsilon, s_1^\epsilon) - I_2(t_2^\epsilon, s_2^\epsilon))$ is bounded from above, and so is $(\varphi_\epsilon(t_1^\epsilon, t_2^\epsilon, s_1^\epsilon, s_2^\epsilon))$ by (4.30). This implies $\bar{t}_1 = \bar{t}_2 =: \bar{t}$ and $\bar{s}_1 = \bar{s}_2 =: \bar{s}$. Sending ϵ to 0 in (4.29) yields

$$M \le \limsup_{\epsilon \to 0} I_1(t_1^\epsilon, s_1^\epsilon) - I_2(t_2^\epsilon, s_2^\epsilon) \le I_1(\bar{t}, \bar{s}) - I_2(\bar{t}, \bar{s}) \le M.$$

Hence, we get

$$\lim_{\epsilon\to 0} I_1(t_1^\epsilon, s_1^\epsilon) - I_2(t_2^\epsilon, s_2^\epsilon) = I_1(\bar{t},\bar{s}) - I_2(\bar{t},\bar{s}) = M > 0 \tag{4.31}$$

and (4.30) implies

$$\lim_{\epsilon\to 0} \varphi_\epsilon(t_1^\epsilon, t_2^\epsilon, s_1^\epsilon, s_2^\epsilon) = 0. \tag{4.32}$$

Let $\bar{\mathcal{P}}^{2,\pm} I_{1,2}(t,s)$ denote the closure of the second order super- resp. subjets of I_1 resp. I_2 at (t,s) (see [40, Section 2] for the definition). By Ishii's Lemma (see [68, Lemma 4.4.6] and [68, Remark 4.4.9], or [40, Theorem 8.3]) there exist $h_1^\epsilon, h_2^\epsilon \in \mathbb{R}_+$ such that

$$\begin{aligned} \left(\frac{1}{\epsilon}(t_1^\epsilon - t_2^\epsilon), \frac{1}{\epsilon}(s_1^\epsilon - s_2^\epsilon), h_1^\epsilon\right) &\in \bar{\mathcal{P}}^{2,+} I_1(t_1^\epsilon, s_1^\epsilon), \\ \left(\frac{1}{\epsilon}(t_1^\epsilon - t_2^\epsilon), \frac{1}{\epsilon}(s_1^\epsilon - s_2^\epsilon), h_2^\epsilon\right) &\in \bar{\mathcal{P}}^{2,-} I_2(t_2^\epsilon, s_2^\epsilon) \end{aligned}$$

and

$$\sigma^2(s_1^\epsilon)h_1^\epsilon - \sigma^2(s_2^\epsilon)h_2^\epsilon \le \frac{3}{\epsilon}|\sigma(s_1^\epsilon) - \sigma(s_2^\epsilon)|^2. \tag{4.33}$$

Then the subsolution property of I_1 and the supersolution property of I_2 imply respectively

$$-\frac{1}{\epsilon}(t_1^\epsilon - t_2^\epsilon) - \frac{1}{2}\sigma^2(s_1^\epsilon)h_1^\epsilon - \frac{I_1(t_1^\epsilon, s_1^\epsilon)}{T - t_1^\epsilon} + \frac{I_1(t_1^\epsilon, s_1^\epsilon)^2}{T - t_1^\epsilon} - (T - t_1^\epsilon)\frac{\lambda(s_1^\epsilon)}{\eta} \le 0$$

and

$$-\frac{1}{\epsilon}(t_1^\epsilon - t_2^\epsilon) - \frac{1}{2}\sigma^2(s_2^\epsilon)h_2^\epsilon - \frac{b_2(t_2^\epsilon, s_2^\epsilon)}{T - t_2^\epsilon} + \frac{b_2(t_2^\epsilon, s_2^\epsilon)^2}{T - t_2^\epsilon} - (T - t_2^\epsilon)\frac{\lambda(s_2^\epsilon)}{\eta} \ge 0.$$

Subtracting the second from the first inequality yields

$$A_1(\epsilon) + A_2(\epsilon) + A_3(\epsilon) \le 0$$

with

$$\begin{aligned} A_1(\epsilon) &= \frac{1}{2}\left(\sigma^2(s_2^\epsilon)h_2^\epsilon - \sigma^2(s_1^\epsilon)h_1^\epsilon\right), \\ A_2(\epsilon) &= (T - t_2^\epsilon)\frac{\lambda(s_2^\epsilon)}{\eta} - (T - t_1^\epsilon)\frac{\lambda(s_1^\epsilon)}{\eta}, \\ A_3(\epsilon) &= \frac{I_2(t_2^\epsilon, s_2^\epsilon)}{T - t_2^\epsilon} - \frac{I_1(t_1^\epsilon, s_1^\epsilon)}{T - t_1^\epsilon} - \frac{I_2(t_2^\epsilon, s_2^\epsilon)^2}{T - t_2^\epsilon} + \frac{I_1(t_1^\epsilon, s_1^\epsilon)^2}{T - t_1^\epsilon}. \end{aligned}$$

Applying Inequality (4.33), the Lipschitz continuity of σ and Equation (4.32) implies $\liminf_{\epsilon\to 0} A_1(\epsilon) \geq 0$. Moreover, continuity of λ yields $\lim_{\epsilon\to 0} A_2(\epsilon) = 0$. A straightforward calculation shows that

$$\begin{aligned} A_3(\epsilon) &= \frac{I_1(t_1^\epsilon, s_1^\epsilon) + I_2(t_2^\epsilon, s_2^\epsilon) - 1}{T - t_2^\epsilon}(I_1(t_1^\epsilon, s_1^\epsilon) - I_2(t_2^\epsilon, s_2^\epsilon)) \\ &\quad + \frac{(1 - I_1(t_1^\epsilon, s_1^\epsilon))I_1(t_1^\epsilon, s_1^\epsilon)}{(T - t_2^\epsilon)(T - t_1^\epsilon)}(t_2^\epsilon - t_1^\epsilon). \end{aligned}$$

Since I_1 and I_2 are bounded from below by one and $I_1(t_1^\epsilon, s_1^\epsilon) - I_2(t_2^\epsilon, s_2^\epsilon)$ is positive by (4.29), we have

$$A_3(\epsilon) \geq \frac{I_1(t_1^\epsilon, s_1^\epsilon) - I_2(t_2^\epsilon, s_2^\epsilon)}{T - t_2^\epsilon} + \frac{(1 - I_1(t_1^\epsilon, s_1^\epsilon))I_1(t_1^\epsilon, s_1^\epsilon)}{(T - t_2^\epsilon)(T - t_1^\epsilon)}(t_2^\epsilon - t_1^\epsilon).$$

Note that the sequence $((1 - I_1(t_1^\epsilon, s_1^\epsilon))I_1(t_1^\epsilon, s_1^\epsilon))$ is bounded, since I_1 is u.s.c. and bounded from below by one. Then (4.31) implies

$$\liminf_{\epsilon\to 0} A_3(\epsilon) \geq \frac{M}{T - \bar{t}}.$$

Putting everything together we obtain the desired contradiction

$$0 \geq \liminf_{\epsilon\to 0} A_1(\epsilon) + A_2(\epsilon) + A_3(\epsilon) \geq \frac{M}{T - \bar{t}} > 0.$$

Part ii) We introduce the function $\phi : [0, T] \times [0, \infty) \to \mathbb{R}$, $\phi(t, s) = e^{-\alpha t}\frac{1+s^{2q}}{T-t}$ with $\alpha \geq \sigma(s)^2 q(2q-1)\frac{s^{2(q-1)}}{1+s^{2q}}$ for all $s \in [0, \infty)$. Note that the linear growth condition on σ guarantees the existence of $\alpha \in \mathbb{R}$. Then for $\epsilon > 0$ we define the function $I_2^\epsilon(t, s) = I_2(t, s) + \epsilon\phi(t, s)$. A straightforward calculation yields

$$-\phi_t - \frac{1}{2}\sigma^2\phi_{ss} + \frac{\phi}{T - t} \geq 0.$$

Since I_2 is bounded from below by one we get

$$\begin{aligned} \epsilon\left(-\phi_t - \frac{1}{2}\sigma^2\phi_{ss} - \frac{\phi}{T - t}\right) + \frac{(I_2^\epsilon)^2}{T - t} &\geq \epsilon\left(-\phi_t - \frac{1}{2}\sigma^2\phi_{ss} + \frac{\phi}{T - t}\right) + \frac{I_2^2}{T - t} \\ &\geq \frac{I_2^2}{T - t}. \end{aligned}$$

Then the supersolution property of I_2 implies that I_2^ϵ is a viscosity supersolution of (4.20) as well. Assume that $M_\epsilon := \sup_{[0,T)\times(0,\infty)} I_1 - I_2^\epsilon > 0$. Then the polynomial growth condition and Inequalities (4.27) and (4.28) imply that M_ϵ is attained in $[0, T) \times (0, \infty)$. But then Part i) implies $M_\epsilon \leq 0$. Hence, we have for every $(t, s) \in [0, T) \times (0, \infty)$

$$I_1(t, s) - I_2(t, s) - \epsilon\phi(t, s) \leq 0.$$

Letting ϵ tend to 0 completes the proof.

□

Proof of Theorem 4.2.4. We check the uniqueness and continuity of I simultaneously, since both are consequences of Proposition 4.2.8. Let $\tilde{I}$ be another viscosity solution of (4.20) with $\tilde{I} \geq 1$, having polynomial growth in s and satisfying (4.22). Then $\tilde{I}_*$ and $\tilde{I}^*$ satisfy (4.22) as well. By Proposition 4.2.6 we have that I^* is a u.s.c. viscosity subsolution of (4.20), I_* is a l.s.c. viscosity supersolution of (4.20) and both satisfy (4.22). Moreover, I_* and I^* as well as $\tilde{I}_*$ and $\tilde{I}^*$ are bounded from below by one and have polynomial growth in s, since I and $\tilde{I}$ have these properties. Hence, by Proposition 4.2.8 we have $\tilde{I}^* \leq I_*$ and $I^* \leq \tilde{I}_*$ in $[0, T) \times (0, \infty)$, which implies

$$\tilde{I}_* \leq \tilde{I}^* \leq I_* \leq I^* \leq \tilde{I}_*.$$

But this means $I_* = I^* = I = \tilde{I}$.

For the representation of the optimal execution strategy we refer the reader to Corollary 4.2.10 in Section 4.2.4 below.

□

4.2.4. Convergence of the discrete-time inflator

The first aim of this section is to characterize inflators in the *discrete-time* version of the model studied in 4.1. We then show that the discrete-time inflator converges to its continuous-time counterpart if model parameters are properly scaled. The convergence allows us to give a simple proof that the candidate for the optimal position path in continuous-time, given in (4.23), is indeed optimal (see Corollary 4.2.10).

We next briefly recall the formulation of the discrete-time optimal liquidation problem with N trading periods. Notice that we *only* discretize the time variable, but not price and position. For $n \leq N - 1$ and $(s, x) \in (0, \infty) \times \mathbb{R}$ the value function is defined by

$$V_n^N(s, x) := \inf_{(z_k) \in \mathcal{A}_k(x)} \mathbb{E}\left[\sum_{k=n}^{N-1} \eta^N z_k^2 + \lambda^N(S_k^N) x_k^2 \middle| S_n^N = s, x_n = x\right]. \tag{4.34}$$

Let us introduce the discretization parameter $h^N = T/N$. The price impact and the risk function are rescaled in the following way (see [47, Page 130] for an explanation)

$$\eta^N = \frac{N}{T}\eta = \frac{\eta}{h^N}, \quad \lambda^N = \frac{T}{N}\lambda = h^N \lambda, \tag{4.35}$$

where η is the price impact parameter and λ the risk function in continuous-time from Section 4.2.3. The discrete-time price process $(S_n^N)_{n \in \mathbb{N}}$ is given by $S_n^N = S_{nh^N}$, where (S_t) is the time-continuous price process introduced in Section 4.2.3. Proposition 4.1.3 yields the following semi-explicit representation of the value function V^N

$$V_n^N(s, x) = a_n^N(s) x^2,$$

where the functions a_n^N are determined by the backward recurrence relation

$$a_{N-1}^N(s) = \eta^N + \lambda^N(s), \quad a_n^N(s) = \frac{\eta^N \mathbb{E}[a_{n+1}^N(S_{n+1}^N)|S_n = s]}{\eta^N + \mathbb{E}[a_{n+1}^N(S_{n+1}^N)|S_n = s]} + \lambda^N(s). \tag{4.36}$$

We define the time-continuous versions of V^N and a^N as follows. For $t \in [nh^N, (n+1)h^N)$ we set $V^N(t,s,x) := V_n^N(s,x)$ and $a^N(t,s) := a_n^N(s)$ for all $s \in (0,\infty)$ and $x \in \mathbb{R}$. As in Section 4.2.3 we rescale a^N by the factor $\frac{T-t}{\eta}$ and define $I^N(t,s) := \frac{T-t}{\eta} a^N(t,s)$ for all $(t,s) \in [0,T) \times (0,\infty)$. Note that the time-homogeneity of (S_t) implies that I^N satisfies

$$I^N(t,s) = (T-t)\left(\frac{\mathbb{E}[I^N(t+h^N, S_{t+h^N})|S_t = s]}{T-t-h^N+h^N\mathbb{E}[I^N(t+h^N, S_{t+h^N})|S_t = s]} + h^N\frac{\lambda(s)}{\eta}\right) \tag{4.37}$$

for $t < T - h^N$ and $s \in (0,\infty)$. Let us introduce the u.s.c. respectively l.s.c. functions

$$\overline{I}(t,s) := \limsup_{\substack{N\to\infty \\ (t',s')\to(t,s)}} I^N(t',s'), \qquad \underline{I}(t,s) := \liminf_{\substack{N\to\infty \\ (t',s')\to(t,s)}} I^N(t',s').$$

Next, we state the main result of this section, which shows that I^N converges to I locally uniformly in (t,s,x) as N tends to ∞.

Theorem 4.2.9. *In $[0,T) \times (0,\infty)$ we have $\overline{I} = \underline{I} = I$. In particular, we have $V^N \to V$ pointwise in $[0,T) \times (0,\infty) \times \mathbb{R}$ as $N \to \infty$.*

In Proposition 4.1.3 we showed that the optimal execution strategy in discrete-time with N trading periods is Markovian: Given an open position $x \in \mathbb{R}$ and a price $s \in (0,\infty)$ in the n-th trading period, the optimal amount to sell is $z_n^N(s,x) = h^N f_n^N(s)x$ with $f_n^N(s) = \frac{I^N(nh^N,s)}{T-nh^N} - h^N\frac{\lambda(s)}{\eta}$. The associated position trajectory $(x_k^N)_{k\geq n}$ (we suppress the dependence on x, s, n and ω here and in the sequel) is given by

$$\begin{aligned} x_k^N &= x_{k-1}^N - z_{k-1}(S_{k-1}^N, x_{k-1}^N) = \left(1 - h^N f_{k-1}^N(S_{k-1}^N)\right) x_{k-1}^N \\ &= x\prod_{i=n}^{k-1}(1 - h^N f_i^N(S_i^N)). \end{aligned} \tag{4.38}$$

For $t \in [kh^N, (k+1)h^N)$ we set $x^N(t) = x_k^N$.

Corollary 4.2.10. *The discrete-time optimal position path converges to the continuous-time optimal trajectory. More precisely, let x^N denote the optimal trajectory for liquidating a position x in N trading periods starting at time t with initial price s. Then we have for every $r \geq t$*

$$\lim_{N\to\infty} x^N(r) = x\exp\left(-\int_0^r \frac{I(u, S_u^{t,s})}{T-u}du\right) =: x(r), \quad a.s.,$$

and

$$\lim_{N\to\infty} V^N(t,s,x) = J(t,s,x,(\dot{x})). \tag{4.39}$$

In particular, this implies $J(t,s,x,(\dot{x})) = V(t,s,x)$.

Proof. Without loss of generality we take $t = 0$. For $r \in [0, T)$ and $N \in \mathbb{N}$ let $n \in \mathbb{N}$ such that $r \in [nh^N, (n+1)h^N)$. Fix $(s, x) \in (0, \infty) \times \mathbb{R}$. From (4.38) we have that

$$\log \frac{x^N(r)}{x} = \sum_{i=0}^{n-1} \log(1 - h^N f_i^N(S_{ih^N})) = h^N \sum_{i=0}^{n-1} \frac{\log(1 - h^N f_i^N(S_{ih^N}))}{h^N f_i^N(S_{ih^N})} f_i^N(S_{ih^N}).$$

Note that Theorem 4.2.9 yields $f_n^N(S_{nh^N}) \to \frac{I(r, S_r^{t,s})}{T-r}$ as $N \to \infty$. Then Lebesgue's theorem of dominated convergence (see Claim III in the proof of Theorem 4.2.9) implies

$$\lim_{N\to\infty} \log \frac{x^N(r)}{x} = -\int_0^r \frac{I(u, S_u^{t,s})}{T-u} du, \quad \text{a.s.}.$$

From Proposition 4.1.3 we know that

$$V^N(0, s, x) = h^N \mathbb{E}\left[\sum_{i=0}^{N-1} \left(\eta(f_i^N(S_{ih^N}))^2 + \lambda(S_{ih^N})\right)(x_i^N)^2\right].$$

Appealing again to Lebesgue's theorem of dominated convergence yields (4.39). □

We will prove Theorem 4.2.9 using the comparison principle Proposition 4.2.8 and the characterization of I provided by Theorem 4.2.4. Thus, one part of the proof is to verify that $\overline{I}$ and $\underline{I}$ satisfy the boundary conditions (4.22). This will be established using the following lemmata. In [47] optimal liquidation with price insensitve risk function in the presence of dark pools was studied. We will make use of a special case of [47, Theorem 2.6.3.], which guarantees that the discrete-time inflator converges to the continuous-time inflator if the risk function does not depend on the price. Let I_c^N denote the discrete-time inflator in the case that $\lambda(s) = c$. Recall the definition of the continuous-time inflator $I_c : [0, T) \to \mathbb{R}$ from Lemma 4.2.2.

Lemma 4.2.11. *Assume that the risk function is constant, $\lambda(s) = c$ for all $s \in [0, \infty)$. Then I_c^N does not depend on the price and satisfies for every $t_0 \in [0, T)$*

$$\lim_{N\to\infty, t\to t_0} I_c^N(t) = I_c(t_0).$$

Lemma 4.2.12. *For $(t, s) \in [0, T) \times (0, \infty)$ and $N \in \mathbb{N}$ let $n \in N$ such that $t \in [nh^N, (n+1)h^N)$. Then we have*

$$I^N(t, s) \le 1 + \frac{(T - nh^N)^2}{\eta} \sup_{t \le r \le T} \mathbb{E}[\lambda(S_r^{t,s})].$$

In particular this implies

$$\overline{I}(t, s) \le 1 + \frac{(T - t)^2}{\eta} \sup_{t \le r \le T} \mathbb{E}[\lambda(S_r^{t,s})].$$

Proof. Let $(t,s) \in [0,T) \times (0,\infty)$, $N \in \mathbb{N}$ and $n \in N$ such that $t \in [nh^N, (n+1)h^N)$. Then we have

$$I^N(t,s) = \frac{T-t}{\eta} V^N(t,s,1) \leq \frac{T-nh^N}{\eta} V_n^N(s,1).$$

Closing the position $x=1$ at time nh^N with constant trading rate $z_k = \frac{1}{N-n}$ for $k \geq n$ is clearly suboptimal. The associated position trajectory is given by $x_k = \frac{N-k}{N-n}$ for $k \geq n$. This implies

$$\begin{aligned}
I^N(t,s) &\leq \frac{N-n}{\eta}\frac{T}{N}\mathbb{E}\left[\sum_{k=n}^{N-1} \eta^N z_k^2 + \lambda^N(S_k^N)x_k^2 \middle| S_n^N = s\right] \\
&= 1 + \frac{N-n}{\eta}\frac{T^2}{N^2}\mathbb{E}\left[\sum_{k=n}^{N-1} \lambda(S_k^N)\left(\frac{N-k}{N-n}\right)^2 \middle| S_n^N = s\right] \\
&\leq 1 + \frac{N-n}{\eta}\frac{T^2}{N^2}\mathbb{E}\left[\sum_{k=n}^{N-1} \lambda(S_k^N) \middle| S_n^N = s\right] \\
&\leq 1 + \frac{(T - \frac{n}{N}T)^2}{\eta} \sup_{n \leq k \leq N-1} \mathbb{E}\left[\lambda(S_k^N) | S_n^N = s\right] \\
&\leq 1 + \frac{(T-nh^N)^2}{\eta} \sup_{t \leq r \leq T} \mathbb{E}[\lambda(S_r^{t,s})].
\end{aligned}$$

□

Lemma 4.2.13. *Let $\lambda_1, \lambda_2 : [0,\infty) \to \mathbb{R}_+$ be two risk functions such that $\lambda_1 \leq \lambda_2$ in $[0,\infty)$. For $N \in \mathbb{N}$ let λ_1^N and λ_2^N denote the rescaled risk functions defined by* (4.35) *and let a_1^N and a_2^N denote the coefficient functions of the associated minimization problem defined by* (4.36)*, respectively. Then we have for every $n \leq N-1$ and $s \in (0,\infty)$*

$$0 \leq a_{2,n}^N(s) - a_{1,n}^N(s) \leq \frac{T}{N}\sum_{k=n}^{N-1} \mathbb{E}[\lambda_2(S_k^N) - \lambda_1(S_k^N) | S_n^N = s].$$

Proof. For $n \leq N-1$ and $s \in (0,\infty)$ let (x_k) be an ϵ-optimal strategy for $V_{2,n}^N(s,1)$. Then we have

$$a_{2,n}^N(s) - a_{1,n}^N(s) = V_{2,n}^N(s,1) - V_{1,n}^N(s,1) \geq \mathbb{E}\left[\sum_{k=n}^{N-1} (\lambda_2^N(S_k) - \lambda_1^N(S_k))x_k^2\right] - \epsilon \geq -\epsilon.$$

Letting ϵ go to 0 yields the first inequality.

Concerning the second inequality let $n \leq N-1$ and $s \in (0,\infty)$. Then we have

$$\begin{aligned}
&a_{2,n}^N(s) - a_{1,n}^N(s) \\
&\quad = \lambda_2^N(s) - \lambda_1^N(s) + \frac{(\eta^N)^2(\mathbb{E}[a_{2,n+1}^N(S_{n+1}^N) - a_{1,n+1}^N(S_{n+1}^N)|S_n^N = s]}{(\eta^N + \mathbb{E}[a_{1,n+1}^N(S_{n+1}^N)|S_n^N = s])(\eta^N + \mathbb{E}[a_{2,n+1}^N(S_{n+1}^N)|S_n^N = s])}.
\end{aligned}$$

Then the first inequality and the fact that $a^N_{1,n+1}, a^N_{2,n+1} \geq 0$ imply

$$a^N_{2,n}(s) - a^N_{1,n}(s) \leq \frac{T}{N}(\lambda_2(s) - \lambda_1(s)) + \mathbb{E}[a^N_{2,n+1}(S^N_{n+1}) - a^N_{1,n+1}(S^N_{n+1})|S^N_n = s].$$

Now a simple induction yields the claim. □

Proof of Theorem 4.2.9. We divide the proof into 6 claims. In Claim I and II we prove that $\underline{I}$ is a viscosity supersolution to (4.20). Claim III delivers the main argument why the same reasoning as in Claims I and II can be applied to verify the subsolution property of $\overline{I}$. In Claims IV-VI we show that $\underline{I}$ as well as $\overline{I}$ satisfy the boundary conditions (4.22) and are bounded from below by one. Then Proposition 4.2.8 implies that we have $\overline{I} = \underline{I}$.

Let us start by verifying the viscosity supersolution property of $\underline{I}$. To this end fix a point $(t_0, s_0) \in [0,T) \times (0,\infty)$ and a function $\varphi \in C^{1,2}([0,T) \times (0,\infty))$ such that $\varphi(t_0,s_0) = \underline{I}(t_0,s_0)$ and $\varphi \leq \underline{I}$ in $[0,T) \times (0,\infty)$. Without loss of generality we assume that (t_0,s_0) is a strict global minimum of $\underline{I} - \varphi$. Moreover, we assume that $\varphi = -c$ for some constant $c > 0$ in the complement of a bounded neighborhood $\mathcal{U}$ of (t_0,s_0).

Claim I. *There exists a sequence* $(N_i, t_i, s_i) \in \mathbb{N} \times [0,T) \times (0,\infty)$ *such that* $\lim_{i\to\infty}(N_i,t_i,s_i) = (\infty, t_0, s_0)$*,* $\lim_{i\to\infty} I^{N_i}(t_i,s_i) = \underline{I}(t_0,s_0)$ *and that* (t_i,s_i) *is a global minimum of* $I^{N_i} - \varphi$ *for all* $i \in \mathbb{N}$*.*

Since $\underline{I}$ is nonnegative we have for all $(t,s) \in \mathcal{U}^c$ and $N \in \mathbb{N}$

$$I^N(t,s) - \varphi(t,s) = I^N(t,s) + c \geq c.$$

By the very definition of $\underline{I}$ there exists a sequence $(N_i, t'_i, s'_i) \in \mathbb{N} \times [0,T) \times (0,\infty)$ such that $\lim_{i\to\infty}(N_i,t'_i,s'_i) = (\infty,t_0,s_0)$ and $\lim_{i\to\infty} I^{N_i}(t'_i,s'_i) = \underline{I}(t_0,s_0)$. In particular, we have $\lim_{i\to\infty} I^{N_i}(t'_i,s'_i) - \varphi(t'_i,s'_i) = 0$. Hence, for i large enough, we have

$$\inf_{[0,T)\times(0,\infty)} I^{N_i} - \varphi = \min_{\bar{\mathcal{U}}} I^{N_i} - \varphi = I^{N_i}(t_i,s_i) - \varphi(t_i,s_i)$$

for a sequence $(t_i,s_i) \in \bar{\mathcal{U}}$. By compactness of $\bar{\mathcal{U}}$ there exists a converging subsequence, also denoted by (t_i,s_i), such that $\lim_{i\to\infty}(t_i,s_i) = (\bar{t},\bar{s})$. This implies

$$\begin{aligned}
\underline{I}(\bar{t},\bar{s}) - \varphi(\bar{t},\bar{s}) &\leq \liminf_{i\to\infty} I^{N_i}(t_i,s_i) - \varphi(t_i,s_i) \\
&= \liminf_{i\to\infty} \inf_{[0,T)\times(0,\infty)} I^{N_i} - \varphi \\
&\leq \liminf_{i\to\infty} I^{N_i}(t'_i,s'_i) - \varphi(t'_i,s'_i) \\
&= \underline{I}(t_0,s_0) - \varphi(t_0,s_0).
\end{aligned}$$

Since (t_0,s_0) is a strict global minimum of $\underline{I} - \varphi$ this implies $(\bar{t},\bar{s}) = (t_0,s_0)$ and $\lim_{i\to\infty} I^{N_i}(t_i,s_i) = \varphi(t_0,s_0)$.

Claim II. *At* (t_0,s_0) *the function* φ *satisfies*

$$0 \leq -\varphi_t - \frac{1}{2}\sigma^2\varphi_{ss} - \frac{\varphi}{T-t_0} + \frac{\varphi^2}{T-t_0} - (T-t_0)\frac{\lambda(s_0)}{\eta}.$$

Let (N_i, t_i, s_i) denote the sequence from Claim I. We introduce the sequence $h_i = T/N_i$ for i large enough such that $t_i + h_i \leq T$. Then we set

$$\rho_i = \mathbb{E}[\varphi(t_i + h_i, S_{t_i+h_i})|S_{t_i} = s_i] - \varphi(t_i, s_i).$$

Since φ is constant outside of $\mathcal{U}$ the derivatives φ_s, φ_{ss} and φ_t are bounded. Itô's Formula implies

$$\rho_i = \mathbb{E}\left[\int_{t_i}^{t_i+h_i} \varphi_t(r, S_r) + \frac{1}{2}\sigma^2(S_r)\varphi_{ss}(s, S_r)dr \middle| S_{t_i} = s_i\right].$$

Appealing to the mean value theorem and Lebesgue's theorem of dominated convergence yields

$$\lim_{i\to\infty} \frac{\rho_i}{h_i} = \varphi_t(t_0, S_0) + \frac{1}{2}\sigma^2(s_0)\varphi_{ss}(t_0, s_0)$$

In particular we have $\rho_i \to 0$ as $i \to \infty$. By Claim I we have

$$\begin{aligned} \mathbb{E}[I^{N_i}(t_i + h_i, S_{t_i+h_i})|S_{t_i} = s_i] &\geq \mathbb{E}[\varphi(t_i + h_i, S_{t_i+h_i})|S_{t_i} = s_i] + I^{N_i}(t_i, S_i) - \varphi(t_i, s_i) \\ &= \rho_i + I^{N_i}(t_i, S_i). \end{aligned}$$

Note that the map

$$\mathbb{R} \ni x \mapsto \frac{x}{T - t - h_i + h_i x}$$

is nondecreasing, since we have $t_i + h_i \leq T$. Equation (4.37) yields

$$I^{N_i}(t_i, s_i) \geq (T - t_i)\left(\frac{\rho_i + I^{N_i}(t_i, s_i)}{T - t_i - h_i + h_i(\rho_i + I^{N_i}(t_i, s_i))} + h_i\frac{\lambda(s_i)}{\eta}\right).$$

Since $\rho_i + I^{N_i}(t_i, S_i)$ is positive for i large enough, muliplying by

$$\frac{(T - t_i - h_i)\eta + h_i(\rho_i + I^{N_i}(t_i, s_i))}{(T - t_i)h_i\eta}$$

yields

$$\begin{aligned} 0 \leq -\frac{\rho_i}{h_i} - \frac{I^{N_i}(t_i, s_i)}{T - t_i} + \frac{I^{N_i}(t_i, s_i)^2}{T - t_i} - (T - t_i - h_i)\frac{\lambda(s_i)}{\eta} \\ - h_i\frac{\lambda(s_i)}{\eta}(I^{N_i}(t_i, s_i) + \rho_i) + \rho_i\frac{I^{N_i}(t_i, s_i)}{T - t_i} \end{aligned}$$

Letting i tend to ∞ then yields the claim

$$0 \leq -\varphi_t - \frac{1}{2}\sigma^2\varphi_{ss} - \frac{\varphi}{T - t_0} + \frac{\varphi^2}{T - t_0} - (T - t_0)\frac{\lambda(s_0)}{\eta}.$$

The proof of the subsolution property goes along the lines of Claim I and II. There is essentially only one difference: While the proof of Claim I relied on the fact that I^N is nonnegative, I^N is in general not bounded from above by a constant independent of N. But as Claim III will show, there exists a polynomial $\psi : (0,\infty) \to \mathbb{R}$ such that $I^N(t,s) \le \psi(s)$ for all $(t,s) \in [0,T) \times (0,\infty)$ and $N \in \mathbb{N}$. Now, to verify the subsolution property of $\overline{I}$ at a point (t_0, s_0) one can assume without loss of generality that the test function is equal to $\psi + c$ for some $c > 0$ outside of some bounded neighborhood $\mathcal{U}$ of (t_0, s_0). Then the same reasoning as in Claim I and Claim II yields the subsolution property.

Claim III. *There exists a polynomial* $\psi : (0,\infty) \to \mathbb{R}$ *such that* $I^N(t,s) \le \psi(s)$ *for all* $(s,t) \in [0,T) \times (0,\infty)$ *and* $N \in \mathbb{N}$.

By Lemma 4.2.12 we have for every $(t,s) \in [0,T) \times (0,\infty)$

$$I^N(t,s) \le 1 + \frac{T^2}{\eta} \sup_{t \le r \le T} \mathbb{E}[\lambda(S_r^{t,s})].$$

Using the polynomial growth of λ and appealing to the moment estimate (4.21) implies

$$\begin{aligned} I^N(t,s) &\le 1 + \frac{T^2}{\eta} C(1 + \sup_{t \le r \le T} \mathbb{E}[(S_r^{t,s})^p] \\ &\le 1 + \frac{T^2}{\eta} C(1 + K(1 + s^p)). \end{aligned}$$

Let us now turn to the boundary conditions.

Claim IV. *For all* $t_0 \in [0,T)$ *we have*

$$\lim_{\substack{t \to t_0 \\ s \searrow 0}} \underline{I}(t,s) = \lim_{\substack{t \to t_0 \\ s \searrow 0}} \overline{I}(t,s) = I_{\lambda(0)}(t_0).$$

We only consider the case $\lambda(0) = \max_{s \ge 0} \lambda(s)$ here, because the proof of the other case relies on the identical arguments. Note that Lemma 4.2.11 implies that $I_{\lambda(0)}(t_0) = \lim_{N \to \infty, t \to t_0} I^N_{\lambda(0)}(t)$. Lemma 4.2.13 implies $I^N(t,s) \le I^N_{\lambda(0)}(t)$. Hence, we have

$$\limsup_{\substack{t \to t_0 \\ s \searrow 0}} \underline{I}(t,s) \le \limsup_{\substack{t \to t_0 \\ s \searrow 0}} \overline{I}(t,s) \le I_{\lambda(0)}(t_0).$$

Appealing to Lemma 4.2.13 and Lebesgue's theorem of dominated convergence yields

$$\begin{aligned} I_{\lambda(0)}(t) - \underline{I}(t,s) &= \liminf_{\substack{N \to \infty \\ (t',s') \to (t,s)}} I^N_{\lambda(0)}(t') - I^N(t',s') \\ &\le \frac{T-t}{\eta} \mathbb{E}\left[\int_t^T \lambda(0) - \lambda(S_r^{t,s}) dr \right]. \end{aligned}$$

Taking the limit $t \to t_0$ and $s \searrow 0$ gives

$$\liminf_{\substack{t\to t_0\\ s\searrow 0}} \overline{I}(t,s) \geq \liminf_{\substack{t\to t_0\\ s\searrow 0}} \underline{I}(t,s) \geq I_{\lambda(0)}(t_0),$$

and hence the result.

Claim V. *$\underline{I}$ and $\overline{I}$ are bounded from below by one and have polynomial growth in s.*

Again we appeal to Lemma 4.2.11 and note that $1 = I_0(t_0) = \lim_{N\to\infty, t\to t_0} I_0^N(t)$. Since λ is nonnegative, Lemma 4.2.13 implies for all $(t,s) \in [0,T) \times (0,\infty)$

$$\eta \leq \underline{I}(t,s) \leq \overline{I}(t,s).$$

The polynomial growth of $\underline{I}$ and $\underline{I}$ is a direct consequence of Claim III.

Claim VI. *For all $s_0 > 0$ we have*

$$\lim_{\substack{t\nearrow T\\ s\to s_0}} \underline{I}(t,s) = \lim_{\substack{t\nearrow T\\ s\to s_0}} \overline{I}(t,s) = 1$$

Since we have $\overline{I} \geq \underline{I} \geq 1$ on $[0,T) \times (0,\infty)$ by Claim V it suffices to show that

$$\lim_{\substack{t\nearrow T\\ s\to s_0}} \overline{I}(t,s) \leq 1.$$

But this is a direct consequence of Lemma 4.2.12. □

4.3. Numerical approximation

In general neither the discrete-time nor the continuous-time liquidation problem presented in Sections 4.1 and 4.2 admit an explicit solution. Hence, the coefficient functions a_k of the discrete-time value function or the continuous-time inflator I have to be approximated numerically. In this section we provide a numerical scheme that approximates the functions a_k. If we choose a sufficiently large number of time steps in the discrete time model it follows from Theorem 4.2.9 that these functions also provide an approximation of the continuous-time inflator.

Recall the backward recursion scheme (4.8) defining the coefficient functions a_k

$$\begin{aligned} a_{N-1}(s) &= \eta(s) + \lambda(s) \\ a_k(s) &= \eta(s)(\eta(s) + Ta_{k+1}(s))^{-1} Ta_{k+1}(s) + \lambda(s), \end{aligned} \tag{4.40}$$

where T denotes the expectation operator $Tf(s) = \mathbb{E}[f(S_{k+1})|S_k = s]$. The challenge is to determine the function Ta_k in every time step k. A natural approximation is to calculate values of a_{k+1} at finitely many points, to interpolate linearly and then to apply the trapezoidal rule. In Subsection 4.3.1 we describe such a piecewise linear approximation in more detail. To simplify the analysis, we consider only the case $d = 1$. Recall that Subsection 4.1.2 shows how a multi-asset liquidation problem can be broken down into several one-dimensional problems. In Subsection 4.3.2 we analyze how function approximation errors propagate through the recursion (4.40). We will see that *positive* operators are benign, guaranteeing that the total error is bounded against the sum of the approximation error made in every recursion step. In Section 4.3.3 we prove error bounds of the linear approximation in dependence of the grid size and the grid endpoints. We show that the error is essentially of order $\mathcal{O}(1/n^2)$, where n is the number of grid points.

4.3.1. Piecewise linear approximation

Given a function $f \in \mathcal{L}^1$ (see 4.1.2 for the definition of $\mathcal{L}^1$) we approximate the expectation Tf by interpolating f piecewise linearly on a finite partition of $\mathbb{R}_{>0}$ with logarithmic scale. Choose a minimal and maximal price $0 < s^{\min} < s^{\max} < \infty$, a grid size $n \geq 1$ and define

$$\Delta = \left\{ s_i = \left(\frac{s^{\max}}{s^{\min}} \right)^{\frac{i}{n}} s^{\min} \middle| i = 0, \dots, n \right\}. \tag{4.41}$$

Let Lf be the piecewise linear function that equals f at each point of Δ. On $\mathbb{R} \setminus [s^{\min}, s^{\max}]$ we set $Lf = 0$. The logscale mesh size is given by $\kappa = (s^{\max}/s^{\min})^{1/n}$. Note that there exist $m_i, \zeta_i \in \mathbb{R}$ such that $Lf(s) = \sum_{i=0}^{n-1} (m_i s + \zeta_i) \mathbf{1}_{\{s_i \leq s \leq s_{i+1}\}}$. Then we have

$$TLf(s_j) = \sum_{i=0}^{n-1} m_i s_j \mathbb{E}[I \mathbf{1}_{\{\kappa^{i-j} \leq I \leq \kappa^{i-j+1}\}}] + \zeta_i \mathbb{P}[\kappa^{i-j} \leq I \leq \kappa^{i-j+1}], \tag{4.42}$$

where I is the multiplicative price increment (see Section 4.1.1). All information we require for the recursion can be stored in advance in the vectors $(\mathbb{E}[I \mathbf{1}_{\{\kappa^l \leq I \leq \kappa^{l+1}\}}])_{-n \leq l \leq n-1}$ and $(\mathbb{P}[\kappa^l \leq I \leq \kappa^{l+1}])_{-n \leq l \leq n-1}$. No integrals need to be computed in the recursion, which implies that the numerical effort of the algorithm is proportional to the number of grid points n.

4.3.2. Error propagation

Define an operator F acting on the set of nonnegative functions on $(0, \infty)$ by

$$Ff(s) = \frac{\eta(s) f(s)}{\eta(s) + f(s)}.$$

Then the recurrence relation (4.40) may be rewritten as

$$\begin{aligned} a_{N-1} &= \eta + \lambda \\ a_k &= FTa_{k+1} + \lambda \end{aligned}$$

The challenge is to calculate Ta_k in every time step. In most examples, including geometric Brownian motion, there is no explicit representation available. This is why one has to approximate T by an operator $\widehat{T}$, acting again on $\mathcal{L}^1$. A common choice is $\widehat{T} = TL$, where L is the piecewise linear approximation operator introduced in Subsection 4.3.1. Alternatively, one can approximate functions with polynomials. Let P be an operator that assigns to any function f a polynomial Pf, and set $\widehat{T} = TP$. A polynomial approximation has the drawback that it is not a positive operator and hence neither $\widehat{T} = TP$. As shown in the next lemma, however, positivity of $\widehat{T}$ is crucial to control the error propagation of the recurrence relation. This is why we opt for the piecewise linear interpolation in Section 4.3.3.

After having chosen a set of approximating operators $(\widehat{T}_k)_{k\leq N-2}$ we set up the following approximation scheme:

$$\begin{aligned} \hat{a}_{N-1} &= \eta + \lambda \\ \hat{a}_k &= F\widehat{T}_k\hat{a}_{k+1} + \lambda \end{aligned} \tag{4.43}$$

Note that we have $Ff(s) \leq \eta(s)$ for $f \geq 0$, which implies $\hat{a}_k \in \mathcal{L}^1$. Hence, the approximation scheme is well-defined if T_k is positive for every $k \leq N-2$. Moreover, F is pointwise nonexpansive, since for $f, g \geq 0$ and $s \in \mathbb{R}_{>0}$ we have

$$|Ff(s) - Fg(s)| = \left| \frac{\eta^2(s)(f(s) - g(s))}{(\eta(s) + f(s))(\eta(s) + g(s))} \right| \leq |f(s) - g(s)|. \tag{4.44}$$

This leads to the following pointwise error estimate.

Lemma 4.3.1. *Suppose that $\widehat{T}_k$ is a positive operator for every $k \leq N-1$. Then the approximation error at time $t_k \leq t_{N-1}$ is bounded by*

$$\begin{aligned} |a_k(s) - \hat{a}_k(s)| &\leq \sum_{l=k}^{N-2} T^{l-k}|(T - \widehat{T}_l)\hat{a}_{l+1}|(s) && (4.45) \\ &= \sum_{l=k}^{N-2} \|(T - \widehat{T}_l)\hat{a}_{l+1}\|_{L^1(\mu_s^{l-k})}. && (4.46) \end{aligned}$$

for every $s \in \mathbb{R}_{>0}$.

The previous lemma guarantees that in order to verify pointwise convergence of the approximation scheme at $s \in \mathbb{R}_{>0}$, it suffices to specify subsets $\mathcal{D}_{s,k} \subset \mathcal{L}^1$ for $k \leq N-2$ such that $\hat{a}_l \in \mathcal{D}_{s,k}$ for all $l > k$ and $\widehat{T}_l f \to Tf$ in $L^1(\mu_s^k)$ for every $f \in \mathcal{D}_{s,k}$ and $l \geq k$.

Proof of Lemma 4.3.1. To prove (4.45) we proceed via induction. The case $k = N-1$ is clear, since $a_{N-1} = \hat{a}_{N-1}$. For $k \leq N-2$ Equation (4.44) and the induction hypothesis imply

$$\begin{aligned} |a_k(s) - \hat{a}_k(s)| &\leq |Ta_{k+1}(s) - \widehat{T}_k\hat{a}_{k+1}(s)| \\ &\leq T|a_{k+1} - \hat{a}_{k+1}|(s) + |(T - \widehat{T}_k)\hat{a}_{k+1}|(s) \\ &\leq \sum_{l=k}^{N-2} T^{l-k}|(T - \widehat{T}_l)\hat{a}_{l+1}|(s). \end{aligned}$$

□

4.3.3. Error bounds for piecewise linear interpolations

In this section we provide error bounds for the piecewise linear approximation described in Section 4.3.1. A linear interpolation operator is positive, and hence, by Lemma 4.3.1, its approximation error does not explode when transmitted through the function recursion.

The error bounds depend on the grid size and the grid endpoints. We show that the error is essentially of order $\mathcal{O}(1/n^2)$, where n is the number of grid points. In particular we obtain that the corresponding approximation scheme converges.

We use notation of Subsection 4.3.1. In particular we refer to Δ as the set of discretization points defined in (4.41). Moreover, let L be the piecewise linear interpolation operator on Δ and let κ be the logscale mesh size. Whenever we want to stress the dependence on n, the number of grid points, we write $\Delta = \Delta_n$, $L_n = L$ and $\kappa_n = \kappa$. Recall that the expectation of a piecewise linear function satisfies (4.42).

For $k = 0, \dots, N-2$ we approximate T by $\widehat{T}_k = TL_{n(k)}$ with $n(k) \in \mathbb{N}$. To derive error bounds of this approximation we define the function space

$$\mathcal{D}_{s,k} := \left\{ f \in L^1(\mu_s^k) \cap C^2(\mathbb{R}_{>0}) \,\middle|\, \int_0^\infty x^2 |f''(x)| \mu_s^k(dx) < \infty \right\}$$

for $k \leq N-2$ and $s \in \mathbb{R}_{>0}$. We suppose that the multiplicative increments I_k have a twice differentiable density, denoted by ψ, and introduce the functions

$$\begin{aligned} \Phi_1(x) &:= 1 + x\frac{\psi'(x)}{\psi(x)} \\ \Phi_2(x) &:= 2 + 4x\frac{\psi'(x)}{\psi(x)} + x^2\frac{\psi''(x)}{\psi(x)}. \end{aligned}$$

Let us state a list of conditions on λ and μ_s^k, which will turn out to be sufficient for the convergence of the approximation scheme.

I The density function satisfies $\psi \in C^2(\mathbb{R}_{>0})$.

II Φ_1 belongs to $L^4(\mu_1^1)$, while $\Phi_2 \in L^2(\mu_1^1)$.

III The risk function $\lambda \geq 0$ belongs to $\mathcal{D}_{s,k}$ for all $k \leq N-2$.

IV There exists $C \in \mathbb{R}_+$ such that for any choice of $0 < s^{\min} < s^{\max}$ and $n \in \mathbb{N}$ the associated interpolation operator L_n satisfies

$$\|L_n\lambda\|_{L^4(\mu_s^k)} < C \qquad \text{for all } k \leq N-1.$$

V For every $k \leq N-2$ we have $\eta \in L^1(\mu_s^k)$.

In the case of a relative price impact, we further assume

VI The function $\operatorname{inv}(x) = 1/x$ belongs to $L^2(\mu_s^k)$ and μ_s^k admits finite fourth momemts for all $k \leq N-2$.

Remark 4.3.2. Let us take a closer look at Condition **IV**. If λ is bounded, then $L_n\lambda$ is bounded uniformly in n. Hence, **IV** is satisfied.

If λ is concave, then $0 \leq L_n\lambda \leq \lambda$. Hence, **IV** is met, if $\lambda \in L^4(\mu_s^k)$ for all $k \leq N-1$.

If λ grows polynomially of order $p \in \mathbb{N}$, i.e. there exist $c_0, c_1 > 0$ such that $\lambda(x) \leq c_1 x^p + c_0$ for all $x \in \mathbb{R}_{>0}$, then $L_n\lambda(x) \leq c_1\kappa_n^p x^p + c_0$. Thus, it suffices to require that μ_k^s admits finite moments of order $4p$ and that n, $s^{\min}$ and $s^{\max}$ are chosen such that κ_n is bounded.

Example 4.3.3. *Let (S_k) be a discrete-time geometric Brownian motion and λ the risk function given by* (4.3) *resp.* (4.4). *Then Assumptions **I**-**VI** are satisfied.*

Indeed, notice that we have

$$\psi(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln(x) + \frac{1}{2}\sigma^2)^2}{2\sigma^2}\right).$$

*Hence, Condition **VI** is met. Moreover, **I** is satisfied with*

$$\begin{aligned}
\psi'(x) &= -\frac{1}{x}\left(\frac{3}{2} + \frac{\log x}{\sigma^2}\right)\psi(x) \\
\psi''(x) &= -\frac{1}{x^2}\left(4\frac{1}{4} + \frac{4\log(x) - 1}{\sigma^2} + \frac{\log(x)^2}{\sigma^4}\right)\psi(x).
\end{aligned}$$

This yields

$$\begin{aligned}
\Phi_1(x) &= -\frac{1}{2} - \frac{\log(x)}{\sigma^2} \\
\Phi_2(x) &= -\frac{1}{4} - \frac{1}{\sigma^2} + \frac{\log(x)^2}{\sigma^4}.
\end{aligned}$$

*Since the normal distribution has finite fourth moments, we have $\int_{\mathbb{R}_{>0}} \log(x)^4 \mu_1^1(dx) < \infty$, which implies Condition **II**.*

*Notice that $\lambda \notin C^2(\mathbb{R}_{>0})$ and hence Condition **III** is not satisfied. We can, however, relax this assumption by requiring that $\lambda \in C^2(\mathbb{R}_{>0} \setminus \Delta_n)$ and that λ admits left- and right-hand second derivatives at every sampling point of Δ_n for all $n \in \mathbb{N}$. For $\lambda(s) = \max(0, as + \bar{s})^2$ we therefore have to require that $\bar{s}/a \in \Delta_n$ for all $n \in \mathbb{N}$.*

*Finally, Conditions **IV** and **V** are easily verified.*

We fix a price $s \in \mathbb{R}_{>0}$ for the rest of this subsection. Due to the special form of $\widehat{T}_k = TL_{n(k)}$, the error estimate of Lemma 4.3.1 becomes

$$|a_k(s) - \hat{a}_k(s)| \quad \leq \quad \sum_{l=k}^{N-2} \|(\operatorname{Id} - L_{n(l)})\hat{a}_{l+1}\|_{L^1(\mu_s^{l-k+1})},$$

We fix $l \in \{0, \dots, N-2\}$ and simply write n instead of $n(l)$ in the sequel. Then it suffices to find error bounds of

$$\|(\mathrm{Id} - L_n)\hat{a}_{l+1}\|_{L^1(\mu_s^k)} = \int_{\mathbb{R}\setminus[s^{\min}, s^{\max}]} |\hat{a}_{l+1}| d\mu_s^k + \int_{[s^{\min}, s^{\max}]} |(\mathrm{Id} - L_n)\hat{a}_{l+1}| d\mu_s^k \quad (4.47)$$

for $k = 1, \dots, N-1$. Since $0 \le \hat{a}_{l+1} \le \eta + \lambda$ and $\eta, \lambda \in L^1(\mu_s^k)$ by Condition **V** the error

$$\int_{\mathbb{R}\setminus[s^{\min}, s^{\max}]} |\hat{a}_{l+1}| d\mu_s^k \le \int_{\mathbb{R}\setminus[s^{\min}, s^{\max}]} \eta + \lambda d\mu_s^k$$

becomes arbitrarily small (independently of $\hat{a}_{l+1}$), if we choose $[s^{\min}, s^{\max}]$ large enough. The next proposition yields that the second summand of (4.47) decays with the order $1/n^2$, again independently of $\hat{a}_{l+1}$.

Proposition 4.3.4. *Assume that Conditions **I**-**V** are satisfied. If $\eta(s) = \eta s$, suppose that **VI** is met as well. Then for every $k \le N-2$ there exists a constant $c_k > 0$ such that*

$$\limsup_{n\to\infty} \quad n^2 \int_{[s^{\min}, s^{\max}]} |(\mathrm{Id} - L_n)\hat{a}_{l+1}| d\mu_s^k \le c_k \log\left(\frac{s^{\max}}{s^{\min}}\right).$$

Note that Proposition 4.3.4 implies the convergence of the linear approximation algorithm. For the proof we need the following two lemmata.

Lemma 4.3.5. *Let $f \in \mathcal{D}_{s,k}$ and $0 < s^{\min} < s^{\max} < \infty$. Then $L_n f$ converges to f in $L^1(\mu_s^k, [s^{\min}, s^{\max}])$ as $n \to \infty$. More precisely, we have*

$$\limsup_{n\to\infty} \quad n^2 \int_{s^{\min}}^{s^{\max}} |(\mathrm{Id} - L_n) f| d\mu_s^k \le \log\left(\frac{s^{\max}}{s^{\min}}\right) \int_0^\infty x^2 |f''(x)| \mu_s^k(dx).$$

Proof. We denote by $\{s_i^n\}_{i=0,\dots,n}$ the grid points of Δ_n. Since $f \in C^2(\mathbb{R}_{>0})$ the error estimate for piecewise linear interpolation yields for $x \in [s^{\min}, s^{\max}]$

$$\begin{aligned} |L_n f(x) - f(x)| &\le \frac{1}{8}\sum_{i=0}^{n-1} (s_{i+1}^n - s_i^n)^2 |f''(\xi_i^n)| \mathbf{1}_{\{s_i^n \le x \le s_{i+1}^n\}} \\ &= \frac{1}{8}(\kappa_n - 1)\sum_{i=0}^{n-1} (s_i^n)^2 |f''(\xi_i^n)| \mathbf{1}_{\{s_i^n \le x \le s_{i+1}^n\}} \end{aligned}$$

with $\xi_i^n \in [s_i^n, s_{i+1}^n]$. This implies

$$\int_{s^{\min}}^{s^{\max}} |L_n f(x) - f(x)| \mu_s^k(dx) \le \frac{1}{8}(\kappa_n - 1)\sum_{i=0}^{n-1} (s_i^n)^2 |f''(\xi_i^n)| \mu_s^k([s_i^n, s_{i+1}^n]).$$

Since f'' is bounded on $[s^{\min}, s^{\max}]$, we have

$$\sum_{i=0}^{n-1} (s_i^n)^2 |f''(\xi_i^n)| \mu_s^k([s_i^n, s_{i+1}^n]) \to \int_{s^{\min}}^{s^{\max}} x^2 |f''(x)| \mu_s^k(dx) \le \int_0^\infty x^2 |f''(x)| \mu_s^k(dx)$$

as $n \to \infty$. Using $\lim_{n\to\infty} n^2(\kappa_n - 1) = \log\left(\frac{s^{\max}}{s^{\min}}\right)$ we obtain the desired result. □

For the second lemma we introduce the function space

$$\mathcal{C} := \{f \in \mathcal{L} | 0 \leq f \leq \eta + \lambda\}.$$

Lemma 4.3.6. *Assume that Conditions **I**-**V** are satisfied. If $\eta(s) = \eta s$, suppose that **VI** is met as well. Then, $F\widehat{T}_l f + \lambda \in \mathcal{C} \cap \mathcal{D}_{s,k}$ for every $f \in \mathcal{C}$, $l \leq N-2$ and $k \leq N-2$. Moreover, for every $k \leq N-2$ there exists a constant $c_k > 0$ such that for every $f \in \mathcal{C}$ we have*

$$\int_0^\infty x^2 (F\widehat{T}_l f + \lambda)''(x) \mu_s^k(dx) \leq c_k. \tag{4.48}$$

Proof. Let $f \in \mathcal{C}$ and $k, l \leq N-2$. Since $\widehat{T}_l$ is a positive operator, we have $\widehat{T}_l f \geq 0$. This implies $0 \leq F\widehat{T}_l f + \lambda \leq \eta + \lambda$ and $F\widehat{T}_l f + \lambda \in L^1(\mu_s^k) \cap \mathcal{L}$ by Condition **V**. Note that for $g \in C^2(\mathbb{R}_{>0})$ with $g \geq 0$ we have $Fg \in C^2(\mathbb{R}_{>0})$ with

$$(Fg)''(x) = \eta^2 \left(\frac{g''(x)}{(g(x)+\eta)^2} - \frac{2g'(x)^2}{(g(x)+\eta)^3} \right) \tag{4.49}$$

in the case of absolute price impact ($\eta(s) = \eta$) and

$$(Fg)''(x) = \eta^2 \left(\frac{x^2 g''(x)}{(g(x)+\eta x)^2} - \frac{(xg'(x)-g(x))^2}{(g(x)+\eta)^3} \right) \tag{4.50}$$

in the case of relative price impact ($\eta(s) = \eta s$). Hence, in order to verify that $F\widehat{T}_l f + \lambda$ is twice continuously differentiable, it remains to show that $\widehat{T}_l f \in C^2(\mathbb{R}_{>0})$. Note that

$$\widehat{T}_l f(r) = \mathbb{E}[L_n f(rI_1)] = \frac{1}{r} \int_0^\infty L_n f(x) \psi\left(\frac{x}{r}\right) dx.$$

Using that $L_n f$ is bounded on $\mathbb{R}_{>0}$ by construction of L_n and that $\Phi_1 \in L^1(\mu_1^1)$ by Condition **II** we get

$$\begin{aligned}
(\widehat{T}_l f)'(r) &= -\frac{1}{r^2} \int_0^\infty L_n f(x) \psi\left(\frac{x}{r}\right) dx - \frac{1}{r^3} \int_0^\infty L_n f(x) x \psi'\left(\frac{x}{r}\right) dx && (4.51)\\
&= -\frac{1}{r} \mathbb{E}[\Phi_1(I_1) L_n f(rI_1)]. && (4.52)
\end{aligned}$$

A similar argumentation using $\Phi_2 \in L^1(\mu_1^1)$ yields

$$(\widehat{T}_l f)''(r) = \frac{1}{r^2} \mathbb{E}[\Phi_2(I_1) L_n f(rI_1)]. \tag{4.53}$$

Thus, $F\widehat{T}_l f \in C^2(\mathbb{R}_{>0})$ and it remains to show that

$$\int_0^\infty x^2 (F\widehat{T}_l f + \lambda)''(x) \mu_s^k(dx) \leq c_k$$

for some $c_k \geq 0$. Since $\lambda'' \in \mathcal{D}_{s,k}$ by assumption **III**, it suffices to consider $(F\widehat{T}_l f)''$. By Equations (4.49), (4.51) and (4.53) we have in the case $\eta(s) = \eta$

$$r^2(F\widehat{T}_l f)''(r) = \eta^2 \left(\frac{\mathbb{E}[\Phi_2(I_1)L_n f(rI_1)]}{(\widehat{T}_l f(r) + \eta)^2} - 2\frac{\mathbb{E}[\Phi_1(I_1)L_n f(rI_1)]^2}{(\widehat{T}_l f(r) + \eta)^3} \right).$$

Since $\widehat{T}_l$ is a positive operator and (I_k) are iid we get

$$|r^2(F\widehat{T}_l f)''(r)| \leq |\mathbb{E}[\Phi_2(I_{k+1})L_n f(rI_{k+1})]| + \frac{2}{\eta}\mathbb{E}[\Phi_1(I_{k+1})L_n f(rI_{k+1})]^2.$$

Using the definition of μ_s^k and applying Jensen's and Hölder's inequality yields

$$\begin{aligned}
&\int_0^\infty x^2(F\widehat{T}_l f)''(x)\mu_s^k(dx) \\
&\qquad \leq \mathbb{E}\left[|\Phi_2(I_{k+1})|L_n f(S_{k+1}) + 2\eta^{-1}\Phi_1(I_{k+1})^2 L_n f(S_{k+1})^2 | S_0 = s\right] \\
&\qquad\qquad \leq \|\Phi_2\|_{L^2(\mu_1^1)}\|L_n f\|_{L^2(\mu_s^{k+1})} + 2\eta^{-1}\|\Phi_1\|_{L^4(\mu_1^1)}\|L_n f\|_{L^4(\mu_s^{k+1})}.
\end{aligned}$$

Since L_n is monotone and equal to the identity on linear functions, we have $L_n f \leq \eta + L_n\lambda$. Hence, Conditions **II** and **IV** imply (4.48).

In the case $\eta(s) = \eta s$ similar considerations yield

$$|r^2(F\widehat{T}_l f)''(r)| \leq |\mathbb{E}[\Phi_2(I_{k+1})L_n f(rI_{k+1})]| + \frac{1}{\eta^3 r}\mathbb{E}[(\Phi_1(I_{k+1}) + 1)L_n f(rI_{k+1})]^2.$$

and

$$\begin{aligned}
&\int_0^\infty x^2(F\widehat{T}_l f)''(x)\mu_s^k(dx) \\
&\qquad \leq \mathbb{E}\left[|\Phi_2(I_{k+1})|L_n f(S_{k+1}) + \eta^{-3}S_k^{-1}((\Phi_1(I_{k+1}) + 1)^2 L_n f(S_{k+1})^2 | S_0 = s\right] \\
&\qquad \leq \|\Phi_2\|_{L^2(\mu_1^1)}\|L_n f\|_{L^2(\mu_s^{k+1})} + \eta^{-3}\sqrt{\mathbb{E}[S_k^{-2}(\Phi_1(I_{k+1}) + 1)^4]}\|L_n f\|_{L^4(\mu_s^{k+1})} \\
&\qquad = \|\Phi_2\|_{L^2(\mu_1^1)}\|L_n f\|_{L^2(\mu_s^{k+1})} + \eta^{-3}\|\operatorname{inv}\|_{L^2(\mu_s^k)}\|\Phi_1 + 1\|_{L^4(\mu_1^1)}\|L_n f\|_{L^4(\mu_s^{k+1})},
\end{aligned}$$

which yields (4.48) by Conditions **II**,**IV** and **VI**. □

Proof of Proposition 4.3.4. Since we have $a_{N-1} \in \mathcal{C}$, Lemma 4.3.6 implies that $\hat{a}_l \in \mathcal{D}_{s,k}$ for every $l, k \leq N-2$. Moreover, it yields that there exist constants $c_k > 0$ such that for every $l \leq N-2$ we have

$$\int_0^\infty x^2 \hat{a}_l''(x)\mu_s^k(dx) \leq c_k.$$

Then Lemma 4.3.5 implies

$$\limsup_{n\to\infty} \quad n^2 \int_{[s^{\min}, s^{\max}]} |(\operatorname{Id} - L_n)\hat{a}_{l+1}| d\mu_s^k \leq c_k \log\left(\frac{s^{\max}}{s^{\min}}\right).$$

□

4.4. Numerical experiments

For risk management purposes energy producers usually sell most, if not all, of the electricity they plan to generate on forward markets. Simultaneously they buy on forward markets the commodities they need for producing the electricity, e.g. coal and natural gas. Power markets in Europe are split into regional areas, all of which have their own forward markets. With so many different areas most of the power forward markets are not very liquid and frequently transactions entail considerable price impacts.

For numerical illustration of the linear approximation algorithm (see Subsection 4.3.1) we first consider an energy trader aiming at closing a long forward position of electricity. We then extend the example and assume that the trader wants to buy, over the same trading period, also the amount of coal needed for the electricity production.

For a more detailed discussion of liquidity in power forward markets we refer to [87], where also *deterministic* liquidation strategies within a Gaussian model are determined.

4.4.1. Liquidation of a long forward power position

A power trader has got the instruction to sell, within the next 50 trading days, a baseload position with delivery in the front year, i.e. the next calendar year. Suppose that the total position amounts to 5 TWh, which corresponds to a delivery at a constant rate of approximately 570 MW per hour throughout the whole front year.

We assume that the forward price process of the front year power baseload is a geometric Brownian motion. The volatility is set to 20% per year and we suppose that the forward is initially traded at a price of $S_0 = 50$ € per MWh.

The liquidity of power forwards strongly depends on the time to the begin of delivery. For our numerical experiments we assume a price impact of 1 €/MWh per TWh sold. More precisely, selling at time 0 an amount of 1 TWh is possible only at a price of 49 € per MWh. The relative price impact parameter therefore satisfies $\eta(s) = 0.02 \frac{1}{\text{TWh}} s$, where s is the power price in € per MWh.

The risk function is assumed to be of the form

$$\lambda(s) = \max(0, c(a - s))^2 \tag{4.54}$$

(cf. Subsection 4.1.1). We interpret a as the price level beyond which the risk vanishes, and c as the price sensitivity. The higher a and c, the more risk-averse the trader. In the following we will set a either equal to S_0, or equal to S_0 multiplied with the exponential of the standard deviation of 5 daily log returns, which is approximately 53.25.

Notice that by Example 4.3.3 the model assumptions guarantee the convergence of the linear approximation algorithm. We choose an equidistant grid for the log price, consisting of 201 discretization points. The distance between two gridpoints is chosen to be equal to the standard deviation of a daily log return. We have implemented the linear approximation algorithm in MATLAB. The calculation time of the optimal decision rules depends linearly on the number of grid points. For 201 discretization points the calculation takes e.g. on a 2.53 GHz Intel Core 2 Duo Processor approximately 0.8 seconds.

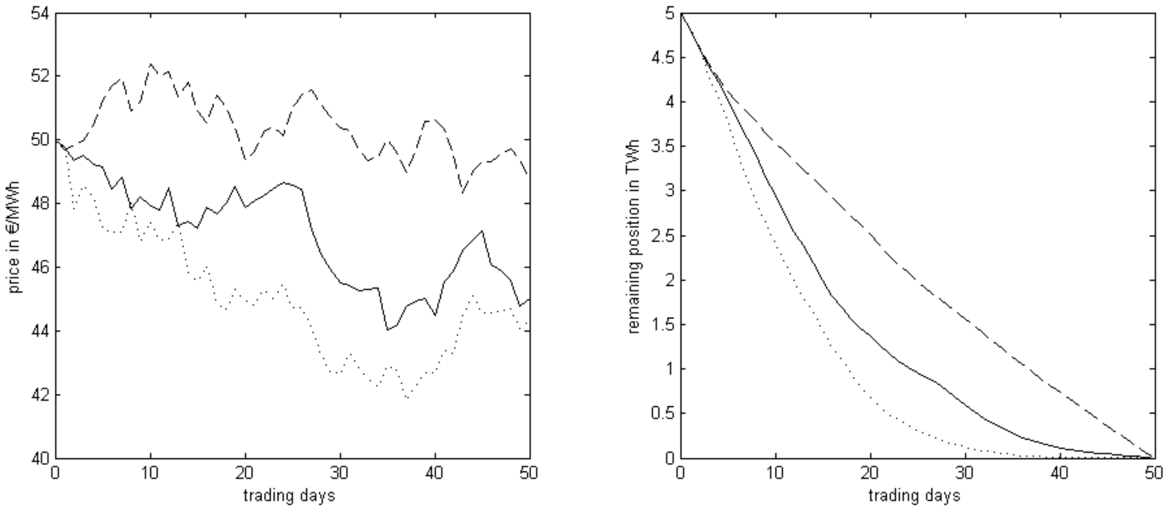

Figure 4.1.: The left hand panel shows three realizations of the price process with initial value $S_0 = 50$ € and a yearly volatility of 20%. The associated liquidation paths are depicted in the right hand panel for $c = 0.03$ and $a = S_0$.

Figure 4.1 illustrates the dependence of the liquidation path on the realized price process. The left hand panel depicts three realizations of the price process, and the right hand panel shows the corresponding optimal liquidation paths of the asset position. The dashed price path moves upward at the beginning and stays above S_0 almost throughout the trading period, and thus the position is closed almost linearly. The solid and dotted price processes have a downward trend over the whole trading period, and thus entail higher initial selling rates.

Figure 4.2 shows how the optimal selling rate depends on the current asset price. Recall that the selling rate is the percentage of the open position that is sold. The left hand panel shows the optimal rate when 50, 25 respectively 10 trading days are left until the position has to be closed. For asset prices smaller than the initial asset price S_0, the rate increases almost linearly as the asset price decreases. Moreover, the rate does not depend on the time left, but is only determined by the large risk exposure due to low asset prices. The high risk entails a quick closure of the remaining position in less than 10 days.

If the asset price exceeds S_0, then the risk function vanishes and it is optimal sell at the rate minimizing the liquidation costs, namely $\frac{1}{\text{\# trading days left}}$. Put differently, whenever the price exceeds the threshold S_0, the position is essentially closed linearly.

The panel on the right hand side of Figure 4.2 demonstrates the dependence of the optimal selling rate on the price sensitivity parameter c. The higher the sensitivity, the faster the selling rate increases as asset prices fall. If the asset price exceeds the threshold S_0, then the difference vanishes and the selling rate converges to the constant rate of $\frac{1}{\text{\# trading days left}}$.

For the risk function λ given by (4.54) we can summarize the optimal selling rate in

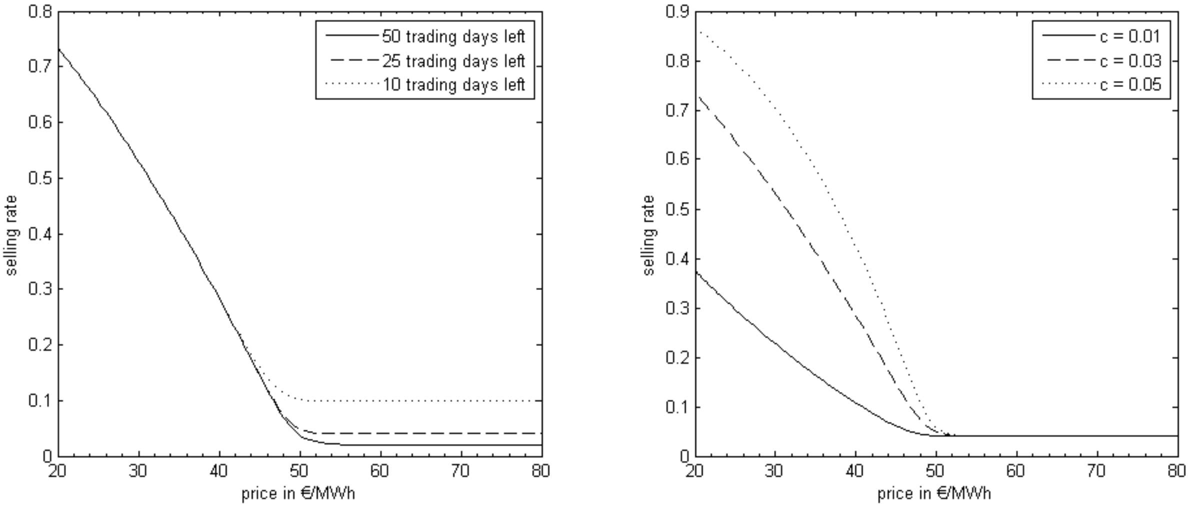

Figure 4.2.: The left hand panel shows the optimal selling rate, as a function of the current price, for various time horizons. The risk threshold parameter is set to $a = S_0$, and the price sensitivity is chosen to be equal to $c = 0.03$. The right hand panel depicts the optimal selling rate function for various sensitivities c. Here we assume that 25 trading days are left.

a simple rule-of-thumb: to avoid downside risk the trader should sell, approximately, at the rate

$$r(S) = \frac{1}{\#\ \text{trading days left}} + c \max(0, S - a),$$

where S is the current power price. The higher c, the faster the position is closed when prices plunge below the reference price a. Note that this rule-of-thumb does not depend on the price volatility σ.

We next turn to a simulation analysis of realized proceeds. To this end we simulate 500.000 price paths, and calculate the optimal liquidation along every realization. Table 4.1 shows statistical features of the optimal liquidation paths obtained for various threshold and sensitivity parameters. A higher threshold a resp. sensitivity c speeds up the selling rate and hence increases liquidation costs. This is confirmed by the third column of Table 4.1. Column 4-7 show that a higher c entails the left hand quantiles of the proceeds to increase. Here $Q(q)$ is the q-quantile of the implementation shortfall, which is the book value minus the realized proceeds. Selling faster when prices are small, implies that the left hand tail of the proceeds becomes thinner. In the case $a = 50$ the right hand tail of the proceeds is not affected by the parameter c, and therefore skewness of the proceeds increases as c increases. In the case $a = 53.25$ the price sensitivity c has still a significant impact if the price lies between S_0 and 53.25. As c increases it is optimal to sell faster for all prices below 53.25. This explains why the skewness slightly falls when c increases from 0.03 to 0.05.

a	c	mean	Q(0.001)	Q(0.01)	Q(0.02)	Q(0.05)	median	std	skewness
50	0.01	249.4831	30.4385	24.9187	22.6899	19.0909	248.8156	12.0493	0.3326
	0.03	249.3340	23.0162	19.2262	17.7383	15.3153	247.9665	10.4723	0.6480
	0.05	249.1471	20.2312	16.9013	15.5760	13.4930	247.5457	9.4249	0.8175
	0.1	248.7049	17.4879	14.5188	13.3885	11.6060	247.2418	7.9205	1.0093
53.25	0.01	249.3924	25.6458	20.8929	19.0192	16.1043	248.3997	10.5286	0.5169
	0.02	249.0665	20.2116	16.5431	15.0896	12.8258	247.9780	8.4513	0.7463
	0.03	248.6863	17.7910	14.5253	13.2748	11.2892	247.8017	7.1252	0.8255
	0.05	247.8973	15.4531	12.6676	11.6338	10.0240	247.3149	5.5554	0.8162

Table 4.1.: Price-sensitive liquidation

λ	mean	Q(0.001)	Q(0.01)	Q(0.02)	Q(0.05)	median	std	skewness
0	249.4969	36.3825	28.2279	25.2388	20.5966	249.1234	12.6347	0.1838
0.001	249.4530	31.5276	24.5447	21.9368	17.9504	249.1278	10.9327	0.1845
0.003	249.2788	26.4470	20.6615	18.4620	15.1503	249.0138	9.0543	0.1747
0.01	248.7534	20.5016	16.0774	14.4718	11.9651	248.5908	6.6788	0.1407

Table 4.2.: Price-**in**sensitive liquidation

For comparison, Table 4.2 depicts the statistical characteristics of the realized proceeds for constant risk function *not* depending on the price. The function λ is assumed to be constant equal to the parameter shown in the first column.

The lower panel of Figure 4.3 shows the histogram of the realized proceeds for $a = S_0$ and $c = 0.05$. Note that the left hand tail is considerably thinner than the right hand tail. For comparison, the upper panel shows the histogram of realized proceeds, for the same set of simulated paths, under price-*in*sensitive risk function. The risk function is assumed to be constant equal to 0.003. Table 4.1 and 4.2 show that in both cases the mean of the proceeds are pretty close, though the 1% quantiles differ considerably. The dashed line in both histograms indicates the level of the 1% quantile. The price-sensitive risk function entails a shift of the quantile to the right. Moreover, the proceeds in the upper panel are symmetrically distributed around the mean, whereas in the price-sensitive case the revenues are right-skewed.

To sum up, we see that by adopting a price-sensitive selling a trader can shift mass from the very left hand side of the revenue distribution into the center. Price-sensitive selling, therefore, makes an extreme shortfall below the book value very unlikely.

4.4.2. Closing a long position of power and a short position of coal

We next demonstrate how our algorithm can be efficiently employed to a multi-asset position. Consider again the trader from Subsection 4.4.1 aiming at closing a baseload front year long position of 5 TWh over the next 50 trading days. Now suppose in addition that the trader is working for an energy company that runs coal power plants

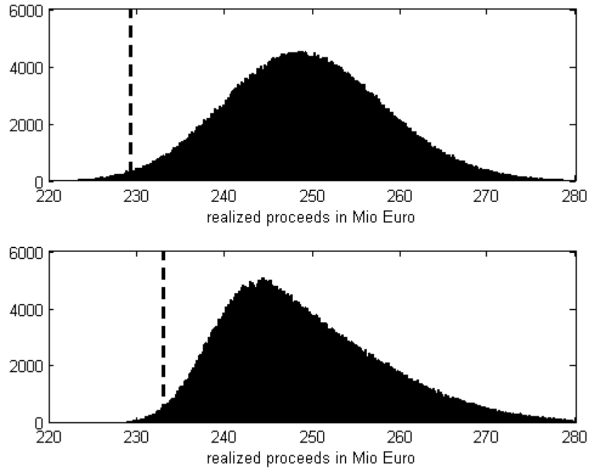

Figure 4.3.: For a set of 500.000 price paths the upper panel depicts a histogram of realized proceeds using a price-**in**sensitive risk function $\lambda = 0.003$. The lower panel shows a histogram of realized proceeds for the same price paths using the price-sensitive risk function $\lambda(s) = \max(0, c(a-s))^2$ with $a = S_0 = 50$€ and $c = 0.05$. The dashed line indicates the respective 1% quantile.

and that he has got the instruction to buy, over the same trading period, the amount of coal needed for the power production.

We assume that the company's coal power plants possess a thermodynamical efficiency of 30% and that the coal type bought has an energy of 6.67 kWh per kg. This means that the coal power plants generate 2 MWh out of 1 ton of coal. Therefore, 2.5 Mio tons of coal are needed for a power production of 5 TWh.

The revenues from a coal power plant are essentially determined from the *dark spread*, i.e. the spread between the power price and the coal price. The risk from an open power and coal position is therefore determined by the dark spread, and not by the *absolute* power and coal price. For determining dark-spread-sensitive liquidation strategies it turns out to be convenient to work with the *geometric dark spread* defined by

$$\mathrm{GDS} = \frac{hS}{K},$$

where S is the power price, K the coal price, and $h = 2$MWh/t the power generation rate of the coal plant.

For our numerical experiment we assume that the coal and the power forward price processes are geometric Brownian motions with volatility $\sigma_S = \sigma_K = 20\%$, and that the correlation between the Brownian motions driving both processes is $\rho = 70\%$. We suppose that the power forward is initially traded at a price of $S_0 = 50$ € per MWh, and the coal forward at a price of $K_0 = 80$ € per ton. Denote by GDS_t the geometric dark spread at time t. Notice that the initial geometric dark spread satisfies $\mathrm{GDS}_0 = 1.25$ and that GDS_t is lognormally distributed. Moreover, the process $(\mathrm{GDS}_t)_{t\geq 0}$ satisfies the Markov property, i.e. the conditional expectation of GDS_T with respect to the market information at some earlier time $t < T$ is a function of GDS_t only (cf. Subsection 4.1.2).

Being more international, coal markets are usually far more liquid than power markets. In order to appeal to Proposition 4.1.5 we employ *absolute* price impact and suppose that the risk matrix λ is a scalar function of the GDS. More precisely, we assume that the absolute power price impact is 1 €/MWh per TWh sold, i.e. $\eta(s) = 1 \frac{€}{\mathrm{MWh\ TWh}}$. The coal price impact is set equal to 0.03 €/t per 10^6 tons bought. The risk matrix λ is the price weighted covariance matrix C of the log returns multiplied with a scalar function of the GDS, namely

$$\lambda(\mathrm{GDS}) = \max(0, c(a - \mathrm{GDS}))^2 \begin{pmatrix} \sigma_S^2 S_0^2 & \rho\sigma_S\sigma_K S_0 K_0 \\ \rho\sigma_S\sigma_K S_0 K_0 & \sigma_K^2 K_0^2 \end{pmatrix}, \tag{4.55}$$

with $c = 1$ and $a = \mathrm{GDS}_0$. We can, therefore split the portfolio into two components with orthogonal risk and price impact (see Proposition 4.1.5 for details). The optimal liquidation of the total portfolio is hence the optimal liquidation of each orthogonal component on its own.

The Markov property and the form of the risk matrix (4.55) imply that the value function of each orthogonal portfolio component is a quadratic function of the size with coefficients depending only on the geometric dark spread. The coefficients can be determined via a one-dimensional function recursion (cf. Subsection 4.1.2).

Figure 4.4 illustrates the dependence of the closure on the geometric dark spread evolution. The left hand panel depicts two realizations of the GDS, and the right hand panel shows the associated closure of the power position (5 TWh) and the coal position (- 2.5 Mio tons). Observe that in both realizations the two positions are closed symmetrically, irrespective of the different price impacts. The coal price is highly correlated with the power price, and hence buying coal faster than selling power would decrease the diversifying effect and hence increase risk.

Finally observe that the overall closing rate increases when the GDS falls. In the solid realization the GDS soon rises above the initial value, and hence the position is closed almost linearly. The dashed realization has a downward trend during the first half of the trading period, which entails a higher risk exposure and hence that the position is closed faster.

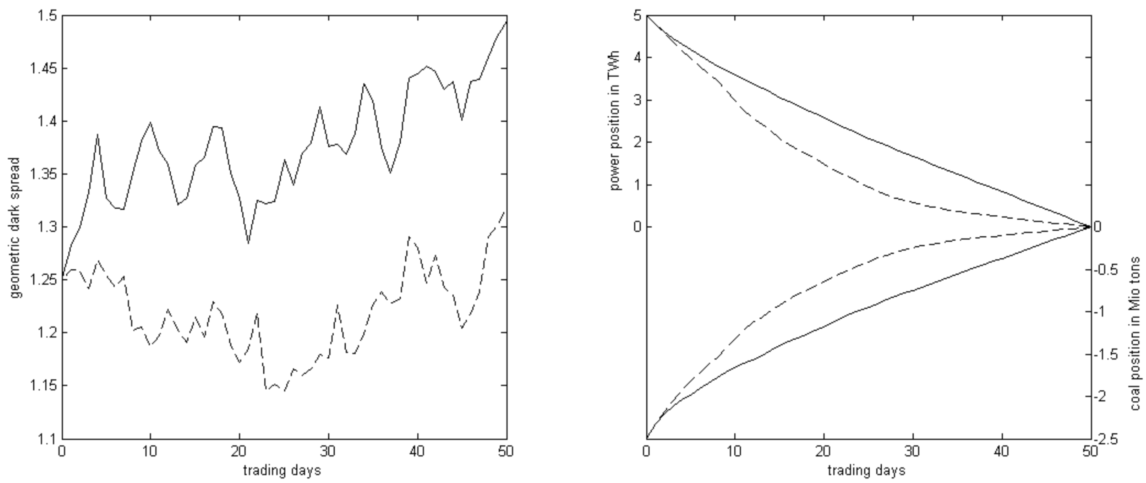

Figure 4.4.: The left hand panel depicts two realizations of the geometric dark spread and the right hand panel shows the associated closure of the power position (5 TWh) and the coal position (- 2.5 Mio tons).

5. Optimal position closure under stochastic liquidity

Liquidity in financial markets usually is not constant over time - it fluctuates and sometimes faces shocks. To benefit from periods when trading is cheap, traders aiming at closing a large asset position need to employ trading strategies that dynamically respond to the evolution of liquidity.

The aim of this chapter is to approach the liquidation problem with a randomly changing price impact by means of backward stochastic differential equations (BSDEs). To this end we consider the following generalization of the continuous-time liquidation problem from Section 4.2. The goal is to minimize the functional

$$J(x) = E\left[\int_0^T (\eta_t|\dot{x}_t|^p + \gamma_t|x_t|^p)dt + 1_{A^c}\xi|x_T|^p\right] \tag{5.1}$$

over all absolutely continuous paths $(x_t)_{t\in[0,T]}$ starting in $\zeta \in \mathbb{R}$ and satisfying the terminal constraint $x_T = 0$ on A. Here $p > 1$ and (η, γ) are two nonnegative stochastic processes that are progressively measurable with respect to the natural filtration $(\mathcal{F}_t)$ generated by a Brownian motion. The random variable $\xi : \Omega \to [0, \infty]$ is $\mathcal{F}_T$-measurable and the set A is defined as $A = \{\xi = \infty\}$. The control strategies x are adapted to $(\mathcal{F}_t)$.

In the framework of an optimal liquidation problem the model parameters can be interpreted as follows. The first term $\int_0^T \eta_t|\dot{x}_t|^p dt$ in (5.1) can be interpreted as the liquidity costs entailed by closing the position, where η is a stochastic price impact factor. The exponent p determines the dependence of the price impact on the trading rate $\dot{x}$. In Chapter 4 we only considered a linear impact ($p = 2$), i.e. the difference between the realized price and the uninfluenced price is given by $\eta_t\dot{x}_t$. Here we allow for the general power law dependence $\eta_t \operatorname{sgn}(\dot{x}_t)|\dot{x}_t|^{p-1}$. If $p < 2$, then the price increment depends in a concave way on the trading rate, else the dependence is convex. We refer the reader to [59] for empirical studies about the shape of the impact function. The second term $\int_0^T \gamma_t|x_t|^p dt$ can be seen as a measure of the risk associated to the open position. For example one can draw inspiration from Section 4.1 and choose $\gamma_t = \lambda(S_t)$, where λ is a function of the asset's forward price S. The random variable ξ allows to relax the constraint of closing the position at all events. On the set $A^c = \{\xi < \infty\}$ there is no terminal constraint, but any nonzero terminal state is penalized by $\xi|x_T|^p$. For instance one can specify a set of outcomes A^c where liquidation is particularly expensive, e.g. $A^c = \{\int_0^T \eta_t dt \geq k\}$ consists of all realizations where the average price impact exceeds some threshold $k > 0$. Choosing $\xi = \infty 1_A$ implies that the position has to be closed only if $\int_0^T \eta_t dt < k$. The case $\xi = \infty$ corresponds to a binding liquidation constraint as in Chapter 4.

In this chapter, the method for solving the control problem (5.1) draws on the notion of backward stochastic differential equations (BSDEs). Its solution is characterized in terms of the minimal solution of the BSDE

$$dY_t = \left((p-1)\frac{Y_t^q}{\eta_t^{q-1}} - \gamma_t \right) dt + Z_t dW_t \tag{5.2}$$

(where $q = 1/(1-\frac{1}{p})$) satisfying the *singular* terminal condition

$$\liminf_{t \nearrow T} Y_t \geq \xi. \tag{5.3}$$

In particular, Y explodes on the set A: $\liminf_{t\nearrow T} 1_A Y_t = \infty$. It is shown that if η and γ satisfy some nice integrability condition, then there exists a minimal supersolution (Y, Z) of the BSDE (5.2) with terminal condition (5.3). We subsequently prove, without any further assumptions, that there exists an optimal control of the problem (5.1) and that it is given by $x_t^* = \zeta e^{-\int_0^t \left(\frac{Y_s}{\eta_s}\right)^{q-1} ds}$. Note that the terminal condition (5.3) is necessary for the constraint $x_T^* = 0$ on A to be satisfied.

The chapter is organized as follows. Section 5.1 precisely describes the modeling set-up and presents the main results. Moreover, it contains a heuristic derivation of why the BSDE (5.2) with singular terminal condition provides a solution of the control problem. In Section 5.2 it is verified that under some nice integrability conditions on η and γ, there exists a minimal supersolution of the BSDE (5.2). Section 5.3 turns to a verification: It is shown that the optimal control and value function can indeed be characterized by the BSDE solution constructed in Section 5.2. Finally, Section 5.4 studies in detail the special case where γ is zero, $\xi = \infty$ and η has uncorrelated multiplicative increments. In exactly this case the optimal control is deterministic.

5.1. Main results

Fix a deterministic, finite time horizon $T > 0$ and a probability space $(\Omega, \mathcal{F}, P)$ which supports a d-dimensional Brownian motion $(W_t)_{0\leq t\leq T}$, where $d \in \mathbb{N}$. Let $(\mathcal{F}_t)_{t\in[0,T]}$ denote the completed filtration generated $(W_t)_{0\leq t\leq T}$. Throughout we assume that $(\eta_t)_{t\in[0,T]}$ and $(\gamma_t)_{t\in[0,T]}$ are nonnegative, progressively measurable stochastic processes. Moreover, let $\xi : \Omega \to [0, \infty]$ be a (not necessarily finite) $\mathcal{F}_T$-measurable random variable. We introduce the set $A = \{\xi = \infty\}$. We assume $p > 1$ and denote by $q = 1/(1-\frac{1}{p})$ its Hölder conjugate. We consider the stochastic control problem to minimize the functional

$$J(x) = E\left[\int_0^T (\eta_t |\dot{x}_t|^p + \gamma_t |x_t|^p) dt + 1_{A^c} \xi |x_T|^p \right] \tag{5.4}$$

over all progressively measurable processes $x : \Omega \times [0, T] \to \mathbb{R}$ that possess absolutely continuous sample paths and satisfy the constraints $x_0 = \zeta \in \mathbb{R}$ and $x_T = 0$ a.s. on A. We denote the set of all these controls by $\mathcal{A}_0$, and define

$$v = \inf_{x \in \mathcal{A}_0} J(x). \tag{5.5}$$

We show that under some nice integrability conditions on η and γ there exists an optimal control $x^* \in \mathcal{A}_0$; i.e. $J(x^*) = v$. Moreover we characterize the optimal control by means of a BSDEs with a singular terminal condition. We define the notion of a solution in the style of [69].

Definition 5.1.1. We say that a pair of progressively measurable processes (Y, Z) with values in $\mathbb{R} \times \mathbb{R}^d$ solves the BSDE (5.2) with singular terminal condition $Y_T = \xi$ if it satisfies

(i) for all $0 \le s \le t < T$: $Y_s = Y_t - \int_s^t \left((p-1)\frac{Y_r^q}{\eta_r^{q-1}} - \gamma_r\right) dr - \int_s^t Z_r dW_r$;

(ii) for all $0 \le t < T$: $E\left[\sup_{0\le s\le t} |Y_s|^2 + \int_0^t |Z_r|^2 dr\right] < \infty$;

(iii) $\lim_{t\nearrow T} Y_t = \xi$, a.s.

If (iii) does not hold true, but $\liminf_{t\nearrow T} Y_t \ge \xi$, then we say that (Y,Z) is a supersolution. We introduce the following spaces of processes. For $i = 1, 2$ and $t \le T$ let

$$\mathcal{M}^i(0,t) = L^i(\Omega \times [0,t], \mathcal{P}, P \otimes \lambda),$$

where λ is the Lebesgue measure and $\mathcal{P}$ denotes the σ-algebra of $(\mathcal{F}_t)$-progressively measurable subsets of $\Omega \times [0, T]$. Throughout we assume that η and γ satisfy the integrability conditions

$$\textbf{(I1)} \qquad \eta \in \mathcal{M}^2(0,T) \text{ and } 1/\eta^{q-1} \in \mathcal{M}^1(0,T),$$

$$\textbf{(I2)} \qquad E\int_0^T (T-s)^p \gamma_s ds < \infty \text{ and } \gamma \in \mathcal{M}^2(0,t) \text{ for all } t < T.$$

In our first main result we prove existence of a minimal solution of the BSDE (5.2).

Theorem 5.1.2. *Assume that Conditions* (**I1**) *and* (**I2**) *are satisfied. Then there exists a minimal supersolution* (Y, Z) *of the BSDE* (5.2) *with singular terminal condition* $Y_T = \xi$.

In the second main result we characterize the value function and the optimal control in terms of the minimal solution.

Theorem 5.1.3. *Suppose Conditions* (**I1**) *and* (**I2**)*, and let* (Y, Z) *be the minimal solution of* (5.2)*. Then*

$$v = Y_0 |\zeta|^p$$

and the optimal control is given by

$$x_t^* = \zeta \exp\left(-\int_0^t \left(\frac{Y_s}{\eta_s}\right)^{q-1} ds\right),$$

for all $t \in [0, T]$.

The following deterministic example illustrates that a violation of the integrability condition $1/\eta^{q-1} \in \mathcal{M}^1(0,T)$ may lead to a minimization problem where no optimal control exists.

Example 5.1.4. *Let $T = 1$, $\eta_t = (1-t)^\beta$ for some $\beta \geq 0$, $\gamma_t = 0$, $\xi = \infty$ and $p = q = 2$. Then we have $1/\eta^{q-1} \in L^1([0,T])$ if and only if $\beta < 1$. In this case Theorem 5.1.3 yields that $x_t = (1-t)^{1-\beta}$ is an optimal control. In the case $\beta \geq 1$ consider the control $x_t = (1-t)^\alpha$ for some $\alpha > 0$. We compute*

$$J(x) = \int_0^1 \eta_t \dot{x}_t^2 dt = \alpha^2 \int_0^1 (1-t)^{2\alpha+\beta-2} dt.$$

Since $\beta \geq 1 > 1 - 2\alpha$ the integral is finite and has the value

$$J(x) = \frac{\alpha^2}{2\alpha + \beta - 1}.$$

Taking the limit $\alpha \searrow 0$ yields $v = 0$, but there exists no control in $\mathcal{A}_0$ attaining this value.

Remark 5.1.5. If $p = 1$, then the control problem also does not possess an optimal control in $\mathcal{A}_0$ (except for some simple cases). For $p = 1$ the right formulation of the problem would be to allow for singular controls; and consequently the description of optimal controls would require different methods.

We prove Theorem 5.1.2 in Section 5.2 (see Theorem 5.2.2) and Theorem 5.1.3 in Section 5.3 (see Theorem 5.3.2). Before tackling the proofs we provide a heuristic derivation of the BSDE (5.2).

Heuristic derivation of the BSDE

Throughout this section we assume $\zeta > 0$. First we show that in this case we can restrict attention to nonincreasing nonnegative controls. To this end we denote the set of controls in $\mathcal{A}_0$ with nonincreasing sample paths by $\mathcal{D}_0$.

Lemma 5.1.6. *Every control $x \in \mathcal{A}_0$ can be modified to a control $\underline{x} \in \mathcal{D}_0$ such that $J(x) \geq J(\underline{x})$. In particular, we have $v = \inf_{x \in \mathcal{D}_0} J(x)$.*

Proof. Let $x \in \mathcal{A}_0$ and define its running minimum cut off at zero by $\underline{x}_t = \min_{0 \leq s \leq t} x_s \vee 0$. Notice that $\underline{x}$ is absolutely continuous since $\underline{x}_t = \int_0^t \dot{x}_s 1_{\{x_s = \underline{x}_s\}} ds$. Hence $\underline{x} \in \mathcal{D}_0$. Observe that $|\dot{\underline{x}}_t| \leq |\dot{x}_t|$, and therefore we have $E\left[\int_0^T \eta_t |\dot{x}_t|^p dt\right] \geq E\left[\int_0^T \eta_t |\dot{\underline{x}}_t|^p dt\right]$. Since $\underline{x} \leq x$ on $\Omega \times [0,T]$ it follows that $E\left[\int_0^T \gamma_t |x_t|^p dt + 1_{A^c} \xi |x_T|^p\right] \geq E\left[\int_0^T \gamma_t |\underline{x}_t|^p dt + 1_{A^c} \xi |\underline{x}_T|^p\right]$. Thus, we obtain $J(x) \geq J(\underline{x})$. □

The next result, a maximum principle, provides a sufficient condition for optimality in (5.5). We remark that we use it only for the heuristic derivation of the BSDE (5.2). The rigorous verification in Section 5.3 will be performed via a penalization.

Proposition 5.1.7 (Maximum principle)**.** *Assume that* $x \in \mathcal{D}_0$ *such that*

(i) $\dot{x}_T = (\frac{\xi}{\eta})^{q-1} x_T$ *on* A^c *and*

(ii) $M_t = p\eta_t |\dot{x}_t|^{p-1} + p\int_0^t \gamma_s x_s^{p-1} ds$ *is a martingale with* $E[M_T^2] < \infty$

Then x *is optimal in* (5.5).

Proof. Let $g(z) = |z|^p$ and $x \in \mathcal{D}_0$ such that $M_t = p\eta_t |\dot{x}_t|^{p-1} + p\int_0^t \gamma_s x_s^{p-1} ds$ is a martingale with $E[M_T^2] < \infty$. Let $y \in \mathcal{D}_0$ and introduce $\theta_t = x_t - y_t$. Then θ satisfies $\theta_0 = \theta_T = 0$ a.s. Furthermore, since x and y are nonincreasing it follows that θ is bounded: $|\theta_t| \leq 2|\zeta|$. Since $\dot{x} \leq 0$ on $\Omega \times [0,T]$ we have $g'(\dot{x}_t) = -p|\dot{x}_t|^{p-1}$. The convexity of g implies for all $t \in [0,T]$

$$g(\dot{x}_t) - g(\dot{y}_t) \leq g'(\dot{x}_t)(\dot{x}_t - \dot{y}_t).$$

Thus, by integration by parts we obtain

$$\begin{aligned}\int_0^T \eta_t (g(\dot{x}_t) - g(\dot{y}_t)) dt &\leq \int_0^T \eta_t g'(\dot{x}_t) d\theta_t = \int_0^T \left(\int_0^t p\gamma_s x_s^{p-1} ds - M_t \right) d\theta_t \\ &= -1_{A^c} \xi g'(x_T)\theta_T + \int_0^T \theta_t dM_t - \int_0^T \gamma_t g'(x_t)\theta_t dt.\end{aligned}$$

Since θ is bounded and M is a martingale with $E[M_T^2] < \infty$ it follows that the integral process $\int_0^\cdot \theta_t dM_t$ is a martingale starting in 0. In particular, it vanishes in expectation. Using again the convexity of g yields for $t \in [0,T]$

$$g(x_t) - g(y_t) \leq g'(x_t)(x_t - y_t).$$

Taking expectations implies optimality of x:

$$\begin{aligned}E\left[\int_0^T \eta_t (g(\dot{x}_t) - g(\dot{y}_t)) dt\right] &\leq -E\left[\int_0^T \gamma_t g'(x_t)\theta_t dt + 1_{A^c} \xi g'(x_T)\theta_T\right] \\ &\leq -E\left[\int_0^T \gamma_t (g(x_t) - g(y_t)) dt + 1_{A^c} \xi (g(x_T) - g(y_T))\right].\end{aligned}$$

□

We next observe that the *relative* control rate $r_t = \frac{\dot{x}_t}{x_t}$ of an optimal control $x \in \mathcal{A}_0$ is independent of the current state x_t. To this end fix $t < T$ and $\zeta_2 > \zeta_1 > 0$. Assume that $(x_s^1)_{t \leq s \leq T}$ is an optimal control to close the position ζ_1 in the period $[t,T]$. Then the homogeneity of $y \mapsto |y|^p$ implies that the control $x_s^2 = \frac{\zeta_2}{\zeta_1} x_s^1$, $s \in [t,T]$, is optimal to close the position ζ_2 in the period $[t,T]$. In particular the relative control rates at time t coincide $\frac{\dot{x}_t^1}{\zeta_1} = \frac{\dot{x}_t^2}{\zeta_2}$. Hence, an optimal control can be represented in feedback form $\dot{x}_t = r_t x_t$, where r_t is the relative control rate, which does not depend on x_t. Using q, the Hölder conjugate of p, we rewrite r_t as $r_t = -\left(\frac{Y_t}{\eta_t}\right)^{q-1}$ for some semi-martingale Y and make the ansatz that an optimal control x is of the form

$$\dot{x}_t = -\left(\frac{Y_t}{\eta_t}\right)^{q-1} x_t \tag{5.6}$$

with $x_0 = 1$. The solution of this pathwise ordinary differential equation is given by

$$x_t = e^{-\int_0^t \left(\frac{Y_s}{\eta_s}\right)^{q-1} ds}. \tag{5.7}$$

Proposition 5.1.7 shows that $x \in \mathcal{A}_0$ is optimal if the process $p\eta|\dot{x}|^{p-1} + p\int_0^\cdot \gamma_s x_s^{p-1} ds$ is a martingale. Since $(\mathcal{F}_t)$ is a Brownian filtration this is equivalent to the existence of a predictable process ϕ such that

$$d(\eta|\dot{x}|^{p-1})_t + \gamma_t x_t^{p-1} dt = \phi_t dW_t.$$

Using the equality $\eta_t|\dot{x}_t|^{p-1} = Y_t x_t^{p-1}$ and applying the integration by parts formula to the product Yx^{p-1} we obtain

$$\begin{aligned} d(\eta|\dot{x}|^{p-1})_t + \gamma_t x_t^{p-1} dt &= x_t^{p-1} dY_t + (p-1) Y_t x_t^{p-2} dx_t + \gamma_t x_t^{p-1} dt \\ &= x_t^{p-1}\left(dY_t - \left((p-1)\frac{Y_t^q}{\eta_t^{q-1}} - \gamma_t\right) dt\right). \end{aligned}$$

Setting $Z_t = \phi_t / x_t^{p-1}$ we see that Y satisfies the BSDE

$$dY_t = \left((p-1)\frac{Y_t^q}{\eta_t^{q-1}} - \gamma_t\right) dt + Z_t dW_t. \tag{5.8}$$

It remains to specify a terminal condition for Y. It follows from Proposition 5.1.7 (i) that on A^c the terminal value of Y is given by $Y_T = \xi$. To meet the state constraint $x_T = 0$ on $A = \{\xi = \infty\}$ the integral $\int_0^t \left(\frac{Y_s}{\eta_s}\right)^{q-1} ds$ necessarily has to explode for $t \nearrow T$. In Theorem 5.3.2 we show that minimal solution of (5.8) satisfying the terminal condition $\liminf_{t\nearrow T} Y_t \geq \xi$ leads indeed to an optimal strategy.

5.2. Construction of a BSDE solution with singular terminal condition

In this section we construct a solution of the BSDE (5.2) with singular terminal condition. To this end we first show existence of solutions to BSDEs with cut off drivers and finite terminal condition $\xi \wedge L$. In a second step we let L tend to infinity and obtain a supersolution with a singular terminal condition. We show that this particular solution is the minimal supersolution of (5.8). We remark that the second step of our construction bears similarities with the existence proof conducted by Popier in [69] resp. [70].

Let us clarify some terminology concerning BSDEs. The pair consisting of the driver and the terminal condition of a BSDE will be referred to as its *parameters*. Given a solution (Y, Z) of a BSDE, we call the first component Y the *solution process* and the second component Z the *martingale component*.

5.2.1. Approximation

Consider the BSDE

$$dY_t^L = \left((p-1)\frac{(Y_t^L)^q}{\eta_t^{q-1}} - (\gamma_t \wedge L)\right) dt + Z_t^L dW_t, \tag{5.9}$$

with terminal condition $Y_T^L = \xi \wedge L$.

Proposition 5.2.1. *Assume that $\eta \in \mathcal{M}^2(0,T)$ and $\frac{1}{\eta^{q-1}} \in \mathcal{M}^1(0,T)$. Then there exists a solution (Y^L, Z^L) to (5.9) with $Z^L \in \mathcal{M}^2(0,T)$. For every $t \in [0,T]$ the random variable Y_t^L is bounded from above*

$$Y_t^L \leq (1+T)L \wedge \frac{1}{(T-t)^p} E\left[\int_t^T (\eta_s + (T-s)^p \gamma_s) ds \middle| \mathcal{F}_t\right]. \tag{5.10}$$

Proof. Let $f(t,y) = -(p-1)\frac{y^q}{\eta_t^{q-1}} + (\gamma_t \wedge L)$ denote the driver of the BSDE (5.9). Define $f^\delta(t,y) = -(p-1)\frac{y^q}{(\eta_t \vee \delta)^{q-1}} + (\gamma_t \wedge L)$ for $\delta > 0$. Being decreasing in y, bounded in ω, the driver $(\omega, t, y) \mapsto f^\delta(t, y \vee 0)$ - which does not depend on z - satisfies all conditions of Theorem 2.2. in [62]. Hence, for every $L > 0$ there exists a solution $(Y^{\delta,L}, Z^{\delta,L})$ to the BSDE with parameters $(f^\delta(t, y \vee 0), \xi \wedge L)$. Moreover, any such solution satisfies

$$E\left[\sup_{0 \leq t \leq T} |Y_t^{\delta,L}|^2 + \int_0^T (Z_t^{\delta,L})^2 dt\right] < \infty. \tag{5.11}$$

For $L = 0$ the solution is given by $(Y^{\delta,0}, Z^{\delta,0}) = (0,0)$. The comparison theorem [62, Theorem 2.4] implies that $Y^{\delta,L}$ is nonnegative and, hence, $Y^{\delta,L}$ is also a solution to the BSDE with parameters $(f^\delta, \xi \wedge L)$.

We can also derive an upper bound for $Y^{\delta,L}$ by appealing to the comparison theorem. Note that we have $f^\delta(t,y) \leq L$ for $y \geq 0$. This implies

$$Y_t^{\delta,L} \leq (1+T)L \tag{5.12}$$

for all $t \in [0,T]$.

We obtain a solution of the BSDE (5.9) by letting δ converge to zero. Indeed, the mapping $\delta \mapsto f^\delta$ is increasing, which implies that $Y^{\delta_1,L} \leq Y^{\delta_2,L}$ if $\delta_1 \leq \delta_2$. In particular we can define Y^L as the decreasing limit of $Y^{\delta,L}$ as $\delta \searrow 0$. For the convergence of the control process $Z^{\delta,L}$, let $(\delta_n)_{n \geq 0}$ be a sequence with $\delta_n \searrow 0$ as $n \to \infty$. Fix $n \geq m$. Then we have $Y^{\delta_n,L} \leq Y^{\delta_m,L}$. For all $0 \leq t \leq T$, Itô's formula leads to

$$\begin{aligned}\int_0^T (Z_s^{\delta_n,L} - Z_s^{\delta_m,L})^2 ds = &- (Y_0^{\delta_n,L} - Y_0^{\delta_m,L})^2 \\ &- 2\int_0^T (Y_s^{\delta_n,L} - Y_s^{\delta_m,L})(Z_s^{\delta_n,L} - Z_s^{\delta_m,L}) dW_s \\ &+ 2\int_0^T (Y_s^{\delta_n,L} - Y_s^{\delta_m,L})(f^{\delta_n}(s, Y_s^{\delta_n,L}) - f^{\delta_m}(s, Y_s^{\delta_m,L})) ds\end{aligned} \tag{5.13}$$

Estimates (5.11) and (5.12) imply

$$E\left[\int_0^T (Y_s^{\delta_n,L} - Y_s^{\delta_m,L})(Z_s^{\delta_n,L} - Z_s^{\delta_m,L})dW_s\right] = 0.$$

By monotonicity of f^{δ_m} and estimate (5.12) we have

$$\begin{aligned}(Y_s^{\delta_n,L} - Y_s^{\delta_m,L})&(f^{\delta_n}(s,Y_s^{\delta_n,L}) - f^{\delta_m}(s,Y_s^{\delta_m,L}))\\ &\le (Y_s^{\delta_n,L} - Y_s^{\delta_m,L})(f^{\delta_n}(s,Y_s^{\delta_n,L}) - f^{\delta_m}(s,Y_s^{\delta_n,L}))\\ &= (p-1)(Y_s^{\delta_m,L} - Y_s^{\delta_n,L})(Y_s^{\delta_n,L})^q \left(\frac{1}{(\eta_s \vee \delta_n)^{q-1}} - \frac{1}{(\eta_s \vee \delta_m)^{q-1}}\right)\\ &\le C\left(\frac{1}{(\eta_s \vee \delta_n)^{q-1}} - \frac{1}{(\eta_s \vee \delta_m)^{q-1}}\right)\end{aligned}$$

for all $s \in [0,T]$ and a constant $C > 0$. Taking expectations in Equation (5.13) yields

$$E\left[\int_0^T (Z_s^{\delta_n,L} - Z_s^{\delta_m,L})^2 ds\right] \le 2CE\left[\int_0^T \left(\frac{1}{(\eta_s \vee \delta_n)^{q-1}} - \frac{1}{(\eta_s \vee \delta_m)^{q-1}}\right) ds\right].$$

The sequence $\left(\frac{1}{(\eta \vee \delta_n)^{q-1}}\right)_{n\ge 0}$ converges in $\mathcal{M}^1(0,T)$ to $\frac{1}{\eta^{q-1}}$ as $n \to \infty$. This implies that $(Z^{\delta_n,L})_{n\ge 0}$ is a Cauchy sequence in $\mathcal{M}^2(0,T)$ and converges to $Z^L \in \mathcal{M}^2(0,T)$. In particular the random variable $\int_t^T Z_r^{\delta_n,L} dW_r$ converges to $\int_t^T Z_r^L dW_r$ in $L^2(\Omega)$ as $n \to \infty$. We obtain almost sure convergence by passing to a subsequence. Taking the limit $n \to \infty$ in

$$Y_t^{\delta_n,L} = \xi \wedge L - (p-1)\int_t^T \frac{(Y_r^{\delta_n,L})^q}{(\eta_r \vee \delta_n)^{q-1}} dr - \int_t^T Z_r^{\delta_n,L} dW_r,$$

and using estimate (5.12) yields that (Y^L, Z^L) satisfies almost surely the BSDE

$$Y_t^L = \xi \wedge L - (p-1)\int_t^T \frac{(Y_r^L)^q}{\eta_r^{q-1}} dr - \int_t^T Z_r^L dW_r.$$

We proceed by deriving the upper bound (5.10). We first estimate $Y^{\delta,L}$ against a linear BSDE with driver

$$g(t,y) = -p\frac{y}{T-t} + \frac{\eta_t \vee \delta}{(T-t)^p} + \gamma_t.$$

By using the inequality

$$(p-1)y^q - pa^{q-1}y + a^q \ge 0,$$

which holds for all $y \ge 0, a \ge 0$, one can show that $f^\delta(t,y) \le g(t,y)$ (take $a = (\eta_t \vee \delta)(T-t)^{-p/q}$). Let $\epsilon > 0$ and denote by Ψ^ϵ the solution process of the BSDE on $[0, T-\epsilon]$ with parameters $(g, Y_{T-\epsilon}^{\delta,L})$. By the solution formula for linear BSDEs we have

$$\Psi_t^\epsilon = E\left[\Gamma_{T-\epsilon} Y_{T-\epsilon}^{\delta,L} + \int_t^{T-\epsilon} \Gamma_s \left(\frac{\eta_s \vee \delta}{(T-s)^p} + \gamma_s\right) ds | \mathcal{F}_t\right],$$

where

$$\Gamma_t = \exp\left(-\int_0^t \frac{p}{T-s}ds\right) = \left(\frac{T-t}{T}\right)^p.$$

The comparison theorem implies

$$Y_t^{\delta,L} \le \Psi_t^\epsilon = \frac{1}{(T-t)^p} E\left[\epsilon^p Y_{T-\epsilon}^{\delta,L} + \int_t^{T-\epsilon} \left((\eta_s \vee \delta) + (T-s)^p \gamma_s\right) ds | \mathcal{F}_t\right] \tag{5.14}$$

for all $t \in [0, T]$ and $\epsilon > 0$. By letting $\epsilon \downarrow 0$ we obtain with dominated convergence

$$Y_t^{\delta,L} \le \frac{1}{(T-t)^p} E\left[\int_t^T \left((\eta_s \vee \delta) + (T-s)^p \gamma_s\right) ds | \mathcal{F}_t\right].$$

By letting $\delta \downarrow 0$ we obtain the upper bound in (5.10).

□

5.2.2. Existence of solutions for BSDEs with singular terminal condition

First we establish the convergence of (Y^L, Z^L) from Proposition 5.2.1 to a pair (Y, Z) which is a solution to the BSDE (5.8) with singular terminal condition $\liminf_{t\nearrow T} Y_t \ge \xi$ in the sense of Definition 5.1.1.

Theorem 5.2.2. *Assume* (**I1**) *and* (**I2**) *hold true. Let* (Y^L, Z^L) *be the solution to* (5.9) *from Proposition 5.2.1. Then there exists a process* (Y, Z) *such that for every* $0 \le t < T$ *the random variable* Y_t^L *converges a.s. to* Y_t *and* Z^L *converges in* $\mathcal{M}^2(0, t)$ *to* Z *as* $L \to \infty$. *The limit process* (Y, Z) *is a supersolution to the BSDE* (5.8) *with singular terminal condition* $Y_T = \xi$.

Proof. The proof is partly a generalization of the arguments in [69] to our setting. Appealing to the comparison theorem [62, Theorem 2.4] yields that $Y^L \le Y^N$ if $N > L$ (Observe that although assumption (ii) of [62, Theorem 2.4] is not satisfied here, the comparison holds, since the process α_t from the proof is nonpositive here as well). By Equation (5.10) for fixed $t < T$ the family of random variables $(Y_t^L, L \ge 0)$ is bounded from above as follows

$$Y_t^L \le \frac{1}{(T-t)^p} E\left[\int_t^T (\eta_s + (T-s)^p \gamma_s) ds | \mathcal{F}_t\right]. \tag{5.15}$$

Hence, for all $t < T$ we can define Y_t as the increasing limit of Y_t^L as $L \to \infty$. Notice that by Conditions (**I1**) and (**I2**) the random variable on the RHS of (5.15) is square integrable. By dominated convergence, therefore, Y_t^L converges to Y_t in $L^2(\Omega)$.

Moreover, the limit process Y satisfies $\liminf_{t\nearrow T} Y_t \geq \liminf_{t\nearrow T} Y_t^L = \xi \wedge L$ for every $L > 0$. Thus Y fulfills the terminal condition $\liminf_{t\nearrow T} Y_t \geq \xi$. For the convergence of (Z^L) let $0 \leq s \leq t < T$. Then Itô's formula implies, for $N, L \geq 0$,

$$(Y_s^N - Y_s^L)^2 + \int_s^t |Z_r^N - Z_r^L|^2 dr = (Y_t^N - Y_t^L)^2 - 2\int_s^t (Y_r^N - Y_r^L)(Z_r^N - Z_r^L)dW_r + 2\int_s^t (Y_r^N - Y_r^L)(f^N(r, Y_r^N) - f^L(r, Y_r^L))dr. \tag{5.16}$$

The monotonicity of the driver $f^L(r,y) = -(p-1)\frac{y^q}{\eta_t^{q-1}} + (\gamma_r \wedge L)$ in y yields for $y, y' \geq 0$

$$(y-y')(f^N(r,y) - f^L(r,y')) \leq (y-y')(f^N(r,y) - f^L(r,y)) = (y-y')(\gamma_r \wedge N - \gamma_r \wedge L),$$

and hence

$$(Y_s^N - Y_s^L)^2 + \int_s^t |Z_r^N - Z_r^L|^2 dr \leq (Y_t^N - Y_t^L)^2 - 2\int_s^t (Y_r^N - Y_r^L)(Z_r^N - Z_r^L)dW_r + 2\int_s^t (Y_r^N - Y_r^L)(\gamma_r \wedge N - \gamma_r \wedge L)dr. \tag{5.17}$$

Since Y^L and Y^N are bounded and $Z^L, Z^N \in \mathcal{M}^2(0,T)$, we have

$$E\left[\int_s^t (Y_r^N - Y_r^L)(Z_r^N - Z_r^L)dW_r\right] = 0.$$

Then estimate (5.17) implies

$$E\left[\int_0^t |Z_r^N - Z_r^L|^2 dr\right] \leq E\left[(Y_t^N - Y_t^L)^2] + 2\int_s^t (Y_r^N - Y_r^L)(\gamma_r \wedge N - \gamma_r \wedge L)dr\right], \tag{5.18}$$

and for a constant C_1

$$\begin{aligned} E\left[\sup_{0\leq s\leq t}(Y_s^N - Y_s^L)^2\right] &\leq E[(Y_t^N - Y_t^L)^2] + C_1 E\left[\sqrt{\int_0^t (Y_r^N - Y_r^L)^2 |Z_r^N - Z_r^L|^2 dr}\right] \\ &\quad +2E\left[\int_0^t (Y_r^N - Y_r^L)(\gamma_r \wedge N - \gamma_r \wedge L)dr\right], \end{aligned} \tag{5.19}$$

where we used the Burkholder-Davis-Gundy inequality. From Young's inequality we derive

$$\begin{aligned} & E\left[\sqrt{\int_0^t (Y_r^N - Y_r^L)^2 |Z_r^N - Z_r^L|^2 dr}\right] \\ \leq\ & E\left[\sup_{0\leq s\leq t}|Y_s^N - Y_s^L| \sqrt{\int_0^t |Z_r^N - Z_r^L|^2 dr}\right] \\ \leq\ & \frac{1}{4C_1} E\left[\sup_{0\leq s\leq t}(Y_s^N - Y_s^L)^2\right] + C_1 E\left[\int_0^t |Z_r^N - Z_r^L|^2 dr\right], \end{aligned}$$

which implies, together with (5.19) and (5.18),

$$\begin{aligned}\frac{3}{4}E\left[\sup_{0\le s\le t}(Y_s^N - Y_s^L)^2\right] &\le C_2 E[(Y_t^N - Y_t^L)^2] \\ &\quad +2C_2 E\left[\int_0^t (Y_r^N - Y_r^L)(\gamma_r\wedge N - \gamma_r \wedge L)dr\right],\end{aligned}$$

where $C_2 = 1 + C_1^2$. Again with Young's inequality we get

$$\begin{aligned}& E\left[\int_0^t (Y_r^N - Y_r^L)(\gamma_r\wedge N - \gamma_r \wedge L)dr\right] \\ \le\ & \frac{1}{4C_2}E\left[\sup_{0\le s\le t}(Y_s^N - Y_s^L)^2\right] + C_2 E\left[\left(\int_0^t |\gamma_r\wedge N - \gamma_r\wedge L|dr\right)^2\right].\end{aligned}$$

Finally we arrive at

$$E\left[\sup_{0\le s\le t}(Y_s^N - Y_s^L)^2\right] \le C_3 E\left[(Y_t^N - Y_t^L)^2 + \int_0^t (\gamma_r\wedge N - \gamma_r\wedge L)^2 dr\right], \tag{5.20}$$

for a constant $C_3 \ge 0$. The RHS of (5.20) converges to zero as $N,\ L \to \infty$. In particular, Inequality (5.18) implies that (Z^L) is a Cauchy sequence in $\mathcal{M}^2(0,t)$ and converges to $Z \in \mathcal{M}^2(0,t)$ for every $t < T$. Moreover, Inequality (5.20) yields that $E\left[\sup_{0\le s\le t} Y_s^2\right] < \infty$. Finally, taking the limit $L \nearrow \infty$ in

$$Y_s^L = Y_t^L - \int_s^t \left((p-1)\frac{(Y_r^L)^q}{\eta_r^{q-1}} - \gamma_r\right)dr - \int_s^t Z_r^L dW_r$$

implies that Y satisfies (5.8) for every $0 \le s \le t < T$. □

Proposition 5.2.3. *The supersolution Y obtained in Theorem 5.2.2 is minimal: If (Y', Z') is another nonnegative supersolution of* (5.8) *with singular terminal condition $Y'_T = \xi$, then $Y'_t \ge Y_t$ a.s. for all $t \in [0,T]$.*

Proof. The proof is an adaptation of [70, Theorem 7] to our setting.

Fix $L > 0$ and let (Y^L, Z^L) denote the solution of (5.9) with terminal condition $Y_T^L = \xi \wedge L$. Let (Y', Z') be a nonnegative solution of (5.8) in the sense of Definition 5.1.1. Set $\Delta_t = Y'_t - Y_t^L$, $\Gamma_t = Z'_t - Z_t^L$ and

$$\alpha_t = \begin{cases} \dfrac{p-1}{\eta_t^{q-1}}\dfrac{(Y'_t)^q - (Y_t^L)^q}{Y'_t - Y_t^L}, & \text{if } \eta_t^{q-1}(Y'_t - Y_t^L) \ne 0 \\ 0, & \text{else.}\end{cases}$$

Note that α is nonnegative. Moreover, we have $\alpha_t \le \frac{(p-1)q}{\eta_t^{q-1}}((Y'_t)^{q-1} \vee (Y_t^L)^{q-1})$. This implies that $\int_0^t \alpha_s ds < \infty$ a.s. for all $t < T$. For every $t < T$ the process (Δ, Γ) solves the linear BSDE

$$d\Delta_s = [\alpha_s \Delta_s - (\gamma_s - L)^+]ds + \Gamma_s dW_s$$

on $[0,t]$ with terminal condition $\Delta_t = Y'_t - Y^L_t$. Hence, by Lemma 5.5.1 in the Appendix the solution Δ admits the explicit representation

$$\Delta_s = E\left[\Delta_t e^{-\int_s^t \alpha_r dr} + \int_s^t e^{-\int_s^u \alpha_r dr}(\gamma_u - L)^+ du|\mathcal{F}_s\right].$$

Since Y' is nonnegative and $Y^L \le (1+T)L$ by Proposition 5.2.1, we have $\Delta_t \ge -(1+T)L$. Thus $\Delta_t e^{-\int_s^t \alpha_r dr}$ is bounded from below by $-(1+T)L$ and we can apply Fatou's lemma to obtain

$$\begin{aligned} Y'_s - Y^L_s &= \Delta_s = \liminf_{t\nearrow T} E\left[\Delta_t e^{-\int_s^t \alpha_r dr} + \int_s^t e^{-\int_s^u \alpha_r dr}(\gamma_u - L)^+ du|\mathcal{F}_s\right] \\ &\ge E\left[\liminf_{t\nearrow T} \Delta_t e^{-\int_s^t \alpha_r dr}|\mathcal{F}_s\right] \ge 0. \end{aligned}$$

Finally, taking the limit $L \nearrow \infty$ yields the claim. □

5.3. Optimal controls

In this section we first consider a variant of the minimization problem (5.5), where we omit the constraint $x_T = 0$ on A in the set of admissible controls but penalize any nonzero terminal state by $(\xi \wedge L)|x_T|^p$. We show that optimal controls for this unconstrained minimization problem admit a representation in terms of the solutions Y^L from Proposition 5.2.1. We then use this result to derive an optimal control for (5.5).

Throughout this section we assume (**I1**) and (**I2**) without further mentioning it.

5.3.1. Penalization

In this section we consider the unconstrained minimization problem

$$v^L = \inf_{x\in\mathcal{A}} J^L(x) = \inf_{x\in\mathcal{A}} E\left[\int_0^T \big(\eta_t|\dot{x}_t|^p + (\gamma_t \wedge L)|x_t|^p\big)dt + \xi \wedge L|x_T|^p\right] \tag{5.21}$$

for some $L > 0$, where we take the infimum over $\mathcal{A}$, the set of all progressively measurable processes $x : \Omega \times [0,T] \to \mathbb{R}$ with absolutely continuous sample paths starting in $x_0 = \zeta$. Next, we show how to obtain a minimizing control for (5.21) from the solution Y^L to (5.9).

Proposition 5.3.1. *Let* (Y^L, Z^L) *be the solution to* (5.9) *from Proposition 5.2.1. Then*

$$x^L_t = \zeta e^{-\int_0^t \left(\frac{Y^L_s}{\eta_s}\right)^{q-1} ds}$$

is optimal in (5.21) *and we have* $v^L = Y^L_0|\zeta|^p$.

Proof. To simplify notation we assume $\zeta = 1$ and set $\gamma_t^L = \gamma_t \wedge L$. Let $g(z) = |z|^p$ and $M_t = pY_t^L(x_t^L)^{p-1} + p\int_0^t \gamma_s^L(x_s^L)^{p-1}dt$. Applying the integration by parts formula to M results in

$$\begin{aligned} dM_t &= p(x_t^L)^{p-1}dY_t^L + p(p-1)Y_t^L(x_t^L)^{p-2}dx_t^L + p\gamma_t^L(x_t^L)^{p-1}dt \\ &= p(x_t^L)^{p-1}Z_t^L dW_t. \end{aligned}$$

Since x^L is bounded and $Z^L \in \mathcal{M}^2(0,T)$, the process M is a martingale. Let $x \in \mathcal{A}$ and introduce $\theta_t = x_t^L - x_t$. Then θ satisfies $\theta_0 = 0$. Similar considerations as in Lemma 5.1.6 imply that we can assume that x is pathwise nonincreasing and hence $|\theta_t| \leq 2$. Furthermore, we have $\eta_t g'(\dot{x}_t^L) = -p\eta_t|\dot{x}_t^L|^{p-1} = -pY_t^L(x_t^L)^{p-1}$. The convexity of g implies for all $t \in [0,T]$

$$g(\dot{x}_t^L) - g(\dot{x}_t) \leq g'(\dot{x}_t^L)(\dot{x}_t^L - \dot{x}_t).$$

Thus, it follows from integration by parts

$$\begin{aligned} \int_0^T \eta_t(g(\dot{x}_t^L) - g(\dot{x}_t))dt &\leq \int_0^T \eta_t g'(\dot{x}_t^L)d\theta_t = \int_0^T \left(\int_0^t p\gamma_s^L x_s^{p-1}ds - M_t\right)d\theta_t \\ &= -(\xi \wedge L)g'(x_T^L)\theta_T + \int_0^T \theta_t dM_t - \int_0^T \gamma_t^L g'(x_t)\theta_t dt \end{aligned}$$

Since M is a square integrable martingale and θ is bounded, we obtain $E\left[\int_0^T \theta_t dM_t\right] = 0$. Using convexity of g once more, we obtain

$$g(x_t^L) - g(x_t) \leq g'(x_t^L)(x_t^L - x_t).$$

This implies optimality of x^L:

$$\begin{aligned} & E\left[\int_0^T \eta_t(g(\dot{x}_t^L) - g(\dot{x}_t))dt\right] \\ \leq\ & -E\left[(\xi \wedge L)g'(x_T^L)\theta_T + \int_0^T \gamma_t^L g'(x_t)\theta_t dt\right] \\ \leq\ & -E\left[(\xi \wedge L)(g(x_T^L) - g(x_T)) + \int_0^T \gamma_t^L(g(x_t^L) - g(x_t))dt\right]. \end{aligned}$$

It remains to verify the identity $v^L = Y_0^L$. To this end we apply the integration by parts formula to the process $Y(x^L)^p$ to obtain

$$d(Y(x^L)^p)_t = -\left(\frac{(Y_t^L)^q}{\eta_t^{q-1}}(x_t^L)^p + \gamma_t^L(x_t^L)^p\right)dt + (x_t^L)^p Z_t^L dW_t.$$

Moreover we have

$$|\dot{x}_t^L|^p = \left(\left(\frac{Y_t^L}{\eta_t}\right)^{q-1} x_t^L\right)^p = \left(\frac{Y_t^L}{\eta_t}\right)^q (x_t^L)^p.$$

Thus we obtain

$$Y_0^L = E\left[\int_0^T \eta_t |\dot{x}_t^L|^p + \gamma_t^L |x_t^L|^p dt + (\xi \wedge L)|x_T^L|^p\right] = J^L(x^L) = v^L.$$

□

5.3.2. The constrained case

We now turn to the constrained case and prove Theorem 5.1.3. For the reader's convenience we restate the theorem here.

Theorem 5.3.2. *Let* (Y, Z) *be the minimal supersolution to* (5.8) *with singular terminal condition* $Y_T = \xi$ *from Theorem 5.2.2. Then* $v = Y_0|\zeta|^p$*; moreover the control* $x_t = \zeta \exp\left(-\int_0^t \left(\frac{Y_s}{\eta_s}\right)^{q-1} ds\right)$ *belongs to* $\mathcal{A}_0$ *and is optimal in* (5.5).

Proof. To simplify notation assume that $\zeta = 1$. As in the proof of Proposition 5.3.1 we introduce $M_t = pY_t x_t^{p-1} + p\int_0^t \gamma_s x_s^{p-1} dt$. Performing integration by parts yields

$$dM_t = x_t^{p-1} Z_t dW_t.$$

Hence, M is a nonnegative local martingale on $[0, T)$ and in particular a nonnegative super-martingale. Thus it converges almost surely in $\mathbb{R}$ as $t \nearrow T$. Since Y satisfies the terminal condition $\liminf_{t\nearrow T} Y_t = \infty$ on A we have that

$$0 \le x_t = \left(\frac{M_t - p\int_0^t \gamma_s x_s^{p-1} ds}{pY_t}\right)^{q-1} \le \left(\frac{M_t}{pY_t}\right)^{q-1} \to 0 \tag{5.22}$$

a.s. on A for $t \nearrow T$. It follows that $x_T = 0$ on A and hence $x \in \mathcal{A}_0$.

Next we apply the integration by parts formula to the process Yx^p to obtain

$$d(Yx^p)_t = -\left(\eta_t |\dot{x}_t|^p + \gamma_t x_t^p\right) dt + x_t^p Z_t dW_t.$$

Since $Z \in \mathcal{M}^2(0, t)$ and $|x_t| \le 1$ we can deduce for $t < T$

$$Y_0 = E\left[\int_0^t \left(\eta_s |\dot{x}_s|^p + \gamma_s x_s^p\right) ds + Y_t x_t^p\right]$$

From Inequality (5.22) we know that $0 \le pY_t x_t^{p-1} \le M_t$. Since M converges a.s. in $\mathbb{R}$ and $x_t \to 0$ on A it follows that $Y_t x_t^p \to 0$ on A as $t \nearrow T$. On A^c we have $\liminf_{t\nearrow T} Y_t x_t^p \ge \xi x_T^p$. This yields $\liminf_{t\nearrow T} Y_t x_t^p \ge 1_{A^c} \xi x_T^p$ a.s. Fatou's lemma implies

$$Y_0 \ge E\left[\int_0^T \left(\eta_s |\dot{x}_s|^p + \gamma_s x_s^p\right) ds + 1_{A^c} \xi x_T^p\right] = J(x). \tag{5.23}$$

Next, note that for every $\overline{x} \in \mathcal{A}_0$ we have $J(\overline{x}) \geq J^L(\overline{x})$. This implies $v \geq v^L$ for every $L > 0$. By Proposition 5.3.1 we have $Y_0^L = v^L$. Minimality of Y implies $Y_0 = \lim_{L \nearrow \infty} Y_0^L = \lim_{L \nearrow \infty} v^L \leq v$. Consequently we obtain with Inequality (5.23)

$$Y_0 \geq J(x) \geq v \geq Y_0$$

and thus optimality of x. □

Remark 5.3.3. The solution Y from Theorem 5.2.2 does not only lead to optimal controls in the case where the liquidation period begins at time $t = 0$. If liquidation starts at an arbitrary time $t < T$ the minimization problem reads

$$V_t = \operatorname{essinf} E\left[\int_t^T \left(\eta_s|\dot{\tilde{x}}_s|^p + \gamma_s|\tilde{x}_s|^p\right)ds + 1_{A^c}\xi|x_T|^p \middle| \mathcal{F}_t\right],$$

where the infimum is taken over all progressively measurable processes $\tilde{x}$ starting in a $\mathcal{F}_t$-measurable random variable ζ and ending in 0 on A. In this case the optimal control is given by

$$x_s = \zeta \exp\left(-\int_t^s \left(\frac{Y_r}{\eta_r}\right)^{q-1} dr\right)$$

and the value is equal to $V_t = Y_t|\zeta|^p$.

In the next proposition we state an integrability condition that allows to identify the minimal supersolution of (5.8) with terminal condition $\xi = \infty$. We employ this result in Section 5.4.

Proposition 5.3.4. *Let* (Y, Z) *be a nonnegative supersolution of* (5.8) *with singular terminal condition* $Y_T = \infty$. *Let* $x_t = \exp\left(-\int_0^t \left(\frac{Y_s}{\eta_s}\right)^{q-1} ds\right)$ *denote the associated position path and assume that* $x^{p-1}Z \in \mathcal{M}^2(0,T)$. *Then* Y *is the minimal supersolution of* (5.8).

Proof. Let $Y^{\min}$ denote the minimal supersolution of (5.8). Without loss of generality we only consider the point in time $t = 0$ and show that $Y_0 = Y_0^{\min}$. For general $t < T$ we refer to Remark 5.3.3 which shows that $Y_t^{\min}$ is the value of the minimization problem starting in time t. We proceed as in the proof of Theorem 5.3.2. Let $M_t = pY_t x_t^{p-1} + p\int_0^t \gamma_s x_s^{p-1} dt$. Then we obtain by integration by parts

$$dM_t = x_t^{p-1} Z_t dW_t.$$

Hence, M is a nonnegative true martingale with $E[M_T^2] < \infty$ and converges a.s. in $\mathbb{R}$ as $t \nearrow T$. Since Y satisfies the terminal condition $\liminf_{t \nearrow T} Y_t = \infty$ we have that $x_t \to 0$ as $t \nearrow T$. Consequently, $x \in \mathcal{A}_0$ and Lemma 5.1.7 implies optimality of x. Again an application of the integration by parts formula yields

$$d(Yx^p)_t = (\eta_t|\dot{x}_t|^p + \gamma_t x_t^p)dt + x_t^p Z_t dW_t.$$

By assumption the process $t \mapsto \int_0^t x_t^p Z_t dW_t$ is a true martingale. Moreover we have $\lim_{t \nearrow T} Y_t x_t^p = 0$ and hence Theorem 5.3.2 implies $Y_0 = J(x) = v = Y_0^{\min}$. □

5.4. Processes with uncorrelated multiplicative increments

In this section we study the special case of the control problem (5.5) where $\gamma = 0$, $\xi = \infty$ and η has uncorrelated multiplicative increments. We first give a rigorous definition of what the latter means.

We say that a positive, progressively measurable process η has uncorrelated multiplicative increments if $E\left[\frac{\eta_t}{\eta_s}|\mathcal{F}_s\right] = E\left[\frac{\eta_t}{\eta_s}\right]$ for all $s \le t < T$. We show that it is precisely this class of processes which leads to *deterministic* optimal controls for the minimization problem (5.5) (with $\gamma = 0$). Moreover we show that if η is a martingale, then it is optimal to close the position at a constant rate.

Observe that any process η where $\frac{\eta_t}{\eta_s}$ is independent of $\mathcal{F}_s$ for $s \le t < T$ has uncorrelated multiplicative increments. The converse does not hold true.

In the next lemma we give an equivalent characterization of processes with uncorrelated multiplicative increments.

Lemma 5.4.1. *A positive, progressively measurable process η has uncorrelated multiplicative increments if and only if the process $\left(\frac{\eta_t}{E[\eta_t]}\right)_{t<T}$ is a martingale. Any such process satisfies $E\left[\frac{\eta_t}{\eta_s}\right] = \frac{E[\eta_t]}{E[\eta_s]}$ for all $s \le t < T$.*

Proof. Let η have uncorrelated multiplicative increments. We first show that for $s \le t < T$ any such η satisfies $E\left[\frac{\eta_t}{\eta_s}\right] = \frac{E[\eta_t]}{E[\eta_s]}$. Indeed, we have

$$E[\eta_t] = E\left[\eta_s E\left[\frac{\eta_t}{\eta_s}\middle|\mathcal{F}_s\right]\right] = E[\eta_s]E\left[\frac{\eta_t}{\eta_s}\right].$$

Next let $M_t = \frac{\eta_t}{E[\eta_t]}$ for $t < T$. For $s \le t < T$ the process M satisfies

$$E[M_t|\mathcal{F}_s] = \frac{1}{E[\eta_t]}E[\eta_t|\mathcal{F}_s] = \frac{1}{E[\eta_t]}E\left[\frac{\eta_t}{\eta_s}\eta_s\middle|\mathcal{F}_s\right] = \frac{\eta_s}{E[\eta_t]}E\left[\frac{\eta_t}{\eta_s}\right] = M_s.$$

For the converse direction, let $M_t = \frac{\eta_t}{E[\eta_t]}$ be a martingale for $t < T$. Then we have for $s \le t < T$

$$E[\eta_t|\mathcal{F}_s] = E[\eta_t]E[M_t|\mathcal{F}_s] = E[\eta_t]M_s = \frac{E[\eta_t]}{E[\eta_s]}\eta_s.$$

Thus the random variable $E\left[\frac{\eta_t}{\eta_s}|\mathcal{F}_s\right]$ is deterministic, which implies $E\left[\frac{\eta_t}{\eta_s}|\mathcal{F}_s\right] = E\left[\frac{\eta_t}{\eta_s}\right]$. □

Lemma 5.4.1 implies that any positive martingale has uncorrelated multiplicative increments. Further examples are provided by the following class of diffusions.

Example 5.4.2. *Let η be a diffusion with linear drift, i.e. η solves*

$$d\eta_t = \mu(t)\eta_t dt + \sigma(t, \eta_t)dW_t,$$

where the drift μ is a deterministic function of time and the stochastic volatility $\sigma : [0,T] \times \mathbb{R} \times \Omega \to \mathbb{R}_+$ is such that $t \mapsto \sigma(t, \eta_t) \in \mathcal{M}^2(0,T)$. Then the process

$$t \mapsto \eta_t \exp(-\int_0^t \mu(r)dr)$$

is a martingale, and hence we have $E[\eta_t|\mathcal{F}_s] = \eta_s \exp(\int_s^t \mu(r)dr)$. This implies that the random variable $E\left[\frac{\eta_t}{\eta_s}|\mathcal{F}_s\right]$ is deterministic. Therefore η has uncorrelated multiplicative increments.

We first show that if the optimal control from Theorem 5.3.2 is deterministic, then the process η has necessarily uncorrelated multiplicative increments.

Proposition 5.4.3. *Let η be positive, progressively measurable and such that $\eta \in \mathcal{M}^2(0,T)$, $1/\eta^{q-1} \in \mathcal{M}^1(0,T)$. Assume that the optimal control $x \in \mathcal{A}_0$ from Theorem 5.3.2 is deterministic. Then η has uncorrelated multiplicative increments.*

Proof. The optimal control from Theorem 5.3.2 satisfies $\dot{x}_t = -\left(\frac{Y_t}{\eta_t}\right)^{q-1} x_t$ where Y is the minimal solution of (5.8) with singular terminal condition $Y_T = \infty$. Since x is deterministic it follows that the nonnegative process $\alpha_t = \left(\frac{Y_t}{\eta_t}\right)^{q-1}$ is deterministic as well. Moreover, we have $\int_0^t \alpha_s ds < \infty$ for all $t < T$. The process Y satisfies the linear BSDE

$$dY_t = (p-1)\alpha_t Y_t dt + Z_t dW_t$$

and hence Lemma 5.5.1 in the Appendix implies for $s \le t < T$

$$\alpha_s^{p-1}\eta_s = Y_s = E\left[Y_t e^{-\int_s^t (p-1)\alpha_r dr}|\mathcal{F}_s\right] = \alpha_t^{p-1} e^{-\int_s^t (p-1)\alpha_r dr} E[\eta_t|\mathcal{F}_s].$$

Consequently, the random variable $E\left[\frac{\eta_t}{\eta_s}|\mathcal{F}_s\right]$ is deterministic for all $s \le t < T$ and hence η has uncorrelated multiplicative increments. □

We next show that the converse of Proposition 5.4.3 holds true as well: If η has uncorrelated multiplicative increments, then there exists an deterministic optimal control for (5.5). Here and in the sequel we assume that $\zeta = x_0 = 1$

Proposition 5.4.4. *Assume that η has uncorrelated multiplicative increments and satisfies the integrability assumptions* (**I1**) *and $\eta_T \in L^2(\Omega)$. Then*

$$Y_t = \frac{1}{\left(\int_t^T \frac{1}{E[\eta_s|\mathcal{F}_t]^{q-1}} ds\right)^{p-1}}$$

is the minimal solution to (5.8) *with singular terminal condition. The deterministic control*

$$x_t = \frac{1}{\int_0^T \frac{1}{E[\eta_s]^{q-1}} ds} \int_t^T \frac{1}{E[\eta_s]^{q-1}} ds$$

is optimal in (5.5). *In particular the optimal control rate is inversely proportional to* $E[\eta_t]^{q-1}$.

Proof. First note that we have by Jensen's inequality

$$\int_t^T \frac{1}{E[\eta_s|\mathcal{F}_t]^{q-1}} ds \geq (T-t)^q \frac{1}{\left(\int_t^T E[\eta_s|\mathcal{F}_t] ds\right)^{q-1}}.$$

This implies that Y is bounded from above as follows

$$Y_t \leq \frac{1}{(T-t)^p} E\left[\int_t^T \eta_s ds | \mathcal{F}_t\right]. \tag{5.24}$$

Next we use the fact from Lemma 5.4.1 that $E[\eta_s|\mathcal{F}_t] = \eta_t E\left[\frac{\eta_s}{\eta_t}\right] = \eta_t \frac{E[\eta_s]}{E[\eta_t]}$ for $s \geq t$ to rewrite Y as

$$Y_t = M_t \frac{1}{\left(\int_t^T \frac{1}{E[\eta_s]^{q-1}} ds\right)^{p-1}}$$

where the process M denotes the martingale $M_t = \frac{\eta_t}{E[\eta_t]}$. Moreover, we have by assumption $E[M_T^2] = E[\eta_T^2]/E[\eta_T]^2 < \infty$. Hence, M is a square integrable martingale. Let $\phi \in \mathcal{M}^2(0,T)$ denote the integrand from its martingale representation. Then we obtain, by integration by parts,

$$\begin{aligned} dY_t &= (p-1)\frac{1}{E[\eta_t]^{q-1}} \frac{M_t}{\left(\int_t^T \frac{1}{E[\eta_s]^{q-1}} ds\right)^p} dt + \frac{\phi_t}{\left(\int_t^T \frac{1}{E[\eta_s]^{q-1}} ds\right)^{p-1}} dW_t \\ &= (p-1)\frac{Y_t^q}{\eta_t^{q-1}} dt + Z_t dW_t, \end{aligned}$$

with

$$Z_t = \frac{\phi_t}{\left(\int_t^T \frac{1}{E[\eta_s]^{q-1}} ds\right)^{p-1}}. \tag{5.25}$$

Hence, we have $Z \in \mathcal{M}^2(0,t)$ for every $t < T$. An application of the Burkholder-Davis-Gundy inequality as in the proof of Theorem 5.2.2 in combination with Inequality (5.24) yields $E[\sup_{0\leq s\leq t} Y_s^2] < \infty$ for all $t < T$. Hence, (Y,Z) is a solution to (5.8) with singular terminal condition $Y_T = \infty$.

The associated path x satisfies

$$x_t = \exp\left(-\int_0^t \left(\frac{Y_s}{\eta_s}\right)^{q-1} ds\right) = \exp\left(-\int_0^t \frac{1}{E[\eta_s]^{q-1}\int_t^T \frac{1}{E[\eta_r]^{q-1}}dr}ds\right)$$
$$= \frac{1}{\int_0^T \frac{1}{E[\eta_s]^{q-1}}ds}\int_t^T \frac{1}{E[\eta_s]^{q-1}}ds.$$

In particular it follows from (5.25) that $x^{p-1}Z \in \mathcal{M}^2(0,T)$ and hence Proposition 5.3.4 yields that Y is the minimal solution of (5.8). Theorem 5.3.2 then implies optimality of x. □

If η is monotone in expectation, then we obtain the following result about the path of the optimal control.

Corollary 5.4.5. *Let η satisfy the assumptions of Proposition 5.4.4. If the mapping $t \mapsto E[\eta_t]$ is nondecreasing (nonincreasing), then the optimal control $x \in \mathcal{A}_0$ from Proposition 5.4.4 is a convex (concave) function of time.*

Proof. The optimal control rate from Proposition 5.4.4 is given by $\dot{x}_t = -\frac{1}{cE[\eta_t]^{q-1}}$ with $c = \int_0^T \frac{1}{E[\eta_s]^{q-1}}ds$. In particular $t \mapsto \dot{x}_t$ is nondecreasing (nonincreasing) if $t \mapsto E[\eta_t]$ is nondecreasing (nonincreasing). □

Proposition 5.4.4 includes the case where η is a martingale as a special case.

Corollary 5.4.6. *Let η be a positive martingale satisfying $1/\eta^{q-1} \in \mathcal{M}^1(0,T)$ and $\eta_T \in L^2(\Omega)$. Then $Y_t = \frac{\eta_t}{(T-t)^{p-1}}$ solves the BSDE (5.8) with singular terminal condition $Y_T = \infty$ and the control with constant control rate $x_t = 1 - \frac{t}{T}$ is optimal in (5.5).*

Proof. The process η^2 is a submartingale and hence $E[\eta_t^2] \le E[\eta_T^2]$ for all $t \le T$, which implies that $\eta \in \mathcal{M}^2(0,T)$. Moreover, Lemma 5.4.1 yields that η has uncorrelated multiplicative increments. Hence, all assumptions of Proposition 5.4.4 are satisfied which yields the claim. □

Another special case of Proposition 5.4.4 is the case where η is a deterministic function of time.

Corollary 5.4.7. *Assume that η is deterministic and satisfies $1/\eta^{q-1} \in L^1([0,T])$, $\eta \in L^2([0,T])$ and $\eta_T < \infty$. Then*

$$Y_t = \left(\frac{1}{\int_t^T \frac{1}{\eta_s^{q-1}}ds}\right)^{p-1}$$

solves (5.8) with singular terminal condition $Y_T = \infty$ and the control

$$x_t = \frac{\int_t^T \frac{1}{\eta_s^{q-1}}ds}{\int_0^T \frac{1}{\eta_s^{q-1}}ds} \tag{5.26}$$

is optimal in (5.5).

Remark 5.4.8. The results about the optimal control in Corollary 5.4.6 and Corollary 5.4.7 hold also true under weaker assumptions on the process η. In the martingale case it suffices to assume that η is a positive martingale with $E[\eta_T^2] < \infty$. Then Proposition 5.1.7 directly implies that the control with constant rate is optimal. In the deterministic case it is straightforward to show that under the integrability condition $1/\eta^{q-1} \in L^1([0,T])$ the function $\eta|\dot{x}|^{p-1}$ is constant for the control x from Equation (5.26). Then again Proposition 5.1.7 implies optimality of x.

A particular example for a process with uncorrelated multiplicative increments is the geometric Brownian motion.

Example 5.4.9. *Assume that η evolves according to a geometric Brownian motion*

$$d\eta_t = \mu\eta_t dt + \sigma\eta_t dW_t$$

with drift $\mu \in \mathbb{R}$, volatility $\sigma > 0$ and initial value $\eta_0 > 0$. In this case

$$\frac{\eta_t}{\eta_s} = e^{\left(\mu - \frac{\sigma^2}{2}\right)(t-s) + \sigma(W_t - W_s)}$$

for $s \le t \le T$ and hence η has uncorrelated multiplicative increments. Moreover we have $E[\eta_t|\mathcal{F}_s] = \eta_s e^{\mu(t-s)}$ and η satisfies the integrability conditions $\eta \in \mathcal{M}^2(0,T)$, $E[\eta_T^2 < \infty]$ and $\int_t^T \frac{1}{E[\eta_s]^{q-1}} ds < \infty$. In the case $\mu = 0$ the price impact process η is a martingale and Corollary 5.4.6 yields that linear closure is optimal in (5.5). *In the case $\mu \neq 0$ Proposition 5.4.4 implies that a solution of* (5.8) *is given by*

$$Y_t = \mu(q-1)^{p-1} \frac{\eta_t}{\left(1 - e^{-\mu(q-1)(T-t)}\right)^{p-1}}$$

and that the optimal control for (5.5) *satisfies*

$$x_t = \frac{e^{-\mu(q-1)t} - e^{-\mu(q-1)T}}{1 - e^{-\mu(q-1)T}}.$$

5.5. Appendix

The next result provides a uniqueness result for solutions of linear BSDEs under mild assumptions on the coefficients.

Lemma 5.5.1. *Let $(\alpha_t)_{0\le t\le T}$ and $(\beta_t)_{0\le t\le T}$ be progressively measurable processes and ξ a $\mathcal{F}_T$-measurable, square integrable random variable. Assume that α is bounded from below and that the integrals $\int_0^t \alpha_s ds$ and $\int_0^t |\beta_s| ds$ are almost surely finite for every $t \in [0,T]$. Any solution (Y,Z) with $Z \in \mathcal{M}^2(0,T)$ to the linear BSDE*

$$dY_t = (\alpha_t Y_t + \beta_t)\, dt + Z_t dW_t$$

with $Y_T = \xi$ admits the representation

$$Y_t = E\left[\xi e^{-\int_t^T \alpha_s ds} - \int_t^T e^{-\int_t^s \alpha_u du} \beta_s ds \middle| \mathcal{F}_t\right].$$

Proof. Let (Y, Z) be a solution. Set

$$\varphi_t = Y_t e^{-\int_0^t \alpha_s ds} - \int_0^t e^{-\int_0^s \alpha_u du} \beta_s ds.$$

Then by integration by parts we obtain

$$d\varphi_t = e^{-\int_0^t \alpha_s ds} Z_t dW_t.$$

Since α is bounded from below and $Z \in \mathcal{M}^2(0, T)$, the integrand belongs to $\mathcal{M}^2(0, T)$ as well. Therefore φ is a martingale and consequently

$$\varphi_t = E[\varphi_T | \mathcal{F}_t] = E\left[\xi e^{-\int_0^T \alpha_s ds} - \int_0^T e^{-\int_0^s \alpha_u du} \beta_s ds | \mathcal{F}_t\right],$$

which yields the claim. □

6. Hedging forward positions: basis risk versus liquidity costs

The aim of this chapter is to analyze the impact of cross-hedging opportunities on liquidation strategies. On forward markets it happens frequently that liquidity increases as time to delivery approaches. The reader is referred to the introduction of this thesis for an illustrating example from energy markets. An immediate closure of an open forward position therefore implies foregoing the option of reducing execution costs. Often there is a proxy market, where a correlated asset is liquidly traded. This proxy market therefore offers the opportunity of cross-hedging the risk inherent in an open position and to reduce liquidity costs. However, prices in the primary and the proxy market are not perfectly correlated. Therefore cross-hedging the open position in a proxy market entails basis risk.

This chapter aims at describing the optimal trade-off between minimizing basis risk and minimizing execution costs. The goal is to provide *simple* and *explicit decision rules* that can guide practitioners in hedging their risk. To this end some simplifying assumptions are made. *Liquidity costs* are interpreted as half of the bid-ask spread. In other words, the liquidity costs for selling, respectively, buying one asset share are equal to the absolute difference of the realized price to the midmarket price. The bid-ask spread is an exogenously given stochastic process. In contrast to Chapter 4 it does *not* depend on the order size. Allowing in addition for a volume-dependent price impact would make it difficult to obtain explicit hedging strategies; one would have to fall back on numerical methods.

There is an investor that has to close a short position in an illiquid asset within a given time horizon $[0, T]$; in addition she can hedge a part of her market risk by investing in a positively correlated second (more liquid) asset. In order to reflect the interpretation above, the *stochastic* liquidity costs are proportional to the amount traded. The risk costs of the investor are given via a nondecreasing function of a quadratic form taking into account the diversification of the portfolio.[1] The aim of the investor is to minimize an additive functional of liquidity and risk costs. Finding optimal position paths in the two markets is formulated as a *singular control problem* (Section 6.1).

Section 6.2 draws on the well-known connection between singular control and optimal stopping, see e.g. [43],[14],[36] and the references therein. However, the one-dimensional results from the literature cannot always be directly applied to the two-dimensional

[1]This includes, e.g., the case where we can interpret the risk costs as the time average value-at-risk of the portfolio. For another choice, the risk costs correspond approximately to the variance of the portfolio for deterministic trading strategies.

optimization problem of *simultaneously* finding the optimal illiquid asset *and* the cross hedge position process. Hence a successive method for reducing the two-dimensional optimization problem to one-dimensional problems is developed. By this means, it is shown that the problem of finding optimal position strategies is equivalent to a family of stopping problems.

The method is applied to three *stylized case studies* in which optimal hedges are derived *explicitly*. In the first case study (see Section 6.3) trading becomes suddenly active at a random time $\tilde{\tau}$: before $\tilde{\tau}$ liquidity costs are constant equal to a high level K^+, and after $\tilde{\tau}$ are equal to a lower level K^-. The liquidity jump is modeled as the first jump time of a Poisson process. In the second case study (see Section 6.4) liquidity costs are *deterministic* nonincreasing functions. The influence of the speed with which liquidity costs decay is studied. Increasing and decreasing speed are distinguished; in other words, costs that are concave, respectively, convex over time. Finally, in Section 6.5, the risk costs are given by a Brownian bridge in order to illustrate the method for a model with more complex stochastic risk costs. The proofs of the results of Sections 6.3, 6.4 and 6.5 are presented in Appendices 6.6.1, 6.6.2 and 6.6.3, respectively.

6.1. A model with stochastic liquidity

Consider an agent aiming at closing a short forward position of an *illiquid asset* (e.g. German natural gas as in the example of the introduction). We suppose that there is an OTC forward market, where one can buy and sell the asset. We further assume that there exists a more standardized and liquidly traded asset that is highly correlated with the asset to be hedged. We will refer to the illiquid asset as the *primary asset*, and to the liquid asset as the *proxy* of the illiquid asset.

Let $x_0 < 0$ be the initial short position of the primary asset. We assume that the agent has to close the position at the latest at time $T > 0$. The agent has the choice between buying the illiquid asset on the forward market *before* time T or on the spot market at time T. The spot price may also involve some transaction costs. We denote by K_t the costs arising from a closure of one unit at $t \in [0,T]$. We assume that K is a nonnegative adapted stochastic process on a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]})$, where $\mathbb{F}$ satisfies the usual conditions of right-continuity and completeness. Moreover we suppose that the paths of K are càdlàg on $[0,T]$ (i.e. they are right-continuous and possess left-hand limits). In the particular case studies, the dynamics of K and the filtration $\mathbb{F}$ will be specified in more detail.

The proxy is assumed to be liquidly traded. Nevertheless, any acquisition of the proxy will entail liquidity costs. For simplicity we assume that liquidity of the proxy is constant over the period $[0,T]$, and we denote by $L \in \mathbb{R}_+$ half the bid ask spread. In addition we assume that the agent must not have any proxy position at time T (e.g. because of a physical settlement).[2]

[2]Cf. the motivating example about different gas market areas in Europe outlined in the introduction: the forward position in Dutch gas has to be closed before the gas is physically delivered. Our model does not treat the case where forwards are cash settled, i.e. where the proxy position does not need

By a *position process* of the primary asset we mean any $\mathbb{F}$-adapted process $X : [0,T] \times \Omega \to \mathbb{R}$ that is càdlàg and satisfies $X_T = 0$. Analogously, a proxy position $Y : [0,T] \times \Omega \to \mathbb{R}$ is a càdlàg $\mathbb{F}$-adapted process satisfying $Y_T = 0$. We suppose that the initial cross hedge position, the proxy position, is zero. We define $X_{0-} = x_0$ and $Y_{0-} = 0$. Any pair (X, Y) satisfying the properties above will be referred to as a *position strategy*. The set of all position strategies will be denoted by $\mathcal{D}(x_0)$.

The overall execution costs entailed by a strategy (X, Y) are given by

$$C(X,Y) = \int_{[0,T]} K_s |dX_s| + L \int_{[0,T]} |dY_s|,$$

where $|dX_s|$ denotes the integral with respect to the total variation of the path X over the whole interval $[0,T]$. Note that the integral includes the boundary of the interval $[0,T]$, which means that

$$\int_{[0,T]} K_s |dX_s| = K_0 |X_0 - X_{0-}| + \int_{(0,T)} K_s |dX_s| + K_T |X_T - X_{T-}|.$$

Throughout we will assume that the forward price processes of the primary asset and the proxy are martingales. Thus, the returns have zero expectation and the liquidation is not affected by any directional views about the price processes (see also Remark 6.1.1). A model with a nonzero drift can result in profits from trading even if no initial position is to be closed; this makes it difficult to differentiate between optimal liquidation and optimal investment. The additional analysis of optimal investment is not the focus of this chapter (cf. the discussion about gas forward markets in Europe above).

The risk associated to a position strategy (X, Y) will be defined by

$$R(X,Y) = \int_0^T g(f(X_s, Y_s)) ds,$$

where f is a quadratic form and $g : [0,\infty) \to [0,\infty)$ is a continuous and nondecreasing function with $g(0) = 0$. More precisely, let

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

be a covariance matrix and

$$f(x,y) = \begin{pmatrix} x & y \end{pmatrix} \Sigma \begin{pmatrix} x \\ y \end{pmatrix} = \sigma_1^2 x^2 + 2\rho\sigma_1\sigma_2 xy + \sigma_2^2 y^2.$$

Throughout we assume that $\rho \geq 0$. One could assume the covariance matrix Σ to be time-dependent, but to simplify notation we refrain from doing so.

For $g(x) = \lambda\sqrt{x}$ with $\lambda \geq 0$ one can interpret $R(X,Y)$ as the time average of the value-at-risk associated to the position process (X,Y) along the trading period $[0,T]$. Indeed,

to be closed before delivery. It can, however, be easily extended to this case.

recall the *variance-covariance method* for estimating the value-at-risk of a portfolio (see e.g. Chapter 20 in [21] for an introduction). To this end suppose for simplicity that the forward price returns over a fixed holding period (e.g. 10 trading days) of the primary asset and the proxy asset are Gaussian, with covariance matrix Σ. The value-at-risk, at level $\alpha \in (0,1)$, of a portfolio with a constant position, during the holding period, of x units of the primary asset and y units of the proxy is given by

$$a\sqrt{\begin{pmatrix} x & y \end{pmatrix} \Sigma \begin{pmatrix} x \\ y \end{pmatrix}}, \tag{6.1}$$

where a is the α-quantile of the standard normal distribution. Many companies, as part of their risk management, estimate the risk of their portfolios with a formula like (6.1) (cf. the RiskMetric methodology [53]).

For $g(x) = x$, one can interpret $R(X,Y)$ as an approximation of the variance of the portfolio value in a market without market frictions. Indeed, assume for a moment that daily price returns are independent and have covariance Σ. For a deterministic strategy (X,Y) the integral $R(X,Y)$ is approximately equal to the variance of the portfolio value over the whole trading period $[0,T]$. In this context it was introduced in [4] for optimal liquidation of a single asset with quadratic price impact costs; within the same framework, the authors of [76] establish the equivalence of mean-variance minimization (for deterministic strategies) and utility maximization of investors with CARA preferences (for dynamic strategies). Subsequently similar quadratic additive risk functionals have been used in different contexts by [32] and [48].

To keep things general, we only assume that g is continuously differentiable on $(0,\infty)$. Moreover we assume that $x \mapsto g(f(x,y))$ is convex for all $y \geq 0$.

We suppose that the agent aims at minimizing the sum of the expected execution costs and the portfolio's risk. For $(X,Y) \in \mathcal{D}(x_0)$ we define the objective functional

$$\begin{aligned} J(X,Y) &= E\left[C(X,Y) + R(X,Y)\right] \\ &= E\left[\int_{[0,T]} K_s|dX_s| + \int_{[0,T]} L|dY_s| + \int_0^T g(f(X_s,Y_s))ds\right], \end{aligned} \tag{6.2}$$

where E is the expectation operator. Notice that the first two integrals in (6.2) include a possible jump at time 0.

The value function is defined by

$$v(x_0) = \inf_{(X,Y)\in\mathcal{D}(x_0)} J(X,Y). \tag{6.3}$$

As usual, we say that a position strategy is *optimal* if it attains the infimum in (6.3).

Optimal strategies in general are not absolutely continuous with respect to the Lebesgue measure. We are, therefore, dealing with a *singular stochastic control problem*.

For the analysis of optimal strategies it is very helpful to keep in mind the optimal cross hedge position in the case where it does *not* cost anything to cross hedge, i.e. if

$L = 0$: for a given primary position X, the hedge position that minimizes the portfolio's risk $R(X, Y)$ is given by

$$Y_t = -\rho \frac{\sigma_1}{\sigma_2} X_t.$$

We define $h = \rho\frac{\sigma_1}{\sigma_2}$ and remark that h is frequently referred to as the *minimum variance hedge ratio* (see e.g. [39], Chapter 3).

We close this section by observing that if forward prices are martingales, then a strategy minimizing execution costs also minimizes the agent's expected *overall costs* for closing the short position.

Remark 6.1.1. We assume that the price of the primary asset $(P_t)_{t\in[0,T]}$ is a continuous martingale. The agent's overall costs from following a position process X amount to $\int_{[0,T]} P_s dX_s + \int_{[0,T]} K_s|dX_s|$. Integrating by parts and using that $X_T = 0$ yields

$$\int_{[0,T]} P_s dX_s = -P_0 x_0 - \int_0^T X_{s-} dP_s.$$

Under suitable integrability assumptions on X the process $t \mapsto \int_0^t X_{s-} dP_s$ is a martingale starting in 0. This implies

$$E\left[\int_{[0,T]} P_s dX_s + \int_{[0,T]} K_s|dX_s|\right] = -P_0 x_0 + E\left[\int_{[0,T]} K_s|dX_s|\right],$$

and, hence, the expected overall costs are the difference of the expected execution costs and the initial book value $P_0 x_0$. Similar considerations hold true for the proxy position, which shows that (under suitable assumptions) minimizing the expected overall costs is equivalent to minimizing just the expected execution costs.

6.2. Optimal positions via optimal stopping

In this section we show that the problem of finding optimal position processes is equivalent to a family of stopping problems. To this end we first show that any optimal primary position path is nondecreasing, and any optimal proxy position is at first nondecreasing and then nonincreasing (Section 6.2.1). This allows us then to encode the optimal position strategy by stopping times (Sections 6.2.2 and 6.2.3). Finally, we present a method to successively determine the optimal position paths (Section 6.2.4).

6.2.1. Optimal position paths are (piecewise) monotone

We first show that the optimal position process X of the primary must be nondecreasing. On the other hand, the optimal position for the proxy is only nondecreasing until the "optimal" hedging position $-hX$ is reached; afterward the process is nonincreasing: at all times it is given by $-hX$. In particular, this confirms the intuition that the optimal hedge is *at most* the minimum variance hedge ratio.

We first require the following notation. For $z \in \mathbb{R}$ denote by $\mathcal{A}(z)$ the set of stochastic processes on $[0,T]$ that are adapted, càdlàg and nondecreasing and that satisfy $Z_t \geq z$ for all $t \in [0,T]$. We set $Z_{0-} = z$. If $z \leq 0$, then we denote by $\mathcal{A}_0(z)$ the subset of processes $Z \in \mathcal{A}(z)$ with $Z_T = 0$.

Proposition 6.2.1. *We have*

$$v(x_0) = \inf\{J(X, I \wedge -hX) | (X, I) \in \mathcal{A}_0(x_0) \times \mathcal{A}(0)\}.$$

Proof. Let (X,Y) be an arbitrary pair of position paths in $\mathcal{D}(x_0)$. Let $\widehat{X}$ be the smallest adapted, càdlàg and nondecreasing process dominating X (i.e. $\widehat{X}_t = \sup_{0\leq s\leq t} X_s$) and define

$$\tilde{Y}_t = Y_t \wedge -h\widehat{X}_t.$$

The execution costs entailed by $(\widehat{X}, \tilde{Y})$ are smaller than the costs of (X,Y). Moreover, we have that $f(\widehat{X}_t, \tilde{Y}_t) \leq f(X_t, Y_t)$, which implies that also the risk associated with $(\widehat{X}, \tilde{Y})$ is smaller than the risk associated with (X,Y). Therefore, $J(\widehat{X}, \tilde{Y}) \leq J(X,Y)$.

Next define a new cross hedge via $\widehat{Y}_t = \sup_{0\leq s\leq t} \tilde{Y}_s \wedge -h\widehat{X}_t$. Then the execution costs entailed by $(\widehat{X}, \widehat{Y})$ are smaller than or equal to the costs of $(\widehat{X}, \tilde{Y})$. Moreover, we have $f(\widehat{X}_t, \widehat{Y}_t) \leq f(\widehat{X}_t, \tilde{Y}_t)$. This shows that there exists a strategy of the form $Y_t = I_t \wedge -h\widehat{X}_t$ with $I \in \mathcal{A}(0)$ such that Y is at least as good as $\tilde{Y}$, i.e. $J(\widehat{X}, Y) \leq J(\widehat{X}, \tilde{Y})$. □

Instead of determining optimal position paths in both assets *simultaneously*, we first look at the problem of finding *reciprocal* optimal positions. This will give us some qualitative insights into the shape of position paths that will allow us to determine an optimal solution, at least in some cases.

We first explain how one can encode positions paths via family of stopping times. For any $Z \in \mathcal{A}(z)$ we define the associated family of stopping times

$$\tau(y) = \tau^Z(y) = \inf\{t \geq 0 | Z_t > y\}, \qquad \text{for all } y \in [z, \infty)$$

with the convention $\inf \emptyset = +\infty$. Observe that the mapping $[z,\infty) \ni y \mapsto \tau(y)$ is right-continuous and nondecreasing, and hence càdlàg. The process Z can be recovered from $(\tau(y))$, namely

$$Z_t = \inf\{y \geq z | \tau(y) > t\}. \tag{6.4}$$

Indeed, any family of stopping times $(\tau(y))_{y\geq z}$, such that $[z,\infty) \ni y \mapsto \tau(y)$ is right-continuous and nondecreasing, defines a process $Z \in \mathcal{A}(z)$ via (6.4). Hence there is a one-to-one correspondence between $\mathcal{A}(z)$ and $\mathcal{T}(z)$, the set of all such families of stopping times. We remark that if a process $Z \in \mathcal{A}(z)$ is bounded from above, say by c, then the process Z is encoded by the subfamily $(\tau(y))_{z\leq y\leq c}$.

We will frequently use the following change of variable formula (see also [71, Theorem 45, Chapter IV]).

Lemma 6.2.2. *For all measurable functions $f : [0,T] \to \mathbb{R}_+$ we have*

$$\int_{[0,T]} f(t)dZ_t = \int_z^{Z_T} f(\tau(y))dy.$$

Proof. We first show the result for indicator functions of the form $f = 1_{[0,t]}$, $t \in [0,T]$. Note that

$$\int_{[0,T]} 1_{[0,t]}(s)dZ_s = Z_t - z.$$

Since $\{Z_t > y\} \subset \{\tau(y) \le t\} \subset \{Z_t \ge y\}$, we have

$$\int_z^{Z_T} 1_{[0,t]}(\tau(y))dy \le \int_z^{Z_T} 1_{\{Z_t \ge y\}}dy = Z_t - z,$$

and

$$\int_z^{Z_T} 1_{[0,t]}(\tau(y))dy \ge \int_z^{Z_T} 1_{\{Z_t > y\}}dy = Z_t - z,$$

which proves that $\int_z^{Z_T} 1_{[0,t]}(\tau(y))dy = Z_t - z$.

The result follows now by a straightforward monotone class argument. □

6.2.2. Optimal primary position via optimal stopping

By Lemma 6.2.1 any optimal cross hedge is the minimum of a nondecreasing process I and the weighted primary position $-hX_t$. In this section we determine the optimal X for a *given* process I. Throughout this subsection we fix $I \in \mathcal{A}(0)$. For any $X \in \mathcal{A}_0(x_0)$ we define an associated cross hedge $Y(X)_t = I_t \wedge -hX_t$, $t \in [0,T]$. We will usually omit the dependence on X and simply write Y for $Y(X)$. Notice that $f(X_t, Y_t) = \tilde{f}(\omega, t, X_t)$, where

$$\tilde{f}(\omega,t,x) = \begin{cases} (1-\rho^2)\sigma_1^2 x^2 & \text{if } -hx \le I_t(\omega), \\ \sigma_1^2 x^2 - 2\rho\sigma_1\sigma_2 I_t(\omega)x + \sigma_2^2 I_t^2(\omega) & \text{else.} \end{cases}$$

We can formulate the problem of finding an optimal X for the given process I as follows:

$$\inf_{X\in\mathcal{A}_0(x_0)} E\left[\int_{[0,T]} K_s dX_s + L\int_{[0,T]} |dI_s \wedge -hX_s| + \int_0^T g(\tilde{f}(s,X_s))ds\right]. \tag{6.5}$$

The next proposition shows how to derive an optimal primary position path X from the solutions of a family of stopping problems.

Proposition 6.2.3. *Let $I \in \mathcal{A}(0)$ and fix the cross hedge $Y_t = I_t \wedge -hX_t$. For all $x \in [x_0, 0]$ let $\tau(x)$ be a solution of the stopping problem*

$$\inf_{\tau\in[0,T]} E\left[2Lh1_{\{\tau>\tau^I(-hx)\}} + K_\tau - \int_0^\tau g'(\tilde{f}(s,x))\tilde{f}_x(s,x)ds\right] \tag{6.6}$$

such that $(\tau(x)) \in \mathcal{T}(x_0)$*. Then the process* X *given by*

$$X_t = \inf\{x \in [x_0, 0] | \tau(x) > t\} \wedge 0$$

is optimal for (6.5)*.*

Remark 6.2.4. We give the following interpretation of the stopping problems (6.6). Instead of finding the optimal entire position path X in the first place, we may also answer the question of when to buy the infinitesimal unit dx located at $x \in [x_0, 0]$ and concatenate the position path afterward. Determining this optimal point in time $\tau(x)$ means to find an optimal tradeoff between three terms. First, the term K_τ represents the marginal costs for buying one unit. Second, the integral $-\int_0^\tau g'(\tilde{f}(s,x))\tilde{f}_x(s,x)ds$ accounts for risk savings: the sooner the unit located at x is bought, the smaller the marginal risk it contributes to the aggregate risk, since $g'(\tilde{f}(s,x))\tilde{f}_x(s,x) \geq 0$. Finally, the term $2Lh1_{\{\tau > \tau^I(-hx)\}}$ represents the costs incurred in the proxy position: If the unit at x is cross hedged ($\tau > \tau^I(-hx)$) we need to account for the costs of $2L$; else there are no cross hedging costs.

Proof. Let $(\tau^I(y))_{y \geq 0}$ be the family of stopping times associated to I and let $(\tau(x))_{x \geq x_0}$ be the family encoding a process $X \in \mathcal{A}_0(x_0)$. The change of variable formula of Lemma 6.2.2 implies that the costs in the primary asset satisfy

$$\int_{[0,T]} K_s dX_s = \int_{x_0}^0 K_{\tau(x)} dx. \tag{6.7}$$

Observe next that $\max_{s \in [0,T]}(I_s \wedge -hX_s) = \sup\{y \geq 0 | \tau(-\frac{y}{h}) > \tau^I(y)\}$. Hence the execution costs in the secondary asset are given by

$$L \int_{[0,T]} |dI_s \wedge -hX_s| = 2L \int_0^{-x_0/h} 1_{\{\tau(-\frac{y}{h}) > \tau^I(y)\}} dy = 2Lh \int_{x_0}^0 1_{\{\tau(x) > \tau^I(-hx)\}} dx.$$

The risk term satisfies

$$\begin{aligned} \int_0^T g(\tilde{f}(s, X_s))ds &= -\int_0^T \left(\int_{X_s}^0 g'(\tilde{f}(s,x))\tilde{f}_x(s,x)dx - g(\tilde{f}(s,0)) \right) ds \\ &= -\int_{x_0}^0 \int_0^{\tau(x)} g'(\tilde{f}(s,x))\tilde{f}_x(s,x)dsdx + \int_0^T g(\tilde{f}(s,0))ds. \end{aligned} \tag{6.8}$$

The previous calculations show that we can write the sum of time integrals in the expectation of (6.5) as an integral with respect to the position variable x. The functional in (6.5) is minimized if and only if the associated stopping times are optimal in (6.6). □

6.2.3. Optimal cross hedges via optimal stopping

The previous subsection describes how to derive optimal primary positions for a *given* proxy process. In this subsection we consider the opposite problem: for a given *primary* position process we characterize the optimal *cross hedge*.

Throughout let X be a fixed primary asset position process. We can formulate the problem of finding an optimal Y as follows:

$$\inf_{I\in\mathcal{A}(0)} E\left[\int_{[0,T]} K_s dX_s + L\int_{[0,T]} |dI_s \wedge -hX_s| + \int_0^T g(\tilde{f}(s,X_s))ds\right]. \tag{6.9}$$

Problem (6.9) can again be reduced to a family of stopping problems.

Proposition 6.2.5. *Let $\bar{y} = -hx_0$. For all $y \in [0,\bar{y}]$ let $\tau(y)$ be the solution of the stopping problem*

$$\inf_{\tau\in[0,T]} E\left[\int_0^\tau g'(f(X_s,y))[f_y(X_s,y)]^- ds - 2L1_{\{\tau=T\}}\right] \tag{6.10}$$

such that $(\tau(y)) \in \mathcal{T}(0)$. Then a cross hedging strategy Y for which the infimum in (6.9) is attained is given by

$$Y_t = I_t \wedge -hX_t,$$

where the process I is the right continuous inverse of $\tau(y)$, i.e. $I_t = \inf\{y \in [0,\bar{y})|\tau(y) > t\}$.

Proof. Let $(\tau(y))_{y\geq 0} = (\tau^I(y))_{y\geq 0} \in \mathcal{T}(0)$ be the family of stopping times which are optimal in (6.10). By $(\tau^X(x))_{x\geq x_0}$ we denote the family of stopping times encoding X. Then $\tau(y)$ is also optimal in the problem

$$\inf_{\tau\in[0,T]} E\left[1_{\{\tau<\tau^X(-\frac{y}{h})\}}\left(2L + \int_\tau^{\tau^X(-\frac{y}{h})} g'(f(X_s,y))f_y(X_s,y)ds\right)\right] \tag{6.11}$$

such that $\tau \in \mathcal{T}(0)$. Indeed, (6.11) can be rearranged as

$$\begin{aligned}
&E\left[1_{\{\tau<\tau^X(-\frac{y}{h})\}}\left(2L + \int_\tau^{\tau^X(-\frac{y}{h})} g'(f(X_s,y))f_y(X_s,y)ds\right)\right] \\
= \;& E\left[1_{\{\tau<\tau^X(-\frac{y}{h})\}}\left(2L - \int_\tau^T g'(f(X_s,y))[f_y(X_s,y)]^- ds\right)\right] \\
= \;& 2L - E\left[\int_0^T g'(f(X_s,y))[f_y(X_s,y)]^- ds\right] \\
&+ E\left[\int_0^\tau g'(f(X_s,y))[f_y(X_s,y)]^- ds - 2L1_{\{\tau\geq\tau^X(-\frac{y}{h})\}}\right],
\end{aligned}$$

where we used the fact that $t \leq \tau^X(-\frac{y}{h})$ is equivalent to $f_y(X_t,y) \leq 0$. Hence, (6.11) is equivalent to

$$\inf_{\tau\in[0,T]} E\left[\int_0^\tau g'(f(X_s,y))[f_y(X_s,y)]^- ds - 2L1_{\{\tau\geq\tau^X(-\frac{y}{h})\}}\right]. \tag{6.12}$$

Since $f_y(X_t, y) \geq 0$ for $t \geq \tau^X(-\frac{y}{h})$, we can restrict ourselves to stopping times taking values in $[0, \tau^X(-\frac{y}{h})) \cup \{T\}$, which implies that $\tau(y)$ is optimal in (6.12).

Next, define $I_t = \inf\{y \in [0, \bar{y}) | \tau(y) > t\}$ and $Y_t = I_t \wedge -hX_t$. Then we have

$$L \int_{[0,T]} |dY_s| = 2L \int_0^{\bar{y}} 1_{\{\tau(y) < \tau^X(-\frac{y}{h})\}} dy.$$

The risk term of the objective functional can be represented as follows:

$$\begin{aligned}\int_0^T g(f(X_s, Y_s))ds &= \int_0^T \left(\int_0^{Y_s} g'(f(X_s, y)) f_y(X_s, y) dy + g(f(X_s, 0)) \right) ds \\ &= \int_0^{\bar{y}} \int_{\tau(y)}^{\tau^X(-\frac{y}{h})} g'(f(X_s, y)) f_y(X_s, y) ds 1_{\{\tau(y) < \tau^X(-\frac{y}{h})\}} dy \\ &\quad + \int_0^T g(f(X_s, 0)) ds.\end{aligned}$$

Hence, the objective functional is given by

$$\begin{aligned}J(0) = E \Bigg[\int_0^{\bar{y}} \Bigg(2L 1_{\{\tau(y) < \tau^X(-\frac{y}{h})\}} + \int_{\tau(y)}^{\tau^X(-\frac{y}{h})} g'(f(X_s, y)) f_y(X_s, y) ds 1_{\{\tau(y) < \tau^X(-\frac{y}{h})\}} \Bigg) dy \\ + \int_0^T g(f(X_s, 0)) ds + \int_{[0,T]} K_s |dX_s| \Bigg].\end{aligned}$$

Since $\tau(y)$ is optimal in (6.11), we obtain that I is optimal as well. □

If the optimal cross hedging strategy Y is nonincreasing after a possible jump at time 0, then I_t is constant and the problem of finding the optimal cross hedge reduces to finding the optimal initial cross hedge level, a considerably simpler problem. The following example shows, however, that this simplification is not always possible.

Example 6.2.6. *Suppose you have a short position of $x_0 < 0$ in the primary asset and you ask a counterparty to make a sell offer at a price that, in your view, includes no liquidity premium. The counterparty is indecisive about whether to make the offer and asks for some time for consideration. Do you cross hedge until the decision? The following example shows that if you think that the counterparty will accept with a high probability, then you do not cross hedge. It also shows that a cross hedging process Y is not necessarily nonincreasing after time 0.*

Consider a nonincreasing cost process K that takes only two values K_+ and K_-, where $K_+ > K_-$. For simplicity we assume $K_- = 0$. Suppose that at a deterministic time $\delta \in (0, T)$ the process jumps from K_+ to the lower level K_- with probability $p \in (0, 1)$. With probability $1 - p$ the process stays constant equal to K_+ and jumps to the lower level only at T. There are only two scenarios for the cost process, and the scenario the process takes is revealed to the agent at δ.

We suppose that $\sigma_1 = \sigma_2 = \sigma > 0$, that $g(x) = \sqrt{x}$ and that

$$\sigma^2(1 - \rho^2)T + 2L\rho < K_+. \tag{6.13}$$

The latter condition implies that it is not optimal to buy a unit of the primary asset before the cost process jumps. The optimal primary asset position is given by

$$X_t = \begin{cases} x_0 & \text{if } K_t = K_+, \\ 0 & \text{if } K_t = K_-, \end{cases}$$

for $t \in [0, T]$. *We suppose that the cross hedging costs* L *are low in comparison to the risk entailed by keeping the position open over the whole trading period. More precisely, assume that*

$$2L < \sigma(T-\delta) \quad \text{and} \quad 2L\frac{\sqrt{1-\rho^2}}{\sqrt{\sigma^2(T-\delta)^2 - 4L^2}} < \rho. \tag{6.14}$$

Condition (6.14) *guarantees that*

$$y_2 = \left(\rho - 2L\frac{\sqrt{1-\rho^2}}{\sqrt{\sigma^2(T-\delta)^2-4L^2}}\right)(-x_0)$$

is positive. By Proposition 6.4.2 it is optimal to cross hedge with a position of y_2 *between* δ *and* T *if there has been no jump at* δ. *Let* Y *be the strategy that is constant equal to* $y \in [y_2, 0]$ *on* $[0, \delta)$, *and constant equal to* y_2 *on* $[\delta, T)$ *if there has been no jump. Define*

$$A(y) = E\left[\int_{[0,T]} K_s|dX_s| + \int_{[0,T]} L|dY_s| + \int_0^T \sqrt{f(X_s, Y_s)}ds\right]$$

and observe that

$$A(y) = 2Lpy + 2L(1-p)y_2 + \delta\sqrt{f(x_0, y)} + (T-\delta)(1-p)\sqrt{f(x_0, y_2)}.$$

Now suppose that

$$p > \frac{\delta\sigma}{2L} \in (0,1). \tag{6.15}$$

Then the derivative $\frac{\partial A}{\partial y}$ *is nonpositive on* $[0, -\rho x_0]$, *which implies that the minimum of* $A(y)$ *on* $[0, -\rho x_0]$ *is attained at* $y = 0$. *This shows that it is optimal not to build up any cross hedge before* δ.

For the parameters $K^+ = 1$, $L = 0.1$, $T = 1$, $\delta = 0.1$, $\sigma = 1$, $\rho = 0.9$ *and* $p = 0.9$ *the conditions* (6.13), (6.14) *and* (6.15) *are satisfied.*

6.2.4. Successive determination of optimal position paths

If the optimal cross hedging process is nonincreasing after time 0, then one can use an iterative procedure for determining optimal positions. This will be the case in the case studies presented in Sections 6.3 and 6.4. In this subsection we describe this iterative procedure.

Assume that the optimal cross hedge $Y(X)$ associated to any $X \in \mathcal{A}_0(x_0)$ is nonincreasing after 0. Then the value function (6.3) satisfies

$$\begin{aligned} v(x_0) &= \inf_{X\in\mathcal{A}_0(x_0)} \inf_{y\geq 0} E\left[\int_{[0,T]} K_s dX_s + 2Ly + \int_0^T g(f(X_s, y\wedge -hX_s))ds\right] \\ &= \inf_{y\geq 0}\left(2Ly + \inf_{X\in\mathcal{A}_0(x_0)} E\left[\int_{[0,T]} K_s dX_s + \int_0^T g(\bar{f}(X_s,y))ds\right]\right), \end{aligned} \tag{6.16}$$

with

$$\bar{f}(x,y) = \begin{cases} (1-\rho^2)\sigma_1^2 x^2 & \text{if } -hx \leq y, \\ \sigma_1^2 x^2 - 2\rho\sigma_1\sigma_2 xy + \sigma_2^2 y^2 & \text{else.} \end{cases}$$

Suppose that we can solve, for any $y \geq 0$, the problem

$$w(x_0,y) := \inf_{X\in\mathcal{A}_0(x_0)} E\left[\int_{[0,T]} K_s dX_s + \int_0^T g(\bar{f}(X_s,y))ds\right]. \tag{6.17}$$

Moreover assume that there exists a $y^* \geq 0$ for which the infimum in

$$v(x_0) = \inf_{y\geq 0}\{2Ly + w(x_0,y)\} \tag{6.18}$$

is attained. Then the optimal primary asset position is given by the process X^* that solves (6.17) for $y = y^*$; the optimal cross hedge position is given by $Y_t^* = y^* \wedge -hX_t^*$ for all $t \in [0,T]$.

The next proposition shows that the optimal solution of the auxiliary problem (6.17) can again be characterized by a family of stopping times.

Proposition 6.2.7. *For all $x \in [x_0, 0]$ let $\tau(x)$ be the solution of the stopping problem*

$$\inf_{\tau\in[0,T]} E\left[K_\tau - \tau g'(\bar{f}(x,y_0))\bar{f}_x(x,y_0)\right].$$

Then an optimal primary position X for (6.17) is given by

$$X_t = \inf\{x \in [x_0,0] | \tau(x) > t\} \wedge 0.$$

Proof. A change of variables as performed in Equations (6.7) and (6.8) implies the result. □

6.2.5. The case of costless cross hedging

In the case without cross hedging costs the problem of finding optimal position paths can be considerably simplified. If $L = 0$, it is optimal to perform a minimum variance hedge in the proxy position $Y = -hX$. The two-dimensional singular control problem (6.3) thus reduces to a one dimensional problem, which again can be formulated as a collection of stopping problems. The results are summarized in the following corollary. We set $G(x) = g((1-\rho^2)\sigma_1^2 x^2)$.

Corollary 6.2.8. *Assume that $L = 0$. For all $x \in [x_0, 0]$ let $\tau(x)$ be the solution of the stopping problem*

$$\inf_{\tau \in [0,T]} E\left[K_\tau - \tau G'(x)\right]$$

such that $(\tau(x)) \in \mathcal{T}(x_0)$. Let X denote its right-continuous inverse

$$X_t = \inf\{x \in [x_0, 0] | \tau(x) > t\} \wedge 0.$$

Then $(X, -hX) \in \mathcal{D}(x_0)$ is optimal in (6.3).

Proof. It follows from Proposition 6.2.5 that for any primary position path X the optimal proxy position path is $-hX$. Then Proposition 6.2.3 implies the result. □

6.3. Case study: Active trading kicks in at a random time

Trading of forwards usually becomes *active* as soon as the *time to the delivery date* falls below a certain time threshold. For example, a month forward may be actively traded during the three months before delivery; but not if the delivery date lies more than three months ahead. The trading community usually latently agrees upon a time at which they start trading a forward. Acquiring a forward before the active trading period calls for an additional liquidity premium. Once the trading has become active, the additional premium is no longer asked for. Liquidity in this case does not increase uniformly, but comes suddenly.

The *precise* time when trading of a particular forward becomes active, however, is often not predictable. Traders can have expectations about when active trading starts, but the precise starting date can be random. In this section we assume that liquidity increases at a random time before maturity, at which active trading kicks in and hence turns the forward market liquid. We have an illiquid trading period before the kick-in date and a liquid one afterwards. For simplicity we assume that the liquidity costs K are constant before respectively after the kick-in date: K jumps at a random time $\tilde{\tau} \in [0, T]$ from a higher level $K_+ > 0$ to a lower level $K_- \in [0, K_+)$. We model $\tilde{\tau}$ as the first jump time of an inhomogeneous Poisson process with nondecreasing jump intensity. More precisely, let ξ be a random variable with standard exponential distribution and $\gamma : \mathbb{R}_+ \to [0, \infty]$ a nondecreasing function with $\gamma \neq 0$. Define $\Gamma(t) = \int_0^t \gamma(s)ds$ and $\tilde{\tau} = \Gamma^{-1}(\xi)$. Notice that $\tilde{\tau}$ has the same distribution as the first jump time of a Poisson process with jump intensity $\gamma(t)$ at time t. We assume that $\Gamma(T) = \infty$; hence, $\tilde{\tau} \leq T$ almost surely (in other words, there is a period with active trading). Moreover, we suppose that the filtration $\mathbb{F}$ is generated by K.

We start by making two observations:

(i) Every optimal liquidation strategy (X, Y) satisfies $X_t = Y_t = 0$ for all $t \geq \tilde{\tau}$.

(ii) Since X is adapted to the filtration generated by K, we have

$$X_t = x(t)1_{\{t<\tilde{\tau}\}} \tag{6.19}$$

for some deterministic, nondecreasing function $x : [0,T] \to \mathbb{R}_-$.

We will first show that the optimal cross hedge is monotone, i.e. it is optimal to build up an initial cross hedge position and then to reduce it simultaneously with the primary asset (Section 6.3.1). This result allows us to use the iterative procedure from Section 6.2.4 for calculating the optimal positions. Proposition 6.2.7 implies that the optimal primary position is of the form $X_t = x1_{\{t<\tilde{\tau}\}}$, with $x \leq 0$ depending on the initial cross hedge position (Section 6.3.2). We then obtain the optimal initial cross hedge as the position y^* for which the infimum in (6.18) is obtained. Since the primary is constant up to the jump time $\tilde{\tau}$, this implies that the optimal cross hedge is given by $Y_t = y^* 1_{\{t<\tilde{\tau}\}}$. The optimal positions, depending on the expected jump time, will be explicitly given in the case where the risk is measured with $g(x) = \lambda\sqrt{x}$; in this case, $\lambda \geq 0$ can be interpreted as a risk aversion parameter (Section 6.3.3). The proofs of the results of this section are presented in Appendix 6.6.1.

6.3.1. Optimal cross hedges are static

In this subsection we show that the stopping times $\tau(y)$ solving the stopping problem (6.10) are either constant equal to 0 or equal to T. In view of Proposition 6.2.5 this means that the optimal cross hedge for any *given* primary position process $X \in \mathcal{A}_0(x_0)$ is nonincreasing after time 0. We can thus use the iterative procedure described in Subsection 6.2.4 for determining optimal strategies.

Proposition 6.3.1. *Let $X \in \mathcal{A}_0(x_0)$ be a primary position path. Then for every $y \in [0, -hx_0]$ there exists an optimal stopping time $\tau(y)$ for* (6.10) *that takes values only in $\{0, T\}$. In particular, there exists an optimal cross hedging strategy Y of the form $Y_t = y^* \wedge -hX_t$ for some $y^* \in [0, -hx_0]$ attaining the infimum in* (6.9).

6.3.2. Optimal primary position paths are static

Proposition 6.3.1 implies that optimal cross hedging strategies are of the form $Y_t = y \wedge -hX_t$ for some $y \in [0, -hx_0]$. We define $Y_{0-} = y$. We can use Proposition 6.2.7 to obtain optimal primary position paths from solving the following family of stopping problems

$$\inf_{\tau\in[0,T]} E\left[K_\tau - \tau\alpha(x)\right], \tag{6.20}$$

where $\alpha(x) = g'(\bar{f}(x,y))\bar{f}_x(x,y)$. We make the following two observations:

(i) Every optimal stopping time fulfills $\tau \leq \tilde{\tau}$ almost surely (else the stopping time $\tau' := \tau \wedge \tilde{\tau}$ performs strictly better since $\alpha(x) \leq 0$).

(ii) Let τ be a stopping time such that $\tau \leq \tilde{\tau}$. For measurability reasons there exists a time $t \in [0, T]$ such that $\tau = \tau_t := t \wedge \tilde{\tau}$.

The next proposition describes the stopping time solving (6.20). To this end define

$$\bar{x} = \max \left\{ x \leq 0 | \alpha(x) \leq -\frac{E[K_+ - K_{\tilde{\tau}}]}{E[\tilde{\tau}]} \right\}.$$

Proposition 6.3.2. *Let $\tau(x) = 0$ for $x \in (-\infty, \bar{x})$ and $\tau(x) = \tilde{\tau}$ for $x \in [\bar{x}, 0]$. Then $\tau(x)$ is an optimal stopping time for* (6.20).

6.3.3. Explicit optimal strategies

In this subsection we derive explicit position paths for a specific choice of the risk function. More precisely, we choose $g(f(x,y)) = \lambda\sqrt{f(x,y)}$ with $\lambda \geq 0$. This corresponds to a first order approximation of the position's value-at-risk. From the preceding sections we know that optimal strategies are static, i.e. $X_t^* = x^* 1_{\{t<\tilde{\tau}\}}$ and $Y_t^* = y^* 1_{\{t<\tilde{\tau}\}}$. We will derive explicit formulas for x^* and y^* in terms of the model parameters. To this end we distinguish several cases.

Proposition 6.3.3. *Let $\Delta K = K_0 - E[K_{\tilde{\tau}}]$ and*

$$M = \frac{\sigma_2}{\sigma_1}\left(\Delta K \rho - \sqrt{(1-\rho^2)(\lambda^2\sigma_1^2 E[\tilde{\tau}]^2 - \Delta K^2)^+}\right).$$

1. *If $\Delta K \leq \lambda\sigma_1\sqrt{1-\rho^2}E[\tilde{\tau}]$, then it is optimal to close the primary position immediately and not to cross hedge, i.e. $x^* = y^* = 0$.*

2. *If $\Delta K \geq \lambda\sigma_1 E[\tilde{\tau}]$, then it is optimal to keep the primary position open and to hedge with $y^* = -\frac{\sigma_1}{\sigma_2}\max\left(0, \rho - 2L\frac{\sqrt{1-\rho^2}}{\sqrt{(\lambda^2\sigma_2^2E[\tilde{\tau}]^2 - 4L^2)^+}}\right)x_0$ units of the proxy.*

3. *Suppose that $\lambda\sigma_1\sqrt{1-\rho^2}E[\tilde{\tau}] \leq \Delta K < \lambda\sigma_1 E[\tilde{\tau}]$. If $M \leq 2L$, then it is optimal to close the primary position immediately and not to cross hedge.*
   If $M \geq 2L$, then it is optimal to keep the primary position open and to hedge with $y^* = -\frac{\sigma_1}{\sigma_2}\left(\rho - 2L\frac{\sqrt{1-\rho^2}}{\sqrt{(\lambda^2\sigma_2^2E[\tilde{\tau}]^2 - 4L^2)^+}}\right)x_0$ units of the proxy.

6.4. Case study: Deterministic liquidity costs

In this section we turn to deterministic continuous liquidity processes K and derive explicit optimal strategies. The optimal position paths are essentially determined by the time decay of liquidity costs.

If the position of the primary and the proxy is constant equal to (x, y) on $[t, T)$, then the associated risk decays linearly in t at a rate of $g(f(x,y))$. The liquidity costs implied by buying one unit of the primary asset decrease at rate $\dot{K}_t$. Suppose now that the initial liquidity costs are high compared to the risk such that it is not optimal to buy a unit of the primary asset at $t = 0$. If $\dot{K}_t$ is nonincreasing, then the costs do not decrease faster than linearly. Hence costs remain high relative to the risk and it is optimal not to buy before T. In Section 6.4.1 we confirm this intuition by showing that it is optimal to close the whole position either immediately or at T if $\dot{K}_t$ is nonincreasing (i.e. if K is concave). In Section 6.4.2 we treat the case where K is convex. The proofs of the results of this section are presented in Appendix 6.6.2.

6.4.1. Concave decay of liquidity costs

Assume that K is concave and deterministic.

Since no randomness is involved in the model set-up, we can restrict ourselves to deterministic execution strategies. First, we consider the problem of finding optimal cross hedging strategies for a fixed primary position path X. Note that for every $y \leq 0$ the mapping

$$t \mapsto \int_0^t g'(f(X_s, y))[f_y(X_s, y)]^- ds - 2L1_{\{t=T\}}$$

is nondecreasing on $[0, T)$ with a possible downward jump at time T. Hence, it attains its minimum at 0 or T. Proposition 6.2.5 implies that optimal cross hedges are of the form $Y_t = y \wedge -hX_t$ for some $y \geq 0$. Next, we use Proposition 6.2.7 and consider the stopping problem

$$\inf_{t \in [0,T]} \left(K_t - tg'(\bar{f}(x, y))\bar{f}_x(x, y)\right).$$

Concavity of K implies that 0 or T are optimal. Hence, the static path $X_t = 1_{\{t<T\}}x(y)$ with

$$x(y) = \max\{x \leq 0 | K_0 \leq K_T - g'(\bar{f}(x, y))\bar{f}_x(x, y)T\}$$

is optimal.

For the specific choice $g(x) = \lambda\sqrt{x}$ we can perform calculations similar to those in Section 6.3.3. The next proposition provides explicit optimal strategies. To simplify notation we define $\Delta K = K_0 - K_T$ and the nonnegative number

$$A = -\frac{\sigma_1}{\sigma_2} \max\left(0, \rho - 2L\frac{\sqrt{1-\rho^2}}{\sqrt{(\lambda^2\sigma_2^2T^2 - 4L^2)^+}}\right) x_0. \tag{6.21}$$

Proposition 6.4.1 (Concave case). *Suppose that $g(x) = \lambda\sqrt{x}$ ($\lambda \geq 0$) and that K is decreasing and concave on $[0, T]$. Then there exists an optimal strategy that is static, i.e. of the form*

$$X_t^* = x^* 1_{[0,T)}(t), \quad Y_t^* = y^* 1_{[0,T)}(t), \tag{6.22}$$

with $x^* \leq 0$ and $y^* \geq 0$. The optimal positions are as follows:

(C1) *If* $\Delta K \leq \lambda\sigma_1\sqrt{1-\rho^2}T$, *then it is optimal to close the primary position immediately and not to cross hedge, i.e.* $x^* = y^* = 0$.

(C2) *If* $\Delta K \geq \lambda\sigma_1 T$, *then it is optimal to keep the primary position open and to hedge with* $y^* = A$ *proxy contracts.*

(C3) *If* $\lambda\sigma_1\sqrt{1-\rho^2}T \leq \Delta K \leq \lambda\sigma_1 T$ *and* $\frac{\sigma_2}{\sigma_1}\left(\Delta K\rho - \sqrt{(1-\rho^2)(\lambda^2\sigma_1^2T^2 - \Delta K^2)}\right) \leq 2L$, *then it is optimal to close the primary position immediately and not to cross hedge.*

(C4) *If* $\lambda\sigma_1\sqrt{1-\rho^2}T \leq \Delta K \leq \lambda\sigma_1 T$ *and* $\frac{\sigma_2}{\sigma_1}\left(\Delta K\rho - \sqrt{(1-\rho^2)(\lambda^2\sigma_1^2T^2 - \Delta K^2)}\right) \geq 2L$, *then it is optimal to keep the primary position open and to hedge with* $y^* = A$ *proxy contracts.*

In the following, we give a brief economic interpretation. In case (C1) the additional liquidity costs from an early closure are smaller than the risk entailed by keeping the primary position open until T. Since the speed of the cost decay is nondecreasing, it cannot be optimal to close the position at an intermediate point between 0 and T. It is, therefore, optimal to close the primary position immediately. Consequently, there is no need for a cross hedge.

In case (C2) the additional liquidity costs exceed the risk, even if no hedge is performed. Again it is not optimal to close the position at an intermediate point between 0 and T; hence it is optimal to keep the primary position open until T. The optimal cross hedge position is static, too and can be derived by a straightforward calculation. Notice that if the costs L for trading the proxy are high, then no cross hedge position is taken.

If the initial liquidity costs lie between $\rho\sigma^2 T$ and $\sigma^2 T$, then the costs for a cross hedge determine whether or not it is optimal to close the primary position immediately. In case (C3) the costs are too high; hence the position is closed at $t = 0$, and no cross hedge is performed. In case (C4) the liquidity costs are low; it is optimal to keep the primary position open and to cross hedge.

From Proposition 6.4.1 we derive a simple decision rule for whether to cross hedge or not. It is described in Corollary 6.4.2 and illustrated in the decision tree in Figure 6.1.

Corollary 6.4.2. *Let the assumptions of Proposition 6.4.1 hold true and*

$$\bar{L} = \frac{\sigma_2}{2\sigma_1}\left(\Delta K\rho - \sqrt{(1-\rho^2)(\lambda^2\sigma_1^2T^2 - \Delta K^2)^+}\right). \tag{6.23}$$

The optimal positions in (6.22) *are as follows:*

1. *If* $L < \bar{L}$, *then it is optimal to keep the primary position open and to cross hedge with* A *units of the proxy; i.e.* $x^* = x_0$ *and* $y^* = A$.

2. *If* $L \geq \bar{L}$, *then it is optimal* not *to cross hedge (i.e.* $y^* = 0$*). Whether it is optimal to immediately close the primary position depends on the size of the cost increment. If* $\Delta K \geq \lambda\sigma_1 T$, *then* $x^* = x_0$. *If* $\Delta K < \lambda\sigma_1 T$, *then* $x^* = 0$.

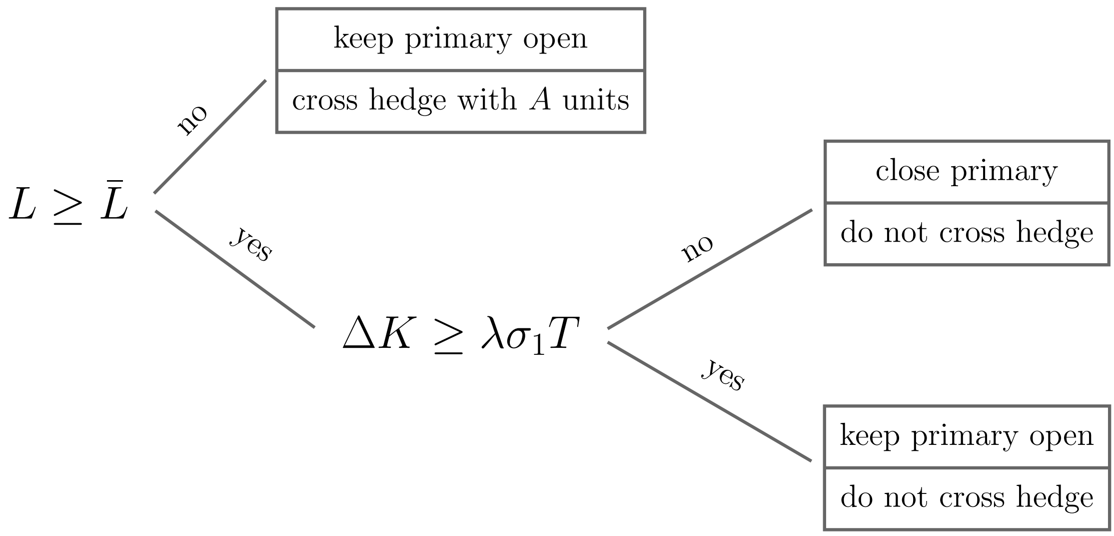

Figure 6.1.: Decision tree when K is concave. The constants A and $\bar{L}$ are defined in (6.21) and (6.23). , respectively.

6.4.2. Convex decay of liquidity costs

If the speed of the cost decay decreases, then it can be reasonable for the agent to postpone the liquidation of the primary position, even if immediate liquidity costs are smaller than the additional risk that this unit entails. Essentially it is optimal to buy as soon as the cost saving rate does not exceed the rate of the additional risk.

In the following we suppose that the liquidity cost process K is strictly convex and $K \in \mathcal{C}^1$, which means that the cost saving rate is decreasing in time. Note that this implies that $\dot{K}$ is strictly increasing and hence invertible. Moreover we assume that there are no cross hedging costs: $L = 0$.[3] We are thus in the framework of Subsection 6.2.5. The optimal stopping problem of Corollary 6.2.8 reads

$$\inf_{\tau \in [0,T]} E[K_\tau - \tau G'(x)], \tag{6.24}$$

where $G(x) = g((1-\rho^2)\sigma_1^2 x^2)$ for $x \in [x_0, 0]$. Note that G is convex and continuously differentiable, and $G'(x) \le 0$ for $x \le 0$.

As no randomness is involved the optimal time solving (6.24) is deterministic and is given by

$$t(x) = \begin{cases} T & \text{if } G'(x) \ge \dot{K}_T, \\ (\dot{K})^{-1}(G'(x)) & \text{if } \dot{K}_0 \le G'(x) < \dot{K}_T, \\ 0 & \text{if } G'(x) < \dot{K}_0. \end{cases}$$

It is straightforward to show that $x \mapsto t(x)$ is continuous and nondecreasing on $(-\infty, 0]$.

[3]This assumption allows us to derive explicit results while it does not contradict the main economic idea of this section.

We set

$$\begin{aligned} b &= \max\{x \le 0 | G'(x) \le \dot{K}_0\}, \\ a &= \min\{x \le 0 | G'(x) \ge \dot{K}_T\}, \end{aligned}$$

where we use the convention $\min \emptyset = 0$ and $\max \emptyset = -\infty$. Note that $0 \ge a \ge b \ge -\infty$. Proposition 6.2.3 implies that the optimal primary asset position trajectory can be recovered as the inverse of the mapping $t(x)$. The next proposition describes the optimal strategy precisely.

Proposition 6.4.3. *Suppose that $L = 0$ and that K is decreasing, continuously differentiable and strictly convex on $[0, T]$. The optimal primary position strategy $(X_t)_{t\in[0,T]}$ of closing $x_0 < 0$ is given as follows:*

1. *At time $t = 0$ it is optimal to buy the amount of $(b - x_0)_+$, i.e. $X_0 = \max\{x_0, b\}$.*
2. *The position is continuously increased between $t(X_0)$ and $t(a)$. More precisely, the optimal position at $t \in [t(X_0), t(a))$ is given by $X_t = (G')^{-1}(\dot{K}_t)$.*
3. *The remaining open position that has to be closed at time T is given by $a \vee x_0$ (note that a may be zero).*

The optimal cross hedge position is $Y_t = -hX_t$.

If the risk is measured with $g(x) = \lambda\sqrt{x}$, then it is optimal to close the primary position in one go. Notice that in this case the risk is linear in the position size. Also the liquidity costs are proportional to the position size. Therefore, as soon as it is optimal to buy *one* unit of the primary asset, it is optimal to close the *whole* primary position immediately.

Corollary 6.4.4 (Strict convex case)**.** *Suppose that the assumptions of Proposition 6.4.3 hold true and that $g(x) = \lambda\sqrt{x}$.*

1. *If $\dot{K}(T) < -\lambda\sigma_1\sqrt{1-\rho^2}$, then the optimal position processes are given by $X^* = x_0 1_{[0,T)}$ and $Y^* = -hx_0 1_{[0,T)}$.*
2. *If $\dot{K}(0) > -\lambda\sigma_1\sqrt{1-\rho^2}$, then the optimal position processes are given by $X^* = Y^* = 0$.*
3. *If $\lambda\sigma_1\sqrt{1-\rho^2} \in [-\dot{K}(T), -\dot{K}(0)]$, then the optimal buying time is given by*

$$t^* = (\dot{K})^{-1}(-\lambda\sigma_1\sqrt{1-\rho^2}),$$

   and $X^* = x_0 1_{[0,t^*)}$ and $Y^* = -hx_0 1_{[0,t^*)}$ are the optimal position processes.

Proof. Notice that in this case the derivative G' is constant equal to $\lambda\sigma_1\sqrt{1-\rho^2}$. In particular, the points a and b are equal either to 0 or $-\infty$. The result now follows directly from Proposition 6.4.3. □

6.5. Case study: Brownian bridge liquidity costs

In this section we suppose that the liquidity cost process K evolves according to a Brownian bridge. We fix an initial value $K_+ > 0$ and a terminal value $0 < K_- \leq K_+$. The volatility is denoted by $\sigma > 0$. Then the dynamics of K are given by

$$dK_t = -\frac{K_t - K_-}{T-t}dt + \sigma dW_t,$$

where W is a Brownian motion. The filtration $\mathbb{F}$ is generated by W. Moreover we assume that there are no cross hedging costs, i.e. $L = 0$.

Notice that the process K can become negative. In order to have a recourse to the method developed in Subsection 6.2.5, we allow only for nondecreasing position processes X, i.e. we restrict the set of primary position processes to $\mathcal{A}_0(x_0)$.

Again we can determine optimal position paths explicitly. To this end we introduce the following notations. We refer to Φ as the cumulative distribution function of the standard normal distribution. We denote by B the solution of the equation

$$\sqrt{2\pi}(1-B^2)e^{\frac{B^2}{2}}\Phi(B) = B$$

($B \approx 0.8399$). Moreover we introduce the process

$$Z_t = \frac{K_- - K_t - \sigma B\sqrt{T-t}}{T-t}.$$

Recall from Subsection 6.2.5 that $G(x) = g((1-\rho^2)\sigma_1^2 x^2)$. We obtain the following result; the proof is presented in Appendix 6.6.3.

Proposition 6.5.1. *If G is strictly convex, then the optimal primary position process X in $\mathcal{A}_0(x_0)$ is given by the running maximum of the process $(G')^{-1}(Z)$ cut off at 0:*

$$X_t = \sup_{s\leq t}\big((G')^{-1}(Z_s)\big) \wedge 0. \tag{6.25}$$

If G is linear (i.e. $g(x) = \sqrt{x}$), then it is optimal to close the whole primary position in one go at the first time when Z falls below the level $-\lambda\sqrt{1-\rho^2}\sigma_1$:

$$X_t = x_0 1_{\{t<\tau\}}$$

with

$$\tau = \inf\{s \geq 0 | Z_s \geq -\lambda\sqrt{1-\rho^2}\sigma_1\}.$$

In both cases the optimal proxy position process is given by $Y_t = -hX_t$.

6.6. Appendix

6.6.1. Proofs of the results of Section 6.3

For the proof of Proposition 6.3.1, we first require some preliminary considerations. We fix a primary position process $X_t = x(t)1_{\{t<\tilde{\tau}\}}$ as in (6.19). For ease of notation we introduce the nonnegative, nonincreasing function

$$\phi(t) = [f_y(x(t), y)]^- g'(f(x(t), y))$$

and the process

$$A_t = \int_0^t \phi(s)1_{\{s<\tilde{\tau}\}} ds - 2L1_{\{t=T\}},$$

where we suppress the dependence on X and y. Then the stopping problem (6.10) can be rewritten as

$$\inf_{\tau\in[0,T]} E[A_\tau]. \tag{6.26}$$

Furthermore we introduce the mapping

$$\beta(t) = \int_t^T \phi(s)P(s < \tilde{\tau}|t < \tilde{\tau})ds.$$

Notice that β is nonincreasing. Indeed, we have $\beta(t) = e^{\Gamma(t)} \int_t^T \phi(s)e^{-\Gamma(s)}ds$, and hence the monotonicity of γ and ϕ imply

$$\begin{aligned}
\beta'(t) &= \gamma(t)e^{\Gamma(t)} \int_t^T \phi(s)e^{-\Gamma(s)}ds - \phi(t) \\
&\leq \phi(t) \left(e^{\Gamma(t)} \int_t^T \gamma(s)e^{-\Gamma(s)}ds - 1 \right) \\
&= -\phi(t)e^{-(\Gamma(T)-\Gamma(t))} \leq 0.
\end{aligned}$$

We proceed with two auxiliary lemmas.

Lemma 6.6.1. *Assume that $\beta(t) \leq 2L$ for all $t \in [0, T]$. Then $\tau = T$ is an optimal stopping time of* (6.26).

Proof. For $s \geq t$ we have

$$E[1_{\{s<\tilde{\tau}\}}|\mathcal{F}_t] = 1_{\{t<\tilde{\tau}\}}P(s < \tilde{\tau}|t < \tilde{\tau}).$$

Thus we obtain for all $t \in [0, T]$

$$E[A_T - A_t|\mathcal{F}_t] = \int_t^T \phi(s)E[1_{\{s<\tilde{\tau}\}}|\mathcal{F}_t]ds - 2L = 1_{\{t<\tilde{\tau}\}}\beta(t) - 2L \leq 0.$$

This implies that the Snell envelope U of A is given by $U_t = E[A_T|\mathcal{F}_t]$ and that $\tau = T$ is optimal. □

Lemma 6.6.2. *Let $a = \inf\{t \geq 0 | \beta(t) \leq 2L\} \leq T$. Then we have for all $t \leq r \leq a$*

$$\int_t^r \phi(s) P(s < \tilde{\tau} | t < \tilde{\tau}) ds \geq 2LP(\tilde{\tau} \leq r | \tilde{\tau} > t). \tag{6.27}$$

Proof. Note first that (6.27) is equivalent to

$$\int_t^r \phi(s) P(s < \tilde{\tau}) ds \geq 2LP(t < \tilde{\tau} \leq r).$$

Since β is nonincreasing we have

$$\begin{aligned} \int_t^r \phi(s) P(s < \tilde{\tau}) ds &= P(t < \tilde{\tau})\beta(t) - P(r < \tilde{\tau})\beta(r) \\ &\geq P(t < \tilde{\tau} \leq r)\beta(r) \\ &\geq 2LP(t < \tilde{\tau} \leq r), \end{aligned}$$

which completes the proof. □

Proof of Proposition 6.3.1. If $\beta(0) < 2L$, then Lemma 6.6.1 implies that $\tau^* = T$ is an optimal stopping time.

For the rest of the proof assume that $\beta(0) \geq 2L$. Let $N \in \mathbb{N}$ and

$$\Delta = \{0 = t_0 < t_1 < \cdots < t_N = a\}$$

be a finite partition of $[0, a]$, where $a = \inf\{t \geq 0 | \beta(t) \leq 2L\} \leq T$. We denote by $(U_t^\Delta)_{0 \leq t \leq T}$ the Snell envelope of the stopping problem

$$\inf_{\tau \in \Delta \cup [a,T]} E[A_\tau].$$

We write $U_i^\Delta = U_{t_i}^\Delta$. Then by Lemma 6.6.1,

$$U_N^\Delta = U_a^\Delta = E[A_T | \mathcal{F}_a],$$

and by definition for $0 \leq i \leq N-1$

$$U_i^\Delta = E[U_{i+1}^\Delta | \mathcal{F}_{t_i}] \wedge A_{t_i}.$$

We next show that

$$U_i^\Delta = A_{t_i} 1_{\{\tilde{\tau} > t_i\}} + E[A_T | \mathcal{F}_{t_i}] 1_{\{\tilde{\tau} \leq t_i\}} \tag{6.28}$$

for all $0 \leq i \leq N$. In particular, this implies $U_0^\Delta = A_0$. Hence, $\tau^* = 0$ is optimal among all stopping times taking values in $\Delta \cup [a, T]$. We prove (6.28) by backwards induction. For $i = N$ we have

$$E[A_T - A_a | \mathcal{F}_a] = -2L 1_{\{\tilde{\tau} \leq a\}}.$$

Hence $E[A_T|\mathcal{F}_a] = A_a$ on $\{\tilde\tau > a\}$, which implies (6.28). Now let now $0 \le i \le N-1$. On $\{\tilde\tau \le t_i\}$ we have $\{\tilde\tau \le t_{i+1}\}$, which implies

$$E[U^\Delta_{i+1} - A_{t_i}|\mathcal{F}_{t_i}] = E[A_T - A_{t_i}|\mathcal{F}_{t_i}] = -L < 0.$$

Hence, $U^\Delta_i = E[A_T|\mathcal{F}_{t_i}]$ on $\{\tilde\tau \le t_i\}$. On $\{\tilde\tau > t_i\}$ we have

$$U^\Delta_{i+1} = A_{t_{i+1}} - 2L1_{\{\tilde\tau \le t_{i+1}\}} = A_{t_i} + \int_{t_i}^{t_{i+1}} \phi(s)1_{\{s<\tilde\tau\}}ds - 2L1_{\{\tilde\tau \le t_{i+1}\}}.$$

This implies on $\{\tilde\tau > t_i\}$

$$E[U^\Delta_{i+1}|\mathcal{F}_{t_{i+1}}] = A_{t_i} + \int_{t_i}^{t_{i+1}} \phi(s)P(s < \tilde\tau|t_i < \tilde\tau)ds - 2LP(\tilde\tau \le t_{i+1}|\tilde\tau > t_i).$$

Equation (6.28) now follows from Lemma 6.6.2.

It remains to show that $\tau^* = 0$ is also optimal among all stopping times taking values in $[0,T]$. Let τ be such a stopping time and define

$$\tau^N = \begin{cases} \frac{k}{N}a & \text{if } \tau \in [\frac{k-1}{N}a\frac{k}{N}a) \text{ for a } 1 \le k \le N, \\ \tau & \text{else.} \end{cases}$$

Notice that $\lim_{N\to\infty} E[A_{\tau^N}] = E[A_\tau]$. The fact that $\tau \ne 0$ a.s. and $E[A_\tau] < E[A_0]$ yield a contradiction to the optimality of $\tau^* = 0$ among the stopping times taking only finitely many values on $[0,a]$. □

Next we provide the proof of Proposition 6.3.2.

Proof of Proposition 6.3.2. Fix $x \le 0$ and consider

$$\begin{aligned} \phi(t) &:= E[K_{\tau_t} - \alpha\tau_t] \\ &= K_+P[\tilde\tau > t] + K_-P[\tilde\tau \le t] - \alpha\left(E[\tilde\tau 1_{\{\tilde\tau \le t\}}] + tP[\tilde\tau > t]\right) \end{aligned}$$

on $[0,T]$. We readily compute

$$P[\tilde\tau \le t] = P[\Gamma^{-1}(\xi) \in [0,t]] = 1 - \exp(-\Gamma(t_0))$$

and

$$P[\tilde\tau > t] = \exp(-\Gamma(t)).$$

Finally,

$$E[\tilde\tau\mathbf{1}_{\{\tilde\tau \le t\}}] = \int \Gamma^{-1}(\xi)\mathbf{1}_{[0,t]}(\Gamma^{-1}(\xi))dP = \int_0^{\Gamma(t)} \Gamma^{-1}(s)\exp(-s)ds.$$

Hence we have

$$\phi'(t) = \exp(-\Gamma(t))\left(-\alpha - \gamma(t)(K_+ - K_-)\right)$$

and $\phi'(t) = 0$ if and only if

$$-\alpha = \gamma(t)(K_+ - K_-).$$

Furthermore,

$$\phi''(t) = -\gamma(t)\phi'(t) + \exp(-\Gamma(t))\left(-\gamma'(t)(K_+ - K_-)\right).$$

Therefore, if γ is strictly increasing on $[0,T]$, the unique local extremum of ϕ is a maximum. Hence, $\phi(t)$ attains its minimum at 0 or T.

Notice that α is nondecreasing on $\mathbb{R}_-$. Therefore, for $x \le \bar{x}$ the minimum is attained at $t = 0$, and for $x > \bar{x}$ it is attained at $t = T$. □

Let us now turn to the proof of Proposition 6.3.3. We introduce the function $H : \mathbb{R}_- \times \mathbb{R}_+ \to \mathbb{R}_+$,

$$H(x,y) = \lambda\sqrt{\bar{f}(x,y)} = \begin{cases} -\lambda\sigma_1 x\sqrt{1-\rho^2} & \text{if } y \ge -hx, \\ \lambda\sqrt{\sigma_1^2 x^2 + 2\rho\sigma_1\sigma_2 xy + \sigma_2^2 y^2} & \text{if } y < -hx. \end{cases}$$

By Proposition 6.3.1, cross hedging strategies are nonincreasing after a possible jump at time 0. Hence, Equation (6.16) yields that the value function is given by

$$v(x_0) = \inf_{y\ge 0} w(x_0,y) + 2Ly$$

with $w(x_0,y)$ defined as in (6.17). Proposition 6.3.2 implies that w is given by

$$w(x_0,y) = K_0(x(y)-x_0)_+ - E[K_{\tilde{\tau}}]\max(x_0,x(y)) + E[\tilde{\tau}]H(\max(x_0,x(y)),y)$$

with

$$x(y) = \max\{x \le 0 | K_0 \le E[K_{\tilde{\tau}}] - H_x(x,y)E[\tilde{\tau}]\}, \tag{6.29}$$

and that the infimum of w is attained for the position process $X_t = (x(y) \vee x_0)1_{\{t<\tilde{\tau}\}}$. The proof of Proposition 6.3.3 consists of computing $x(y)$ and $w(x_0,y)$ explicitly for the particular cases and determining $y^* \ge 0$ satisfying $v(x_0) = w(x_0,y^*) + 2Ly^*$ afterward.

Proof of Proposition 6.3.3. Before proving the statements we notice that

$$H_x(x,y) = \begin{cases} -\lambda\sigma_1\sqrt{1-\rho^2} & \text{if } y \ge -hx, \\ \lambda\frac{\sigma_1^2 x + 2\rho\sigma_1\sigma_2 y}{\sqrt{\sigma_1^2 x^2 + 2\rho\sigma_1\sigma_2 xy + \sigma_2^2 y^2}} & \text{if } y < -hx. \end{cases}$$

We now show the three statements separately.

1. Since H_x is bounded from above by $-\lambda\sigma_1\sqrt{1-\rho^2}$, we have $x(y) = 0$. Hence, $x^* = 0$. This implies $w(x_0,y) = -K_0 x_0$ and $y^* = 0$.

2. Since H_x is bounded from below by $-\lambda\sigma_1$, we have $x(y) = -\infty$, which implies $x^* = x_0$. Hence, we have $w(x_0, y) = -E[K_{\tilde{\tau}}]x_0 + E[\tilde{\tau}]H(x_0, y)$. Note that

$$H_y(x_0, y) = \lambda\sigma_2^2 \frac{hx_0 + y}{\sqrt{\sigma_1^2 x_0^2 + 2\rho\sigma_1\sigma_2 x_0 y + \sigma_2^2 y^2}}$$

for $y < -hx_0$, which is increasing in y with $H_y(x_0, -hx_0) = 0$ and $H_y(x_0, 0) = -\lambda\rho\sigma_2$. So, if $2L \geq \lambda\sigma_2\rho E[\tilde{\tau}]$, then $w(x_0, y) + 2Ly$ attains its minimum at $y^* = 0$. Else, y^* is the solution of $E[\tilde{\tau}]H_y(x_0, y^*) = -2L$ on $[0, -hx_0]$, i.e.

$$y^* = -\frac{\sigma_1}{\sigma_2}\left(\rho - 2L\frac{\sqrt{1-\rho^2}}{\sqrt{(\lambda^2\sigma_2^2 E[\tilde{\tau}]^2 - 4L^2)^+}}\right)x_0.$$

3. Note that by Equation (6.29), $x(y)$ is given implicitly by the solution of $\Delta K = -H_x(x(y), y)E[\tilde{\tau}]$ on $(-\infty, -y/h]$; hence, $x(y) = -\alpha y$ with

$$\alpha = \frac{\sigma_2}{\sigma_1}\left(\rho + \frac{\sqrt{1-\rho^2}}{\sqrt{\lambda^2\sigma_1^2 E[\tilde{\tau}]^2 - \Delta K^2}}\Delta K\right) \in [1/h, \infty).$$

For $y \in [0, -x_0/\alpha]$ (which is equivalent to $x_0 \leq x(y)$) we have

$$\begin{aligned} w(x_0, y) + 2Ly &= K_0(-x_0 - \alpha y) + \alpha E[K_{\tilde{\tau}}]y + E[\tilde{\tau}]H(-\alpha y, y) + 2Ly \\ &= -K_0 x_0 + (-\alpha\Delta K + \lambda E[\tilde{\tau}]\sqrt{\sigma_1^2\alpha^2 - 2\rho\alpha\sigma_1\sigma_2 + \sigma_2^2} + 2L)y \\ &= K_0 x_0 + my, \end{aligned}$$

with

$$m = \frac{\sigma_2}{\sigma_1}\left(-\Delta K\rho + \sqrt{1-\rho^2}\sqrt{\lambda^2 E[\tilde{\tau}]^2\sigma_1^2 - \Delta K^2}\right) + 2L.$$

For $y \in [-x_0/\alpha, -hx_0]$ (or equivalently $x_0 \geq x(y)$) we have

$$\begin{aligned} w(x_0, y) + 2Ly &= -E[K_{\tilde{\tau}}]x_0 + E[\tilde{\tau}]H(x_0, y) + 2Ly \\ &= -E[K_{\tilde{\tau}}]x_0 + \lambda E[\tilde{\tau}]\sqrt{\sigma_1^2 x_0^2 + 2\rho\sigma_1\sigma_2 x_0 y + \sigma_2^2 y^2} + 2Ly. \end{aligned}$$

We verify the following claim at the end of the proof.

Claim: *The mapping $y \mapsto w_y(x_0, y) + 2L$ is continuous and nondecreasing on $[0, -hx_0]$. Moreover it is constant equal to m on $[0, -x_0/\alpha]$ and satisfies*

$$w_y(x_0, -hx_0) + 2L = 2L > 0.$$

If $m \geq 0$, then $y \mapsto w(x_0, y) + 2Ly$ is nondecreasing on $[0, -hx_0]$. This implies $y^* = 0$ as well as $x^* = x(y^*) = 0$. Else y^* is the solution of

$$E[\tilde{\tau}]H_y(x_0, y^*) = -2L$$

on $[-x_0/\alpha, -hx_0]$. A straightforward calculation yields

$$y^* = -\frac{\sigma_1}{\sigma_2}\left(\rho - 2L\frac{\sqrt{1-\rho^2}}{\sqrt{(\lambda^2\sigma_2^2 E[\tilde{\tau}]^2 - 4L^2)^+}}\right)x_0.$$

Note that $y^* \geq -x_0/\alpha$ implies $x(y^*) = -\alpha y^* \leq x_0$. Hence, we have $x^* = x_0$.

It remains to prove the claim. Since $y \mapsto H(x_0, y)$ is convex, monotonicity follows immediately. It is hence sufficient to show continuity of $y \mapsto w_y(x_0, y) + 2L$; more precisely,

$$\lim_{y \searrow -x_0/\alpha} w_y(x_0, y) + 2L = m. \tag{6.30}$$

We have

$$w_y(x_0, y) = \lambda E[\tilde{\tau}]\frac{\rho\sigma_1\sigma_2 x_0 + \sigma_2^2 y}{\sqrt{\sigma_1^2 x_0^2 - 2\rho\sigma_1\sigma_2 x_0 y + \sigma_2^2 y^2}} \to -\lambda E[\tilde{\tau}]\frac{\rho\sigma_1\sigma_2\alpha - \sigma_2^2}{\sqrt{\sigma_1^2\alpha^2 - 2\rho\sigma_1\sigma_2\alpha + \sigma_2^2}}$$

as $y \searrow -x_0/\alpha$. By its defining property, α satisfies

$$\begin{aligned}
\Delta K &= -H_x(-\alpha y, y)E[\tilde{\tau}] \\
&= -\lambda E[\tilde{\tau}]\frac{\rho\sigma_1\sigma_2 - \sigma_1^2\alpha}{\sqrt{\sigma_1^2\alpha^2 - 2\rho\sigma_1\sigma_2\alpha + \sigma_2^2}} \\
&= \frac{\lambda E[\tilde{\tau}]}{\alpha}\left(\sqrt{\sigma_1^2\alpha^2 - 2\rho\sigma_1\sigma_2\alpha + \sigma_2^2} + \frac{\rho\sigma_1\sigma_2\alpha - \sigma_2^2}{\sqrt{\sigma_1^2\alpha^2 - 2\rho\sigma_1\sigma_2\alpha + \sigma_2^2}}\right).
\end{aligned}$$

But this is equivalent to

$$\begin{aligned}
\lambda E[\tilde{\tau}]\frac{\rho\sigma_1\sigma_2\alpha - \sigma_2^2}{\sqrt{\sigma_1^2\alpha^2 - 2\rho\sigma_1\sigma_2\alpha + \sigma_2^2}} &= \alpha\Delta K - \lambda E[\tilde{\tau}]\sqrt{\sigma_1^2\alpha^2 - 2\rho\sigma_1\sigma_2\alpha + \sigma_2^2} \\
&= -m + 2L,
\end{aligned}$$

which implies (6.30).

□

6.6.2. Proofs of the results of Section 6.4

Proposition 6.4.1 is obtained by performing calculations similar to those in the proof of Proposition 6.3.3.

Proof of Corollary 6.4.2. Observe that if $\Delta K \leq \sqrt{1-\rho^2}\lambda\sigma_1 T$, then $\bar{L} \leq 0$. Since $L \geq 0$, we also have $\bar{L} \leq L$. The following implication, therefore, holds true:

$$L < \bar{L} \Longrightarrow \Delta K > \sqrt{1-\rho^2}\lambda\sigma_1 T. \tag{6.31}$$

Next we show

$$L \geq \bar{L} \text{ and } \Delta K \geq \lambda\sigma_1 T \Longrightarrow A = 0. \tag{6.32}$$

Assume that $L \geq \bar{L}$ and $\Delta K \geq \lambda\sigma_1 T$. In order to prove (6.32) it suffices to show that

$$L \geq \frac{\rho}{2}\frac{\sqrt{(\lambda^2\sigma_2^2 T^2 - 4L^2)^+}}{\sqrt{1-\rho^2}}. \tag{6.33}$$

Notice that $L \geq \bar{L} \geq \frac{\rho}{2}\lambda\sigma_2 T$. Moreover,

$$\frac{\rho}{2}\frac{\sqrt{(\lambda^2\sigma_2^2 T^2 - 4L^2)^+}}{\sqrt{1-\rho^2}} \leq \frac{\rho}{2}\frac{\sqrt{(\lambda^2\sigma_2^2 T^2 - \rho^2\lambda^2\sigma_2^2 T^2)^+}}{\sqrt{1-\rho^2}} = \frac{\rho}{2}\lambda\sigma_2 T,$$

which yields Inequality (6.33).

We now prove the statements of Corollary 6.4.2. Assume first that $L \geq \bar{L}$. Implication (6.32) shows that in case (C2) of Proposition 6.4.1 we have $A = 0$ and hence that $y* = 0$. It is therefore never optimal to cross hedge in this case. The primary position is kept open if $\Delta K \geq \lambda\sigma_1 T$ (case (C2)). If $\Delta K < \lambda\sigma_1 T$, then it is optimal to close the primary position immediately (cases (C1) and (C3)).

Next assume that $L < \bar{L}$. From Implication (6.31) we know that in this case $\Delta K > \sqrt{1-\rho^2}\lambda\sigma_1 T$. Cases (C2) and (C4) further imply that $x^* = x_0$ and $y^* = A$. □

6.6.3. Proof of the result of Section 6.5

Proof of Proposition 6.5.1. By Corollary 6.2.8 we need to consider for fixed $x \in [x_0, 0]$ the stopping problem $\inf_{\tau\in[0,T]} E[\xi_\tau^x]$ with $\xi_t^x = K_t - tG'(x)$. The process ξ^x satisfies

$$\begin{aligned} d\xi_t^x = dK_t - G'(x)dt &= -\left(\frac{K_t - K_-}{T-t} + G'(x)\right)dt + \sigma dW_t \\ &= -\frac{\xi_t^x - (K_- - TG'(x))}{T-t}dt + \sigma dW_t. \end{aligned}$$

Hence ξ^x is as well a Brownian bridge starting in K_+ and ending in $K_- - TG'(x)$ at time T. By [24] or [77] an optimal stopping time for $\inf_{\tau\in[0,T]} E[\xi_\tau^x]$ is given by

$$\begin{aligned} \tau(x) &= \inf\{t \geq 0 | \xi_t^x \leq K_- - TG'(x) - \sigma B\sqrt{T-t}\} \\ &= \inf\{t \geq 0 | Z_t \leq G'(x)\}. \end{aligned}$$

Note that convexity of G implies that the family of stopping times $(\tau(x))_{x\in[x_0,0]}$ is non-decreasing in x. Hence, by Corollary 6.2.8 the optimal primary position path can be recovered as the right-continuous inverse of $(\tau(x))_{x\in[x_0,0]}$. If G is strictly convex, then G' is invertible and we obtain Equation (6.25). If G is linear, we have $G'(x) = -\lambda\sqrt{1-\rho^2}\sigma_1$ and the stopping problems do not depend on x. Therefore it is optimal to close the position in one go at time $\tau(0)$. By Corollary 6.2.8 the optimal cross hedge is given by $Y_t = -hX_t$. □

Bibliography

[1] A. Alfonsi, A. Fruth, and A. Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2010.

[2] R. Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1):1–18, 2003.

[3] R. Almgren. Optimal trading with stochastic liquidity and volatility. *SIAM Journal on Financial Mathematics*, 3(1):163–181, 2012.

[4] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3(1):5–39, 2000.

[5] S. Ankirchner, M. Jeanblanc, and T. Kruse. BSDEs with singular terminal condition and control problems with constraints. *arXiv Preprint 1305.6541*, 2013.

[6] S. Ankirchner, P. Kratz, and T. Kruse. Hedging forward positions: Basis risk versus liquidity costs. *SIAM Journal on Financial Mathematics*, 4(1):668–696, 2013.

[7] S. Ankirchner and T. Kruse. Optimal trade execution under price-sensitive risk preferences. *Quantitative Finance*, 13(9):1395–1409, 2013.

[8] S. Ankirchner and T. Kruse. Price-sensitive liquidation in continuous-time. *SSRN Preprint 1968430*, 2013.

[9] P. Bank and A. Fruth. Optimal order scheduling for deterministic liquidity patterns. *Preprint*, 2012.

[10] A. Bensoussan and J. L. Lions. *Applications of variational inequalities in stochastic control*. Elsevier, 1982.

[11] D. Bergemann and J. Välimäki. The dynamic pivot mechanism. *Econometrica*, 78(2):771–789, 2010.

[12] D. Bertsimas and A. W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

[13] L. Bo and R. Yao. Strong comparison result for a class of reflected stochastic differential equations with non-Lipschitzian coefficients. *Frontiers of Mathematics in China*, 2(1):73–85, 2007.

[14] F. Boetius and M. Kohlmann. Connections between optimal stopping and singular stochastic control. *Stochastic Processes and their Applications*, 77(2):253–281, 1998.

[15] M. Bossy, E. Gobet, and D. Talay. Symmetrized Euler scheme for an efficient approximation of reflected diffusions. *Journal of Applied Probability*, 41(3):877–889, 2004.

[16] B. Boyer, T. Mitton, and K. Vorkink. Expected idiosyncratic skewness. *Review of Financial Studies*, 23(1):169–202, 2010.

[17] P. Carr, R. Jarrow, and R. Myneni. Alternative characterizations of American put options. *Mathematical Finance*, 2(2):87–106, 1992.

[18] M. Chaleyat-Maurel, N. El Karoui, and B. Marchal. Réflexion discontinue et systèmes stochastiques. *The Annals of Probability*, 8(6):1049–1067, 1980.

[19] M. G. Crandall and P. L. Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American Mathematical Society*, 277(1):1–42, 1983.

[20] C. Dellacherie and P. A. Meyer. *Probabilities and potential.* North-Holland Publishing Company, 1978.

[21] H. P. Deutsch. *Derivatives and Internal Models.* Palgrave Macmillan, 2009.

[22] R.K. Dixit and R.S. Pindyck. *Investment under uncertainty.* Princeton University Press, 2008.

[23] S. Ebert and D. Wiesen. Testing for prudence and skewness seeking. *Management Science*, 57(7):1334–1349, 2011.

[24] E. Ekström and H. Wanntorp. Optimal stopping of a Brownian bridge. *Journal of Applied Probability*, 46(1):170–180, 2009.

[25] N. El Karoui and M. Chaleyat-Maurel. Un problème de réflexion et ses applications au temps local et aux équations différentielles stochastiques sur R. Cas continu. *Astérisque*, 52:117–144, 1978.

[26] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Mathematical Finance*, 7(1):1–71, 1997.

[27] B. Felix, O. Woll, and C. Weber. Gas storage valuation under limited market liquidity: An application in germany. *The European Journal of Finance*, 19(7):1–19, 2012.

[28] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*, volume 25 of *Stochastic Modelling and Applied Probability.* Springer, New York, second edition, 2006.

[29] P.A. Forsyth, J.S. Kennedy, S.T. Tse, and H. Windcliff. Optimal trade execution: A mean quadratic variation approach. *Journal of Economic Dynamics and Control*, 36(12):1971 – 1991, 2012.

[30] A. Fruth. *Optimal order execution with stochastic liquidity.* PhD thesis, Technische Universtität Berlin, 2011.

[31] A. Fruth, T. Schöneborn, and M. Urusov. Optimal trade execution and price manipulation in order books with time-varying liquidity. *To appear in Mathematical Finance*, 2013.

[32] N. Garleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *To appear in Journal of Finance*, 2013.

[33] J. Gatheral and A. Schied. Optimal trade execution under geometric Brownian motion in the Almgren and Chriss framework. *International Journal of Theoretical and Applied Finance*, 14(03):353–368, 2011.

[34] P. Graewe, U. Horst, and E. Séré. Smooth solutions to portfolio liquidation problems under price-sensitive market impact. *arXiv Preprint 1309.0474*, 2013.

[35] R. Guesnerie and J.J. Laffont. A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm. *Journal of Public Economics*, 25(3):329–369, 1984.

[36] X. Guo and P. Tomecek. Connections between singular control and optimal switching. *SIAM Journal on Control and Optimization*, 47(1):421–443, 2008.

[37] X. Guo and M. Zervos. Optimal execution with log-linear price impact and transaction costs. *To appear in SIAM Journal on Financial Mathematics*, 2013.

[38] R. A. Horn and C. R. Johnson. *Matrix analysis.* Cambridge University Press, 1990.

[39] J. C. Hull. *Options, Futures, and Other Derivatives.* Prentice Hall, 7th edition, 2008.

[40] H. Ishii, M. G. Crandall, and P. L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1):1–67, 1992.

[41] S. Jacka. Optimal stopping and the American put. *Mathematical Finance*, 1(2):1–14, 1991.

[42] S. Jacka and J. Lynn. Finite-horizon optimal stopping, obstacle problems and the shape of the continuation region. *Stochastics: An International Journal of Probability and Stochastic Processes*, 39(1):25–42, 1992.

[43] I. Karatzas. A class of singular stochastic control problems. *Advances in Applied Probability*, 15(2):225–254, 1983.

[44] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*. Springer, 2nd edition, 1991.

[45] I. Kim. The analytic valuation of American options. *Review of Financial Studies*, 3(4):547–572, 1990.

[46] D. B. Kotlow. A free boundary problem connected with the optimal stopping problem for diffusion processes. *Transactions of the American Mathematical Society*, 184:457–478, 1973.

[47] P. Kratz. *Optimal liquidation in dark pools in discrete and continuous time*. PhD thesis, Humboldt Universtität zu Berlin, 2011.

[48] P. Kratz and T. Schöneborn. Portfolio liquidation in dark pools in continuous time. *To appear in Mathematical Finance*, 2013.

[49] T. Kruse and P. Strack. Inverse optimal stopping. *Working Paper*, 2013.

[50] T. Kruse and P. Strack. Optimal stopping with private information. *SSRN Preprint 2291937*, 2013.

[51] H. Kunita. Stochastic differential equations based on Lévy processes and stochastic flows of diffeomorphisms. In *Real and stochastic analysis*. Birkhäuser, 2004.

[52] A. Løkka. Optimal liquidation in a limit order book for a risk-averse investor. *To appear in Mathematical Finance*, 2013.

[53] J. Longerstaey and M. Spencer. Riskmetrics technical document. *Morgan Guaranty Trust Company of New York*, 1996.

[54] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic theory*. Oxford University Press, 1995.

[55] J. J. McCall. Economics of information and job search. *The Quarterly Journal of Economics*, 84(1):113–126, 1970.

[56] R. McDonald and D. Siegel. The value of waiting to invest. *The Quarterly Journal of Economics*, 101(4):707–727, 1986.

[57] H. P. McKean. Appendix: A free boundary problem for the heat equation arising from a problem in mathematical economics. *Industrial Management Review*, 6(2):32–39, 1965.

[58] P. Milgrom and I. Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.

[59] B. Mönch. *Strategic trading in illiquid markets*. Springer, 2005.

[60] A. A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1 – 32, 2013.

[61] T. Önskog and K. Nyström. Weak approximation of obliquely reflected diffusions in time-dependent domains. *Journal of Computational Mathematics*, 28(5):579–605, 2010.

[62] É. Pardoux. BSDEs, weak convergence and homogenization of semilinear PDEs. *NATO ASI Series C Mathematical and Physical Sciences-Advanced Study Institute*, 528:503–550, 1999.

[63] A. Pavan, I. Segal, and J. Toikka. Dynamic mechanism design: A Myersonian approach. *Working Paper*, 2013.

[64] A. F. Perold. The implementation shortfall: Paper versus reality. *The Journal of Portfolio Management*, 14(3):4–9, 1998.

[65] G. Peskir. On the American option problem. *Mathematical Finance*, 15(1):169–181, 2005.

[66] G. Peskir. The Russian option: Finite horizon. *Finance and Stochastics*, 9(2):251–267, 2005.

[67] G. Peskir and A. Shiryaev. *Optimal Stopping and Free-Boundary Problems*. Birkhäuser, 2006.

[68] H. Pham. *Continuous-time Stochastic Control and Optimization with Financial Applications*. Springer, 2009.

[69] A. Popier. Backward stochastic differential equations with singular terminal condition. *Stochastic Processes and their Applications*, 116(12):2014–2056, 2006.

[70] A. Popier. Backward stochastic differential equations with random stopping time and singular final condition. *The Annals of Probability*, 35(3):1071–1117, 2007.

[71] P. Protter. *Stochastic Integration and Differential Equations*. Springer, 2005.

[72] M. Rutkowski. Stochastic integral equations with a reflecting barrier. *Demonstratio Mathematica*, 13(2):483–507, 1980.

[73] Y. Saisho. Stochastic differential equations for multi-dimensional domain with reflecting boundary. *Probability Theory and Related Fields*, 74(3):455–477, 1987.

[74] Y. Sannikov. A continuous-time version of the principal-agent problem. *The Review of Economic Studies*, 75(3):957–984, 2008.

[75] A. Schied. A control problem with fuel constraint and Dawson-Watanabe superprocesses. *To appear in the Annals of Applied Probability*, 2013.

[76] A. Schied, T. Schöneborn, and M. Tehranchi. Optimal Basket Liquidation for CARA Investors is Deterministic. *Applied Mathematical Finance*, 17(6):471–489, 2010.

[77] L. A. Shepp. Explicit solutions to some problems of optimal stopping. *The Annals of Mathematical Statistics*, 40(3):993–1010, 1969.

[78] A. N. Shiryaev. *Optimal stopping rules*. Springer, 2007.

[79] L. Slominski and T. Wojciechowski. Stochastic differential equations with jump reflection at time-dependent barriers. *Stochastic Processes and their Applications*, 120(9):1701 – 1721, 2010.

[80] G. J. Stigler. Information in the labor market. *The Journal of Political Economy*, 70(5):94–105, 1962.

[81] S. T. Tse, P. A. Forsyth, J. S. Kennedy, and H. Windcliff. Comparison between the mean-variance optimal and the mean-quadratic-variation optimal trading strategies. *To appear in Applied Mathematical Finance*, 2013.

[82] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.

[83] S. Villeneuve. On threshold strategies and the smooth-fit principle for optimal stopping problems. *Journal of Applied Probability*, 44(1):181–198, 2007.

[84] A. Wald. *Sequential analysis*. John Wiley and Sons, 1947.

[85] S. Watanabe. On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions. *Kyoto Journal of Mathematics*, 11(1):169–180, 1971.

[86] N. Williams. Persistent private information. *Econometrica*, 79(4):1233–1275, 2011.

[87] O. Woll and C. Weber. Portfolio Optimization in Electricity Trading with Limited Liquidity. *SSRN Preprint 1657014*, 2007.

<u>List of Research Papers</u>

Chapter 2 is based on

- T. Kruse and P. Strack. Optimal stopping with private information. SSRN Preprint 2291937, 2013.

Chapter 3 is a revised version of

- T. Kruse and P. Strack. Inverse optimal stopping. Working Paper, 2013.

Chapter 4 is based on

- S. Ankirchner and T. Kruse. Optimal trade execution under price-sensitive risk preferences. Quantitative Finance, 13(9):1395-1409, 2013.

and

- S. Ankirchner and T. Kruse. Price-sensitive liquidation in continuous-time. SSRN Preprint 1968430, 2013.

Chapter 5 is based on

- S. Ankirchner, M. Jeanblanc, and T. Kruse. BSDEs with singular terminal condition and control problems with constraints. arXiv Preprint 1305.6541, 2013.

Chapter 6 is based on

- S. Ankirchner, P. Kratz, and T. Kruse. Hedging forward positions: Basis risk versus liquidity costs. SIAM Journal on Financial Mathematics, 4(1):668-696, 2013.