Risk-averse Nash equilibria for power trading with market frictions



Julien Vaes

Worcester College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

This thesis is dedicated to my beloved grandmother.

À ceux qui nous ont fait rêver dans le passé et dont nous rêverons à l'avenir.

Acknowledgements

First of all, I would like to thank my supervisor, Professor Raphael Hauser, for giving me the opportunity to conduct this research with him in such a great environment. I thank him for challenging me with great questions, for always finding the right words to keep me motivated, and for his humane approach to supervision by letting me develop independently as a researcher.

Secondly, I am extremely grateful to my family. I would like to take the opportunity here to tell my parents that I am extremely proud of them for being able to build this exceptional family relationship that I can always count on. I thank them for welcoming me and taking care of me during the writing of this thesis, it brought back many good memories. I am immensely grateful to my older brother Urbain for taking the time to go through the thesis, for always being interested in discussing my mathematical problems, and for being enthusiastic about the research ideas I could submit to him. I am also grateful to my younger brother Clément for making me feel useful when he called because he needed my help, for making the 3 months of lockdown such a pleasant experience, and for letting me win a few times at tennis to keep me motivated while writing the thesis.

I would also like to thank my friends at the Alan Turing Institute and Oxford (in particular office S2.33 and the Walton Well road Family) who made my research life so enjoyable. I could not say enough words to thank my friends from Centrale and my Belgian friends known as *les Blaireaux*. They were always there to share good times and to make me feel at home when I returned to Belgium. A special thank you to Audrey whose love and encouragement were essential to the success of this thesis.

I would also like to thank the Alan Turing Institute for funding my research and for its presence during all these years. Finally, I am grateful to professors Vincent Legat, Anthony Papavasiliou and Paul Van Dooren for their excellent letters of recommendation that made this life experience possible.

Abstract

The electricity sector in the UK has only been deregulated relatively recently, which means that today there is still limited competition in the electricity market, although the situation has been improving in recent years. As a result of the limited competition, the electricity market is far from perfect as it offers limited liquidity and suffers from high volatility of electricity prices. Therefore, hedging against risk through electricity-related financial derivatives is of paramount importance for electricity producers and wholesalers, who wish to secure the sale or purchase of electricity from minutes to years before a given delivery period. In this thesis, we focus on hedging possibilities by considering the competition and uncertainty inherent in the electricity market.

We first consider the case of the optimal execution problem under an uncertain volume target. We motivate this approach by drawing a parallel with the situation of an electricity wholesaler who must buy electricity but does not yet know the actual demand of its end-users. The trade execution framework is a good compromise between accurate modelling of market frictions and simplicity of the model. We propose two approaches to address this problem: one is based on a one-shot optimisation problem under uncertainty with a predetermined recourse rule, while the other solves a dynamic program where risk is estimated with a dynamic risk measure.

We then turn our attention to the case of a market where several traders simultaneously hedge against risk. We model this situation as a non-cooperative Nash game, where the trades of one player affect the cost functions of all the other players because of market frictions and limited liquidity. We prove that if the market is viable and thus free of price manipulation, then a unique optimal strategy exists for any risk-averse player. We then show how the market equilibrium can be derived.

In the last part of this thesis, we consider the problem of solving generalised variational inequality problems, which is a classical way of reformulating generalised Nash games in order to find subsets of solutions. We introduce a new numerical method that combines the dual extrapolation method with the cutting plane method. This method is shown to be efficient by means of careful numerical experiments.

Contents

Notations and conventions **iv**

Introduction **1**

1 Background material **6**

1.1 Optimal execution problem 6

1.1.1 Framework 7

1.1.2 Price dynamics 8

1.1.3 Trading costs 10

1.1.4 Optimal risk-averse strategies 10

1.1.5 A necessary condition for market viability 11

1.2 Risk measures and coherent risk measures 13

1.3 Game theory 14

1.3.1 Framework 15

1.3.2 Variational inequality 18

1.3.3 Existence of a variational equilibrium 22

1.3.4 Uniqueness of a variational equilibrium 22

2 Optimal trade execution with uncertain volume target **24**

2.1 Introduction 24

2.2 Model 26

2.2.1 Set of optimal strategies 29

2.2.2 Comparison with other approaches 31

2.3 Numerical results 36

2.3.1 Optimal strategies and impact of the liquidity profile 37

2.3.2 Model performance 41

2.3.3 Comparison of the performance of the different frameworks 42

2.3.4 Real world application: power trading 47

2.4 Discussion 51

2.5 Conclusion 53

3 Trade execution with uncertain volume target under a dynamic coherent risk measure **54**
3.1 Introduction . . . 54
3.2 Framework . . . 56
3.2.1 Price dynamics . . . 57
3.2.2 Volume target dynamics . . . 58
3.2.3 Uncertainty structure . . . 58
3.2.4 Trading decisions and trading cost . . . 59
3.2.5 Risk minimisation . . . 60
3.3 Dynamic programming equations . . . 64
3.4 Scenario tree . . . 68
3.5 Practical implementation . . . 75
3.6 Numerical results . . . 80
3.6.1 Synthetic data . . . 80
3.6.2 Real world application: power trading . . . 85
3.7 Conclusion . . . 89

4 Nash equilibrium for risk averse traders with uncertain volume targets **98**
4.1 Introduction . . . 98
4.2 Framework . . . 99
4.2.1 Price dynamics . . . 100
4.2.2 Volume target dynamics . . . 101
4.2.3 Trading strategy . . . 102
4.2.4 Trading cost . . . 103
4.2.5 Risk minimisation . . . 103
4.2.6 Trading game . . . 104
4.2.7 Equivalent game formulation with differentiable cost function . . . 105
4.2.8 A necessary condition for market viability . . . 108
4.3 Nash equilibrium . . . 111
4.3.1 Existence of a Nash equilibrium . . . 113
4.3.2 Uniqueness of the Nash equilibrium . . . 113
4.4 Numerical results . . . 118
4.4.1 Risk-neutral traders . . . 119
4.4.2 Risk-averse traders . . . 119
4.5 Conclusion . . . 121

5 Solving strongly monotone generalised variational inequalities: theoretical and numerical approaches **124**
5.1 Introduction . . . 124
5.2 Preliminaries and notation . . . 126

5.3 Generalised variational inequality . . . 130
5.3.1 Monotonicity property: interpretation . . . 131
5.3.2 Theoretical approach: differential inclusion reformulation . . . 132
5.4 Methods and algorithms for solving strongly monotone GVI . . . 135
5.4.1 Projected subgradient-type (PS) method . . . 138
5.4.2 Original dual extrapolation (ODE) method . . . 140
5.4.3 Simplified dual extrapolation (SDE) method . . . 147
5.4.4 Numerical results (Part I) . . . 151
5.4.5 Cutting plane (CP) method . . . 155
5.4.6 Partial projected subgradient-type method with cutting planes (PPS-CP) 159
5.4.7 Cutting plane method and dual extrapolation method . . . 163
5.4.8 Numerical results (Part II) . . . 168
5.5 Discussion and future research avenues . . . 169
5.6 Conclusion . . . 172

Conclusions **174**

A Support materials related to Chapter 3 **176**
A.1 Auxiliary results . . . 176
A.2 Proofs . . . 178

B Support materials related to Chapter 4 **182**
B.1 Auxiliary results . . . 182
B.2 Proofs . . . 183

C Support materials related to Chapter 5 **187**
C.1 Auxiliary results . . . 187
C.2 Proofs . . . 189

Notations and conventions

Conventions

We use the following notation:

- Vectors and matrices are noted with bold letters, *e.g.* $\mathbf{M}, \boldsymbol{v}$, and their elements with normal font, *e.g.* $M_{i,j}, v_k$.
- Sets are noted with calligraphic capital letters, *e.g.* $\mathcal{S}$.
- Given any discrete random process $\{S_{t_k}\}_{k\in\mathbb{N}}$, we use S_k as a shorthand for S_{t_k}.

Sets

$[\![i, j]\!]$	Set of all the integer numbers between i and j, *i.e.* $\{i, i+1, \ldots, j-1, j\}$
$\mathbb{R}_+$	Set of non-negative real numbers, *i.e.* $\{x \in \mathbb{R} \,\vert\, x \geq 0\}$
$\mathbb{R}_{++}$	Set of positive real numbers, *i.e.* $\{x \in \mathbb{R} \,\vert\, x > 0\}$
$\mathbb{N}_+$	Set of non-negative natural numbers, *i.e.* $\{x \in \mathbb{N} \,\vert\, x \geq 0\}$
$\mathbb{N}_{++}$	Set of positive natural numbers, *i.e.* $\{x \in \mathbb{N} \,\vert\, x > 0\}$
$\prod_{\nu=1}^{n} \mathcal{X}_\nu$	Cartesian product of the sets $\{\mathcal{X}_\nu\}_{\nu\in\{1,\ldots,n\}}$
$\mathbf{bd}(\mathcal{X})$	Boundary of the set $\mathcal{X}$
$\mathbf{cl}(\mathcal{X})$	Closure of the set $\mathcal{X}$
$\mathbf{conv}(\mathcal{X})$	Convex hull of the set $\mathcal{X}$
$\mathrm{Ext}(\mathcal{P})$	Set of extreme points of this polyhedron $\mathcal{P}$
$\mathcal{B}(\boldsymbol{x}\,;\,\epsilon)$	Open ball of radius $\epsilon > 0$ centred at $\boldsymbol{x}$
$\mathcal{B}[\boldsymbol{x}\,;\,\epsilon]$	Closed ball of radius $\epsilon > 0$ centred at $\boldsymbol{x}$
$x \in_R \mathcal{X}$	The element x is chosen uniformly randomly from the set $\mathcal{X}$

Functions

$\boldsymbol{\Psi} : \mathbb{R}^d \rightrightarrows \mathbb{R}^p$	Set-valued map from $\mathbb{R}^d$ to $\mathbb{R}^p$
$\operatorname{sgn}(x)$	Sign of x, *i.e.* $\operatorname{sgn}(x) = 1$ if $x > 0$, -1 if $x < 0$ and 0 otherwise
$\langle \boldsymbol{x}\,,\, \boldsymbol{y} \rangle$	Dot product of $\boldsymbol{x}$ with $\boldsymbol{y}$
$g(\boldsymbol{x}) \geq 0 \perp f(\boldsymbol{y}) \geq 0$	Complementary condition: at least one of the inequalities $g(\boldsymbol{x}) \geq 0$ and $f(\boldsymbol{y}) \geq 0$ must be an equality
$\mathbb{E}^{\mathbb{P}}[A]$	Expectation of the random variable A with respect to $\mathbb{P}$

Matrices

$\boldsymbol{x} < \boldsymbol{y}$	A comparison between vectors should be understood componentwise: $\boldsymbol{x} < \boldsymbol{y} \Leftrightarrow \forall k : x_k < y_k$
$\boldsymbol{x}_{[\![i,j]\!]}$	Vector composed of all the elements $[\![i, j]\!]$ of vector $\boldsymbol{x}$
$\lambda_{\min}(\mathbf{M})$	Smallest eigenvalue of matrix $\mathbf{M}$, in magnitude
$\lambda_{\max}(\mathbf{M})$	Largest eigenvalue of matrix $\mathbf{M}$, in magnitude
$\sigma_{\min}(\mathbf{M})$	Smallest singular value of matrix $\mathbf{M}$, in magnitude
$\sigma_{\max}(\mathbf{M})$	Largest singular value of matrix $\mathbf{M}$, in magnitude
$\mathbf{1}_m$	Vector of size m composed of all ones

Game theory

ν	Player index
θ_ν	Payoff function of player ν
$\boldsymbol{\Theta} \coloneqq \{\theta_\nu\}_{\nu \in [\![1,n]\!]}$	Payoff functions of all the players
$\boldsymbol{\partial\Theta}$	Pseudo-subdifferential set-valued map of the payoff functions
$\boldsymbol{y}_\nu$	Vector of the elements of $\boldsymbol{y}$ related to player ν
$\boldsymbol{y}_{-\nu}$	Vector of the elements of $\boldsymbol{y}$ related to any player other than player ν
$\mathcal{Y}$	Joint feasible set
$\boldsymbol{\mathcal{Y}}_\nu$ or $\boldsymbol{\mathcal{Y}}_\nu(\boldsymbol{y}_{-\nu})$	Reaction set of player ν given $\boldsymbol{y}_{-\nu}$
$\boldsymbol{\mathcal{Y}}$ or $\boldsymbol{\mathcal{Y}}(\boldsymbol{y})$	Reaction set, *i.e.* $\prod_{\nu=1}^{n} \boldsymbol{\mathcal{Y}}_\nu(\boldsymbol{y}_{-\nu})$
$\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$	Game with joint feasible set $\mathcal{Y}$ and payoff functions $\boldsymbol{\Theta}$

Generalised Variational Inequality Problem

$\operatorname{GVI}(\mathbf{K}, \boldsymbol{\Phi})$	Generalised variational inequality problem with parameters $\mathbf{K}$ and $\boldsymbol{\Phi}$
$\operatorname{SOL}(\operatorname{GVI}(\mathbf{K}, \boldsymbol{\Phi}))$	Set of solutions for $\operatorname{GVI}(\mathbf{K}, \boldsymbol{\Phi})$

Introduction

In the second half of the 1990s, the UK electricity sector was deregulated and electricity became a traded commodity (Schwarz, 2022; Green, 2006a). The disappearance of the monopoly in the electricity industry resulted in an increase in the volatility of electricity prices, which has led electricity market players to seek to hedge against this risk by trading financial derivatives. Electricity is not a commodity like any other because of its particularities related to production, storage and transmission. These operational difficulties, combined with the fact that a trader needs to demonstrate financial strength and technical knowledge to be granted access to the electricity market, lead to a market with limited liquidity when compared to other commodities. Therefore, it is essential for electricity market participants, such a power producers and wholesalers, to trade electricity through financial derivatives well in advance of any delivery period in order to circumvent the high trading costs that can be caused by limited market liquidity (Carrion et al., 2007). This is particularly relevant in the current geopolitical situation, where studies predict an energy shortage in the winter of 2022 due to the current geopolitical instability in the world (Benton et al., 2022).

Taking risk into account is extremely important in the electricity market, as the extreme volatility of electricity prices implies that even well-capitalised companies can go bankrupt in the event of large changes in electricity prices (Bessembinder and Lemmon, 2002). To properly hedge against uncertain future electricity prices, it is interesting to study how these prices can be derived and estimated, which is not possible with the classical no-arbitrage approach (Varian, 1987) since the cost-of-carry relationship between spot and forward prices does not have to be satisfied in the case of electricity derivatives (Lucia and Schwartz, 2002). The main methods that have been proposed in the literature to estimate the prices of financial derivatives in the electricity market can be classified into three categories: non-structural, structural and game theory approaches (Troha, 2015). Non-structural approaches neglect the possible reasons that lead to price fluctuations, and instead are limited to engineering a stochastic process that produces changes with a statistical behaviour similar to those observed in the collected data (see for example the approaches developed in Lucia and Schwartz (2002); Espen Benth and Meyer-Brandis (2009); Cartea and Figueroa (2005)). These models are based on stochastic processes whose parameters may be difficult to estimate and neglect the operational constraints of market participants. Therefore, although these models can provide reasonable short-term forecasts, it is difficult to develop a long-term price forecasting model that takes into account the complexity

of the electricity sector. To avoid these problems, some researchers have turned to structural models which, instead of relying exclusively on stochastic variables, also take into account various details defining the state of the market, such as fuel prices, generation capacity constraints, power plant maintenance forecasts, etc. Models of this type are proposed, for instance, in the work of Barlow (2002), whose model applies the principle that supply matches demand, and of Coulon and Howison (2009); Carmona et al. (2013), who have formulated a bid-stack approach capable of accommodating different types of electricity generators. The last type of approach we discuss is game theory models. The application of game theory to the pricing of electricity derivatives was initiated by Bessembinder and Lemmon (2002) to establish a relationship between forward and spot electricity prices. In their work, the authors assumed that participants in the electricity market are risk averse and wish to maximise a trade-off between the expectation and variance of the profit generated. The greater the risk aversion of the participant, the greater the weight assigned to the variance over the expectation. To estimate risk, the covariance matrix of future prices is estimated from the data and the optimal quantity to trade is given with the hedging strategy proposed by Anderson and Danthine (1980). Later, Bühler and Müller-Merbach (2011) extended the two-period model proposed by Bessembinder and Lemmon (2002) in order to fit with a multi-period framework.

Game-theoretic approaches are particularly interesting because they allow a plethora of details to be taken into account, such as the individual physical properties of any electricity wholesaler or producer (*e.g.* production constraints, risk tolerance, efficiency), and thus model reality with great accuracy. Moreover, in such models, electricity prices are often determined on the basis of a market equilibrium. In this spirit, Troha (2015); Troha and Hauser (2014a,b,c) formulated a game-theoretic model including the main prosumers of the UK market with a high level of detail, such as operational constraints (ramping constraints, start-up cost, etc.) or risk aversion of the participants which were estimated based on the historical data (Kraljic et al., 2019), and derived a prediction on the term structure of UK electricity prices. The authors prove the existence and uniqueness of a competitive equilibrium (Debreu, 1952; Arrow and Debreu, 1954) and derive future contract prices based on the shadow prices, which correspond to the dual variables of the market clearing constraints associated to each type of contract. This holds true as the authors have proven the equivalence between the competitive equilibrium and the social welfare optimum, also know as the *invisible hand* paradigm (Smith, 2014). Although the model presented in Troha (2015) demonstrates promising results, as the estimated term structure matches well with historical prices, the model makes questionable or unrealistic assumptions: Firstly, it assumes a perfect forecast of future electricity demand, imports through interconnectors and renewable energy generation. It is therefore not subject to any source of uncertainty. Secondly, risk is only partially taken into account as it is modelled by a penalty term that depends on the covariance matrix of electricity prices estimated on past data as in the approaches of Bessembinder and Lemmon (2002); Bühler and Müller-Merbach (2011). The resulting optimisation problem solved by each prosumer is not stochastic and the determined

traded quantities are therefore deterministic and do not adapt to available future information such as updated forecasts of electricity demand, power generator failures, electricity production from renewable sources, etc. Thirdly, the market is assumed to be perfect and market participants are assumed to be price takers. However, we know that this is actually not the case as some market participants have market power, although this tends to be less and less the case due to increased competition (Green, 2006a,b). Finally, the market is supposed to be complete and therefore does not suffer from liquidity problems, but limited liquidity can make it more challenging to price financial derivatives as shown in de Maere d'Aertrycke and Smeers (2013).

With regard to risk trading, there are two main approaches in the scientific literature: statistical diversification and portfolio replication (Ralph and Smeers, 2011). The former consists of measuring risk according to some criterion, for example the mean-variance trade-off of portfolio returns (Markowitz, 1952). The second assumes that the market is complete in the sense that there are enough financial instruments to replicate the uncertain payoffs of the endowments of any player. In financial parlance, if a market is complete, there is a probability density function, called a risk-neutral probability measure, such that the price of a forward contract is equal to the expectation of its payoffs. In both methods, a risk-averse agent optimises their trading strategy on the basis of exogenous data on the distribution of future states and the associated price and payoff realisation for each financial instrument, which are consequently independent of the trading strategies of other players. To circumvent this strong assumption, Ralph and Smeers (2011) propose to derive the prices of energy-related financial instruments on the basis of a market equilibrium. Therefore, prices are not predetermined but arise from the market equilibrium and as a consequence depend indirectly on the decisions of all market participants; they are said to be endogenous. Their approach is motivated by two facts: First, risk sharing improves the total risk of the system due to the heterogeneity of individual risk. Second, when agents use coherent risk measures, marginal views of risk are described by PDFs. As a consequence, the market formulates the prices of the financial derivatives (or prices of risk) as a PDF that simultaneously captures all agents views as risk neutral at the margin (Ralph and Smeers, 2011). A similar approach of finding endogenous prices based on a market equilibrium has been applied to various problems in the energy sector (see Ralph and Smeers (2015); Philpott et al. (2016); Ferris and Philpott (2022); de Maere d'Aertrycke and Smeers (2013)). In these works, it is first assumed that the risk market is complete in the sense that there are Arrow-Debreu securities to hedge against any future scenario. In de Maere d'Aertrycke and Smeers (2013), the market completeness assumption is then lifted by introducing liquidity constraints by limiting the number of positions that can be traded for any risky asset. This leads to shared constraints between players and turns the game into a generalised Nash equilibrium problem, which often have multiple equilibria.

In this thesis, our aim is to move away from the limiting assumptions made in Troha (2015) in order to capture the uncertainty inherent in the power sector and to better understand the strategic hedging decisions of market participants. Understanding the hedging behaviour

of market participants is of paramount importance as it directly influences the competitive equilibrium, and thus prices. In order to lift these assumptions, we will assume that an electricity wholesaler is subject to the uncertainty related to the final demand for electricity of end-users and wishes to hedge against the risk of electricity price volatility by trading forward contracts. A similar situation exists for electricity suppliers who are for example subject to power generator failures. We assume that each player estimates risk via a coherent risk measure, which is assumed to be evaluated based on a predetermined physical probability measure on future scenarios. Finally, we assume that the market is neither perfect nor complete. There are different ways to enforce the limited liquidity of a market: de Maere d'Aertrycke and Smeers (2013), for example, impose a limited trading capacity of each type of contract. In trading parlance, the liquidity of a market is often translated as the bid-ask spread and the impact of transactions on the price dynamics (temporary and permanent). We use the optimal execution model presented in Almgren and Chriss (2001) as a trading model to capture some of the micro-structure of the market. Indeed, on the one hand, this model implicitly incorporates the limited liquidity of the market via the temporary impact of trades on the transaction price, which leads to an additional cost compared to the fair price of the trade. This is the effect of walking through the limit order book when submitting a market order and can be compared with the *merit order* principle in the electricity sector, where the best priced orders are matched first. The permanent impact of trades in the execution problem of Almgren and Chriss (2001) can, on the other hand, be paralleled with the market power of market participants, as it corresponds to a price change caused by trades. Hence, even though the underlying price process is assumed to be given, we assume that the trades have an impact on prices dynamics, which implies that the resulting fair price process will be affected by the market equilibrium. The approach can thus be seen as a mid-point between statistical diversification and endogenous pricing. Although more complex optimal execution models exist, we restrict ourselves to the one formulated by Almgren and Chriss (2001) with linear impact functions, as the objective is to integrate the model into the broader framework of electricity trading while being subject to operational constraints. This choice is made in order to keep a good balance between simplicity/ease of use and consistency of modelling. In doing so, we also avoid any shared constraints and do not assume that the decisions of one player constrain the set of strategies of another player via a limited capacity constraint. As a result, the model remains a Nash game (and not a generalised one) with a greater chance of having a unique equilibrium. We finally consider the absence of speculators since purchasers prefer forward contracts backed by producers with market power; the volume of trading in electricity futures exchanges being low compared with bilateral (over-the-counter) trades (Ruddell et al., 2018).

Thesis outline

First, we provide in Chapter 1 some background material articulated in three section that focus on the trade execution problem, risk measures and game theory. In this thesis, we explicitly mention at the beginning of each research chapter the sections on which the content is based. The research I conducted is then articulated as follows:

In Chapter 2, we consider the case of the optimal execution problem when a trader is subject to uncertainty arising from price dynamics on the one hand and volume to be traded on the other. This models the situation of an electricity wholesaler who has to buy electricity while not yet knowing the actual demand of its end users. Our approach is based on a one-shot optimisation problem under uncertainty with a predetermined recourse rule. The resulting optimal trading strategy is shown to be unique if the underlying market is free of price manipulation. The method is easy to implement, yields competitive results compared to current alternatives and can be integrated into a broader framework without difficulties.

Our focus shifts to dynamic programming in Chapter 3. We reformulate the execution problem under both sources of uncertainty using a coherent dynamic risk measure and show that the size of the scenario tree can be significantly reduced under appropriate assumptions. We prove that the value function of each step is a convex sequence and that a large part of the feasible set at any state can be pruned given the optimal decision of another state of the same stage. As a result, we prove that the dynamic program can be solved with quadratic computational complexity, thus avoiding the combinatorial explosion traditionally associated with such an approach. The efficiency of the method is then compared with the one-shot model formulated in Chapter 2.

In Chapter 4 we consider the trade execution problem where several traders simultaneously hedge against risk. We formulate this problem as a non-cooperative Nash game where the optimal trading strategy is obtained with the one-shot approach developed in Chapter 2, with the difference that the trading costs of one player are affected by the trading strategies of all other players due to market frictions. An equivalent game is then derived by lifting the feasible set of each player. We then show the uniqueness of a risk-neutral equilibrium by reformulating the game as a generalised variational inequality problem and by proving that the pseudo-gradient map associated with the lifted game is strictly monotone. The existence of a risk-averse equilibrium is proved, and we propose a conjecture for the uniqueness of a risk-averse equilibrium motivated by numerical experiments.

Finally, in Chapter 5, we focus on methods for solving strongly monotone generalised variational inequality problems (GVIPs). GVIPs are extremely useful since they are a classical way of reformulating generalised Nash games to find subsets of solutions (see Chapter 1). We introduce a new numerical method that has its roots in the dual extrapolation method for variational inequality problems (VIP). The method inherits the fast convergence of the dual extrapolation method. It enables to reach a neighbourhood of the solution and, when combined with a cutting plane type method, is shown to achieve exact convergence.

Chapter 1

Background material

This chapter provides some background material articulated in three sections that focus on the trade execution problem, risk measures and game theory. We will explicitly mention at the beginning of each research chapter of this thesis the sections on which the content is based.

1.1 Optimal execution problem

This section gives a brief overview of the classical optimal execution problem. It introduces the concepts that will be used in the following chapters of this thesis:

- Chapter 2: Optimal trade execution with uncertain volume target;
- Chapter 3: Trade execution with uncertain volume target under a dynamic coherent risk measure;
- and Chapter 4: Nash equilibrium for risk averse traders with uncertain volume targets.

The optimal execution problem refers to the problem of finding the best trading strategy to ensure the transition from one portfolio to another within an allocated period of time. Such trading strategy consists of buy and sell orders exchanged on the dedicated market. This problem of optimal execution is germane to markets where liquidity is insufficient. Indeed, in insufficiently liquid markets, price dynamics are sensitive to large trades. Hence, a rational trader should take into consideration the impact of their own trades, which is assumed adverse to their trading position in the sense that it increases their trading cost. In rest of this section, we formulate the trade execution problem from the point of view of a trader who desires to acquire their position within a fixed time window. This problem is equivalent to the liquidation problem presented in the seminal paper of Almgren and Chriss (2001) up to a change of sign in the temporary and permanent price impacts induced by the trades as presented hereafter.

We should mention that more complex models for the execution problems exist in the literature (Cartea et al., 2015; Gatheral and Schied, 2012; Gueant, 2016). However, as justified

previously, the objective of this work is to capture some of the micro-structure of the market while keeping the model relatively simple, which can then be used in a broader model, where several players compete with each other and are subject to private operational constraints.

1.1.1 Framework

In this work, we will assume that the market has only two types of orders: limit orders (LOs) and market orders (MOs). The collection of all LOs constitutes the limit order book. Each LO reflects the desire of a trader to buy or sell at a given price up to a certain quantity (Cartea et al., 2015). As represented in Figure 1.1, the part of the LOB constituted of all sell limit orders is called the ask side, whereas the part of buy orders is called the bid side. The difference between the best ask and bid prices is called the bid-ask spread and is denoted by b_t at time t. The arithmetic average between the best ask and bid prices is called the mid-price. It will be considered here as the ground truth underlying price of the asset and is denoted by S_t at time t; we call it the fair price.

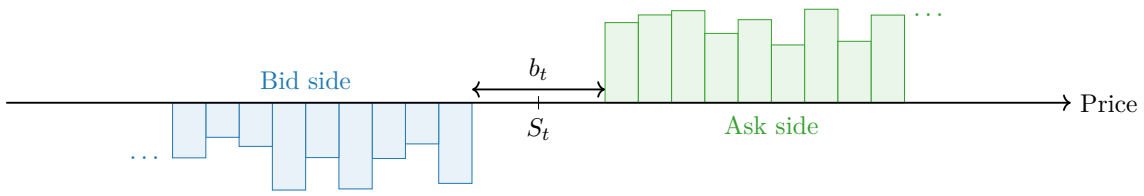


FIGURE 1.1: Illustration of a limit order book

We assume that traders are price takers and that their trading strategies consist exclusively of market orders, which correspond to the desire to acquire a certain quantity of shares at the best price. Market orders are executed by walking the order book in the sense that they are first matched with the LOs at the best price and then paired with the next most affordable LOs until reaching the desired trading volume. It is when the quantity to be traded is too large, in the sense that it cannot be executed in one go, that the execution problem takes on its full meaning. Indeed, if the volume to acquire $d > 0$ is significantly larger than the average size of the limit orders in the book, the matching process will have to dig deep into the LOB and the trader will consequently pay a price per share significantly higher than to the fair price S_t. In such a case, it is interesting to divide the volume to be traded into sub-volumes and to spread their execution over a (fixed) period of time called trading (time) window.

We assume that this trading window spans the time interval $[\, t_s, t_f \,]$ with $t_s < t_f$, where t_s denotes the time at which the trader starts to trade and t_f the time at which the trader must have traded the entire position. Similarly to Almgren and Chriss (2001), the trading window is assumed to be subdivided into m trading periods $[t_{k-1}, t_k)$ for $k \in [\![1, m]\!]$ where $t_0 = t_s$ and $t_m = t_f$. For $k \in [\![1, m]\!]$, we denote by $\tau_k := t_k - t_{k-1}$ the length of the k-th trading period. By a slight abuse of notation, τ_k will also be used to refer to the k-th trading period (see Figure 1.2).

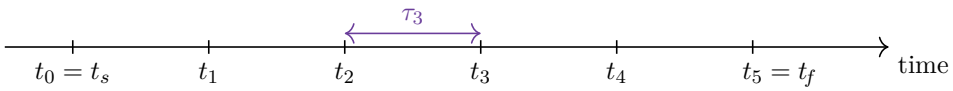


FIGURE 1.2: Illustration of the trading window with 5 trading periods

1.1.2 Price dynamics

Like Almgren and Chriss (2001), we consider that the price dynamics follows an arithmetic random walk, where the price step between decision times t_{k-1} and t_k varies according to the distribution of a random variable ξ_k. Moreover, we assume that the trades have an impact on the price dynamics. It is important that this impact is taken into account since trades executed at the beginning of the trading window have an impact on the fair price of subsequent trades. The nature of the impact of the trades on the price dynamics is of two kinds: first, there is the permanent price impact which refers to the shift of the mid-price due to the temporary exhaustion of the order book. Then, there is the temporary price impact which refers to the price gap between the transaction price and the fair price that is caused by walking the order book. The random variable process $(\xi_k)_{k>0}$ models, among others things, the impact of the trades of the other traders on the market, some public news or information about the economic situation, etc (Ohnishi and Shimoshimizu, 2020).

When the temporary and permanent price impacts induced by the trades are inserted in the price dynamics, we obtain that the prices evolve as follows for $k \in [\![1, m]\!]$:

$$S_k = S_{k-1} + \xi_k + \tau_k g\left(\frac{q_k}{\tau_k}\right), \tag{1.1a}$$

$$\widetilde{S}_k = S_{k-1} + h\left(\frac{q_k}{\tau_k}\right), \tag{1.1b}$$

where S_k is the fair price at decision time t_k, q_k is the volume of shares traded during trading period τ_k (bought if $q_k > 0$ and sold if $q_k < 0$), $\widetilde{S}_k$ is the transaction price for the trades executed during trading period τ_k, and finally g and h respectively model the permanent and temporary price impacts as a function of the average trading rate over the trading interval τ_k.

Gatheral and Schied (2012, Corollary 1) proved that the continuous formulation of this model is free of price manipulations if the following conditions are satisfied: *(i)* g is a linear nondecreasing function, *i.e.* $g(v) = \gamma v$ with $\gamma \geq 0$, and *(ii)* the function $f : x \mapsto xh(x)$ is convex. These conditions are empirically observed as shown in Almgren et al. (2005). For the discrete formulation formulated in Equation (1.1), Huberman and Stanzl (2004, Proposition 2) provide the following necessary conditions for a market with time-independent price-impact functions to be viable:

(i) the function g is linear;

(ii) and the following inequality holds

$$h\left(\frac{q}{\tau}\right) - h\left(\frac{-q}{\tau}\right) \gtreqless g\left(\frac{q}{\tau}\right), \text{ for } q \gtreqless 0, \tag{1.2}$$

where q is the volume traded and τ is the length of each trading period, *i.e.* $\forall k \in [\![1, m]\!]$: $\tau_k = \tau$; an interpretation of this condition is provided later on in Section 1.1.5.

Almgren and Chriss (2001) assume linear permanent and temporary impact functions. We make similar assumptions to the difference that we allow for the liquidity parameters to vary over time. Hence, we have time-dependent price-impact functions. The permanent and temporary impact functions g_k and h_k related to the trading period τ_k are defined as follows

$$g_k(v) = \gamma_k v, \tag{1.3a}$$

$$h_k(v) = \epsilon_k \operatorname{sgn}(v) + \eta_k v, \tag{1.3b}$$

where v is the average trading rate over the trading period τ_k and $\operatorname{sgn}(v)$ is the sign of v. This choice of functions is relevant in that parameters ϵ_k and η_k might be considered as the fixed and variable costs of trading. A reasonable estimate for ϵ_k is the sum of half the bid-ask spread and the trading fees (Almgren and Chriss, 2001). The parameter η_k can be interpreted as the gradient of a linear model for the volumes of orders in the limit order book as a function of the deviation from the best limit order. We assume that this parameter is the same when taking short or long positions, which may not be the case in a real market, but is justified according to statistical analysis of market data (Cartea and Jaimungal, 2016). The price dynamics with linear impact functions is then given by the following set of equations (Figure 1.3 provides a visual interpretation of the assumption of a linear temporary price impact function):

$$S_k = S_{k-1} + \xi_k + \gamma_k q_k, \tag{1.4a}$$

$$\widetilde{S}_k = S_{k-1} + \epsilon_k \operatorname{sgn}(q_k) + \frac{\eta_k}{\tau_k} q_k. \tag{1.4b}$$

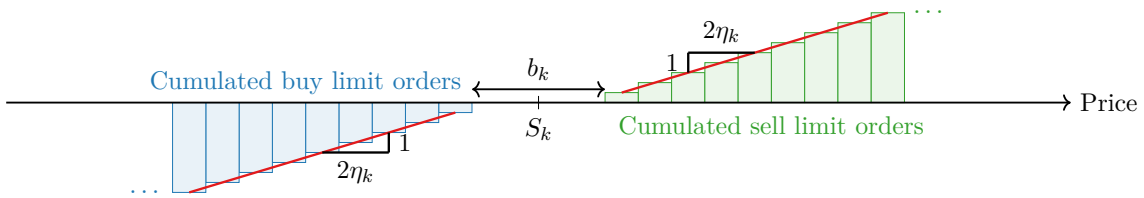


FIGURE 1.3: Illustration of the assumption of a linear temporary impact function.

Equations (1.4a) and (1.4b) capture both the inherent random evolution of the prices as well as the effect of trades on the price dynamics. In the literature, more complex models exist for the price dynamics, which include, for instance, a transient price impact or nonlinear impact functions (Lorenz and Schied, 2013; Almgren, 2003; Curato et al., 2017), but we favour the one formulated here because its simplicity allows us to integrate it into a broader model while providing meaningful interpretations.

Under the assumption of linear impact functions, we can estimate the additional cost caused by the temporary impact of the trades on the transaction price when trading an amount $q_k > 0$ of shares over the trading period τ_k as follows:

$$\epsilon_k q_k + \int_0^{q_k} \frac{2\eta_k}{\tau_k} q \, \mathrm{d}q = \epsilon_k q_k + \frac{\eta_k}{\tau_k} q_k^2 = q_k h_k\left(\frac{q_k}{\tau_k}\right). \tag{1.5}$$

1.1.3 Trading costs

We define the execution strategy $\mathbf{q} \coloneqq [q_1, \ldots, q_m]$ of a trader who must trade a total volume d over the trading window as the volumes q_k of shares to be traded on each trading period τ_k, $k \in [\![1, m]\!]$. To such a strategy, we define the associated *trading costs* or *implementation shortfall* as the following random variable Almgren (2010):

$$\begin{aligned}\mathcal{C}(\mathbf{q}) \coloneqq \sum_{k=1}^{m} q_k \widetilde{S}_k - dS_0 &= \sum_{k=1}^{m} \left(\xi_k + \tau_k g\left(\frac{q_k}{\tau_k}\right)\right)\left(d - \sum_{i=1}^{k} q_i\right) + \sum_{k=1}^{m} q_k h\left(\frac{q_k}{\tau_k}\right), \\ &= \sum_{k=1}^{m} (\xi_k + \gamma_k q_k)\left(d - \sum_{i=1}^{k} q_i\right) + \sum_{k=1}^{m} \epsilon_k \left|q_k\right| + \frac{\eta_k}{\tau_k} q_k^2, \end{aligned} \tag{1.6}$$

where S_0 denotes the asset price at the start of the trading period. In words, the trading costs correspond to the difference between *(i)* the costs incurred at the end of the execution period by following the strategy $\mathbf{q}$ with *(ii)* the costs ideally obtained in an infinitely liquid market where the entire position d is traded at the start of the execution period.

We should emphasise that, unlike most papers in the literature, Equation (1.6) formulates the trading costs from the perspective of a buyer that acquires an amount d of shares over the imparted trading window $[t_s, t_f]$.

1.1.4 Optimal risk-averse strategies

We have seen in the previous section that the trading costs is a random variable. It is consequently sensible for a trader to consider the uncertainty related to the trading costs when defining their trading strategy. To do so, Bianchi and Lo (1998) assumed that a trader defines their strategy as the one that minimises the expected trading costs. Then, by adopting the risk-return trade-off defined by Markowitz (1952) for portfolio hedging, Almgren and Chriss (2001) added the nuance of risk that was missing from the optimal execution problem stated in Bianchi and Lo (1998). Based on the mean-variance framework of Markowitz (1952) and given a risk-aversion parametrised by $\lambda_{\mathbb{V}\mathrm{ar}} \geq 0$, the optimal trading strategy deriving from the model of Almgren and Chriss (2001) is obtained by solving the following optimisation problem:

$$\underset{\mathbf{q}}{\text{minimise}} \quad \mathbb{E}[\mathcal{C}(\mathbf{q})] + \lambda_{\mathbb{V}\mathrm{ar}}\, \mathbb{V}\mathrm{ar}[\mathcal{C}(\mathbf{q})] \tag{1.7a}$$

$$\text{subject to} \quad \sum_{k=1}^{m} q_k = d. \tag{1.7b}$$

Almgren and Chriss (2001) showed that, the more risk-averse a trader, the quicker the acquisition of the desired position should be carried out in order to avoid the risk associated with the price dynamics. These strategies are static and time consistent: They are determined in advance of trading in the sense that they only depend on the information available prior to the start of the execution period. Also, if a trader recomputes the optimal strategy mid-course, the future trading plan remains unchanged since future price fluctuations are assumed independent of the past realisations. Nevertheless, we should mention that if a trader defines as part of their

initial strategy a mid-course updating rule, *aggressive-in-the-money* strategies in the sense of Kissell and Malamut (2005) lead to better risk-return trade-offs (Almgren and Lorenz, 2007; Lorenz and Almgren, 2011).

As the variance penalises both the advantageous and adverse trading costs outcomes without any distinction, some works in the literature suggest to measure risk differently, such as with the *expected utility* (Schied and Schöneborn, 2009; Schöneborn, 2016), or with the α-*Conditional Value-at-Risk* (CVaR_α) (Butenko et al., 2005; Feng et al., 2012; Krokhmal and Uryasev, 2007).

1.1.5 A necessary condition for market viability

As previously mentioned, Huberman and Stanzl (2004) provide the necessary condition (1.2) for a market to be viable if the trading periods are homogeneous, *i.e.* a market with $g_k = g$, $h_k = h$ and $\tau_k = \tau$, for every $k \in [\![1, m]\!]$. In the case of linear impact functions, *i.e.* Equations (1.3a) and (1.3b), Condition (1.2) becomes

$$\left(\epsilon \operatorname{sgn}\left(\frac{q}{\tau}\right) + \eta\left(\frac{q}{\tau}\right)\right) - \left(\epsilon \operatorname{sgn}\left(\frac{-q}{\tau}\right) + \eta\left(\frac{-q}{\tau}\right)\right) \gtreqless \gamma q, \text{ for } q \gtreqless 0. \tag{1.8}$$

This condition intuitively means that the temporary cost incurred by buying and reselling any same amount q of shares on different trading periods should always be greater than the cost saving opportunity engendered by the price shift caused by the permanent impact. For any vector $\boldsymbol{\epsilon} \coloneqq [\epsilon_1, \ldots, \epsilon_m]^{\mathrm{T}}$ such that $\epsilon_k \geq 0$ for all $k \in [\![1, m]\!]$, Condition (1.8) is equivalent to $\eta > \frac{1}{2}\gamma\tau$, which is the sufficient condition for having a unique optimal trading strategy in the model of Almgren and Chriss (2001).

With the rationale that the market should not allow any price manipulation, which is defined as a (deterministic) round-trip strategy with negative expected cost, Condition (1.8) can be generalised in the more general framework where the impact functions g_k and h_k depend on the trading period τ_k.

Definition 1.1 (Round-trip strategy). *A* round-trip strategy $\mathbf{q} \neq \mathbf{0}$ *is a strategy where the sum of all the trades executed during the execution period equates to zero,* i.e. $\sum_{k=1}^{m} q_k = 0$.

Based on Definition 1.1 and on the price manipulation definition from Huberman and Stanzl (2004, Definition 1), we define a price manipulation as follows:

Definition 1.2 (Price manipulation). *A (risk neutral)* price manipulation *is a round-trip strategy* $\mathbf{q}$ *with a strictly negative expected cost,* i.e.

$$\mathbb{E}[\mathcal{C}(\mathbf{q})] < 0. \tag{1.9}$$

Lemma 1.1 (Vaes and Hauser (2022, Lemma 2.3)). *If a market is free of price manipulation, then the matrix* $\mathbf{M} \coloneqq \mathbf{E} - \mathbf{\Gamma}$ *is positive semi-definite,* i.e. $\mathbf{M} \succeq 0$, *where*

$$\mathbf{E} = \begin{bmatrix} \frac{2\eta_1}{\tau_1} + \frac{2\eta_m}{\tau_m} & \frac{2\eta_m}{\tau_m} & \cdots & \frac{2\eta_m}{\tau_m} \\ \frac{2\eta_m}{\tau_m} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{2\eta_m}{\tau_m} \\ \frac{2\eta_m}{\tau_m} & \cdots & \frac{2\eta_m}{\tau_m} & \frac{2\eta_{m-1}}{\tau_{m-1}} + \frac{2\eta_m}{\tau_m} \end{bmatrix}$$

$$\textit{and } \mathbf{\Gamma} = \begin{bmatrix} 2\gamma_1 & \gamma_2 & & \gamma_i & & \gamma_{m-1} \\ \gamma_2 & 2\gamma_2 & & \vdots & & \vdots \\ & & \ddots & \vdots & & \vdots \\ \gamma_i & \cdots & \cdots & 2\gamma_i & & \vdots \\ & & & & \ddots & \vdots \\ \gamma_{m-1} & \cdots & \cdots & \cdots & \cdots & 2\gamma_{m-1} \end{bmatrix}. \quad (1.10)$$

Proof. To prove Lemma 1.1, we show that if matrix $\mathbf{M}$ is not positive semi-definite, then we can find a price manipulation. Given an outcome $\omega \in \Omega$, the trading cost of a round-trip strategy $\mathbf{q}$ is given by

$$\mathcal{C}(\mathbf{q}\,;\,\omega) = \sum_{k=1}^{m} \left[(\xi_k(\omega) + \gamma_k q_k) \left(-\sum_{i=1}^{k} q_i \right) \right] + \sum_{k=1}^{m} \left(\epsilon_k \left| q_k \right| + \frac{\eta_k}{\tau_k} q_k^2 \right). \quad (1.11)$$

As $\mathbf{q}$ defines a round-trip strategy, the trading cost can be expressed in terms of the first $m-1$ components of $\mathbf{q}$ by using the fact that $q_m = -\sum_{k=1}^{m-1} q_k$:

$$\mathcal{C}(\mathbf{q}\,;\,\omega) = \sum_{k=1}^{m-1} \left[(\xi_k(\omega) + \gamma_k q_k) \left(-\sum_{i=1}^{k} q_i \right) \right] + \sum_{k=1}^{m-1} \left(\epsilon_k \left| q_k \right| + \frac{\eta_k}{\tau_k} q_k^2 \right) + \left(\epsilon_m \left| \sum_{k=1}^{m-1} q_k \right| + \frac{\eta_m}{\tau_m} \left(\sum_{k=1}^{m-1} q_k \right)^2 \right). \quad (1.12)$$

This trading cost can be split in two parts: *(i)* $\mathrm{Lin}(\mathbf{q}\,;\,\omega)$ which grows linearly with $\mathbf{q}$ and depends on the uncertainty outcome ω, and *(ii)* a quadratic term in $\mathbf{q}$ independent of ω:

$$\mathcal{C}(\mathbf{q}\,;\,\omega) = \mathrm{Lin}(\mathbf{q}\,;\,\omega) + \sum_{k=1}^{m-1} \left(\frac{\eta_k}{\tau_k} q_k^2 \right) + \frac{\eta_m}{\tau_m} \left(\sum_{k=1}^{m-1} q_k \right)^2 - \sum_{k=1}^{m-1} \left(\gamma_k q_k \left(\sum_{i=1}^{k} q_i \right) \right) \quad (1.13)$$

$$= \mathrm{Lin}(\mathbf{q}\,;\,\omega) + \frac{1}{2} \mathbf{q}_{[\![1,m-1]\!]}^T \, \mathbf{M} \, \mathbf{q}_{[\![1,m-1]\!]}, \quad (1.14)$$

where $\mathbf{q}_{[\![1,m-1]\!]}$ is the vector of the first $m-1$ components of $\mathbf{q}$. If matrix $\mathbf{M}$ is not positive semi-definite, there exists a direction $\boldsymbol{d} \in \mathbb{R}^{m-1}$ along which the quadratic form is concave for every $\omega \in \Omega$ as $\mathbf{M}$ is independent of ω; any normalised eigenvector associated to one of the negative eigenvalues of $\mathbf{M}$ is such a direction. In the following, let ζ^- be a negative eigenvalue of $\mathbf{M}$ and $\boldsymbol{v}$ an associated normalised eigenvector. As a consequence, there exists a $\kappa \in \mathbb{R}$, such that the round-trip strategy $\widetilde{\mathbf{q}} \coloneqq \left[\kappa v_1, \ldots, \kappa v_{m-1}, -\kappa \left(\mathbf{1}^{\mathrm{T}} \boldsymbol{v} \right) \right]^{\mathrm{T}}$ is a price manipulation, where $\mathbf{1}$ is a vector of ones. Indeed, we can find a $\kappa \in \mathbb{R}$, such that $\forall \omega \in \Omega : \mathcal{C}(\widetilde{\mathbf{q}}\,;\,\omega) < 0$, which implies that $\widetilde{\mathbf{q}}$ is a price manipulation, *i.e.* $\mathbb{E}[\mathcal{C}(\widetilde{\mathbf{q}})] < 0$. The existence of such a κ is justified by the fact that when $\kappa \to \infty$, and thus when $\|\widetilde{\mathbf{q}}\|_2 \to \infty$, the trading cost is dominated by its quadratic part, *i.e.* $\mathcal{C}(\widetilde{\mathbf{q}}\,;\,\omega) \simeq \frac{1}{2} \widetilde{\mathbf{q}}_{[\![1,m-1]\!]}^{\mathrm{T}} \, \mathbf{M} \, \widetilde{\mathbf{q}}_{[\![1,m-1]\!]} = \frac{1}{2} \kappa^2 \zeta^- < 0$. □

1.2 Risk measures and coherent risk measures

Given a linear space $\mathcal{Z}$ of measurable functions defined on an appropriate probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a mapping $\varrho : \mathcal{Z} \to \mathbb{R} \cup \{+\infty\}$ is a *risk measure* if it satisfies Axioms 1.1 and 1.2 for all $Z, W \in \mathcal{Z}$ (Föllmer and Schied, 2016, Definition 4.1).

Axiom 1.1 (Monotocity).

$$Z \leq W \Rightarrow \varrho[Z] \leq \varrho[W]\,. \tag{1.15}$$

Axiom 1.2 (Translation invariance).

$$\forall x \in \mathbb{R}: \quad \varrho[Z + x] = \varrho[Z] + x. \tag{1.16}$$

Given $Z \in \mathcal{Z}$ and a risk measure ϱ, $\varrho[Z]$ can be interpreted as the price we would be willing to pay in order to be free of the future risk associated to Z.

Coherent risk measure

A risk measure ϱ is said to be a *coherent risk measure* (CRM) if it additionally satisfies Axioms 1.3 and 1.4 for all $Z, W \in \mathcal{Z}$ (Artzner et al., 1999):

Axiom 1.3 (Convexity).

$$\forall \lambda \in [0,1]: \quad \varrho[\lambda Z + (1-\lambda)\, W] \leq \lambda\, \varrho[Z] + (1-\lambda)\, \varrho[W]\,. \tag{1.17}$$

Axiom 1.4 (Positive homogeneity).

$$\forall \lambda \geq 0: \quad \varrho[\lambda Z] = \lambda\, \varrho[Z]\,. \tag{1.18}$$

This subclass of risk measures is particularly interesting as they have a convenient dual representation: In the case of a finite sample space Ω, any real valued coherent risk measure ϱ has the following *dual representation*:

$$\varrho[Z] = \max_{\mu \in \mathcal{D}} \mathbb{E}^{\mu}[Z]\,, \tag{1.19}$$

where $Z \in \mathcal{Z}$, where $\mathcal{D}$ is called the *risk-set*, which is a closed bounded convex subset of $\mathcal{P}_{\Omega}$, the set of probability measures on Ω (Shapiro et al., 2009, Theorems 6.4 and 6.6), and where $\mathbb{E}^{\mu}[Z]$ denotes the expectation of the random variable Z with respect to μ. The probability measure $\mu^* \in \mathcal{D}$ maximising (1.19) is called the *worst case probability measure*.

Polyhedral coherent risk measure

A coherent risk measure is said to be a *polyhedral coherent risk measure* (PCRM), if, additionally to satisfying Axioms 1.1 to 1.4, the risk-set $\mathcal{D}$ is a polyhedron (Kirilyuk, 2004). This class of risk measures contains the most well known coherent risk measures such as spectral measures. If ϱ is a PCRM, then the linear program (1.19) has an optimal solution located at one of the

extreme points of $\mathcal{D}$, *i.e.* $\mu^* \in \text{Ext}(\mathcal{D})$. Therefore, the dual representation of ϱ can be simplified to

$$\varrho[Z] = \max_{\mu \in \text{Ext}(\mathcal{D})} \mathbb{E}^{\mu}[Z]\,. \tag{1.20}$$

One of the most famous spectral risk measures is the α-Conditional Value-at-Risk (CVaR_α) with $\alpha \in (0,1)$, which is defined as follows (Rockafellar and Uryasev, 2000, 2002):

$$\text{CVaR}_\alpha[Z] = \frac{1}{\alpha}\int_{1-\alpha}^{1} \text{VaR}_\gamma[Z]\,\mathrm{d}\gamma, \tag{1.21}$$

where $\text{VaR}_\gamma[Z]$ is the *value-at-risk* defined as follows:

$$\text{VaR}_\gamma[Z] = \inf\left\{\kappa \in \mathbb{R} \,|\, F_Z(\kappa) > 1-\gamma\right\}, \tag{1.22}$$

where

$$F_Z(\kappa) = \mathbb{P}[Z \leq \kappa] \tag{1.23}$$

is the cumulative distribution function of Z. In words, the CVaR_α risk measure focuses on the proportion α of extreme realisations of Z and can be interpreted as the expectation of Z conditional on exceeding the threshold $\text{VaR}_\alpha[Z]$. When Ω is finite, the risk-set associated to CVaR_α can be formulated as follows:

$$\mathcal{D}_{\text{CVaR}_\alpha} \coloneqq \left\{ \mu \in \mathcal{P}_\Omega \,\middle|\, \forall \omega \in \Omega : 0 \leq \mu(\omega) \leq \frac{\mathbb{P}[\omega]}{\alpha} \right\}. \tag{1.24}$$

We should note that the expectation is also a PCRM where the risk-set contains only the probability measure $\mathbb{P}$.

Finally, the risk measure defined as the trade-off parametrised by λ between the expectation and CVaR_α is denoted by $\mathbb{E}\text{-CVaR}_{\alpha,\lambda}$ and is defined as follows:

$$\mathbb{E}\text{-CVaR}_{\alpha,\lambda}[Z] \coloneqq (1-\lambda)\,\mathbb{E}[Z] + \lambda\,\text{CVaR}_\alpha[Z]\,. \tag{1.25}$$

A parameter value of $\lambda = 0$ corresponds to a risk-neutral player, whereas a value of $\lambda = 1$ corresponds to an exclusively risk focused player. $\mathbb{E}\text{-CVaR}_{\alpha,\lambda}$ is a PCRM and consequently also accepts a dual representation with associated risk-set

$$\mathcal{D}_{\mathbb{E}\text{-CVaR}_{\alpha,\lambda}} \coloneqq \left\{ \mu \in \mathcal{P}_\Omega \,\middle|\, \exists \upsilon \in \mathcal{P}_\Omega \text{ such that } \forall \omega \in \Omega : \begin{cases} \mu(\omega) = (1-\lambda)\,\mathbb{P}[\omega] + \lambda\upsilon(\omega) \\ 0 \leq \upsilon(\omega) \leq \frac{\mathbb{P}[\omega]}{\alpha} \end{cases} \right\}. \tag{1.26}$$

1.3 Game theory

This section gives a brief overview of game theory and introduces the concepts that will be used in the following chapters of this thesis:

- Chapter 4: Nash equilibrium for risk averse traders with uncertain volume targets;
- and Chapter 5: Solving strongly monotone generalised variational inequalities: theoretical and numerical approaches.

We now define what a theoretical game is and the associated notion of a generalised Nash equilibrium (GNE). Then, we show that a subset of GNEs can be obtained as the solutions of a generalised variational inequality (GVI), also called variational equilibrium (VE). Finally, we conclude by providing theoretical results on the existence and uniqueness of a GNE.

In the middle of the 20th century, Nash (1950, 1951) introduced the notion of non-cooperative *Nash equilibrium problem* (NEP), where a finite number of n players compete with one another and respectively desire to minimise their individual objective function. This objective function, also called *payoff function*, represents the preference of the player with regards to the possible outcomes and may depend on the decision variables of all players. In a NEP, the strategy set of a player, which corresponds to the set of all feasible strategies for that player, is independent of the strategies of the other players. To cope with the situation where the strategy set of a player also depends on the strategies of other players, Debreu (1952) introduced the notion of *generalised Nash equilibrium problem* (GNEP). A *(general) Nash equilibrium* (G)NE is a solution to the (G)NEP and is a vector of strategy for all players such that no player has interest in unilaterally changing their strategy.

The study of the existence of generalised Nash equilibria was initiated by Arrow and Debreu (1954) and has since then received much attention in the literature; we refer to Dutang (2013) for a survey on the existence theorems. In some applications, such as in the power market where the market clearing constraint must be satisfied, GNEP only makes sense if there is a unique GNE. Unfortunately, generalised Nash equilibrium problems have most often non unique solutions (Facchinei and Kanzow, 2010). To analyse the uniqueness of a GNE, GNEPs are often formulated as *quasi-variational inequalities* (QVIs); under appropriate assumptions, the uniqueness of a GNE is then equivalent to the uniqueness of the solution of the corresponding QVI (Bensoussan, 1974; Fischer et al., 2014). For jointly convex GNEPs, a relatively well behaved subclass of GNEPs, sufficient conditions exist for the uniqueness of the GNE as we will discuss in the rest of this section. Nowadays, a plethora of algorithms exists in order to find a GNE. This topic is however out of the scope of this thesis; we refer the interested reader to the excellent survey papers on the subject written by Facchinei and Kanzow (2010) and Fischer et al. (2014).

1.3.1 Framework

We assume a n-player game. Each player $\nu \in [\![1, n]\!]$ controls their strategy vector

$$\boldsymbol{y}_\nu \coloneqq \begin{pmatrix} y_{\nu,1} \\ \vdots \\ y_{\nu,d_\nu} \end{pmatrix} \in \mathbb{R}^{d_\nu}, \tag{1.27}$$

where d_ν denotes the number of decision variables of player ν. The vector formed of all the strategies is denoted by

$$\boldsymbol{y} \coloneqq \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix} \in \mathbb{R}^d, \tag{1.28}$$

with $d = \sum_{\nu=1}^{n} d_\nu$. In this work, we interchangeably use the notations $(\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu})$ and $\boldsymbol{y}$ to designate the variables of all the players. The former notation is used when we desire to emphasise the distinction between the decision variables $\boldsymbol{y}_\nu$ of the ν-th player with the ones of the competing players, *i.e.* $\boldsymbol{y}_{-\nu}$.

Each player ν has an objective function, also called *payoff function*, to minimise, which we denote by $\theta_\nu : \mathbb{R}^d \to \mathbb{R}; (\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu}) \mapsto \theta_\nu(\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu})$. The payoff function of a player ν depends on the variables of all players and should not be understood as a two arguments function.

The strategy of a player $\nu \in [\![1, n]\!]$ is assumed to be constrained to belong to a set $\mathcal{X}_\nu$, which represents the private constraints of the player that only depends on the player's own variables $\boldsymbol{y}_\nu$; $\mathcal{X}_\nu$ is thus independent of the variables $\boldsymbol{y}_{-\nu}$ of the other players. The Cartesian product of these sets is denoted by $\mathcal{X} \coloneqq \prod_{\nu=1}^{n} \mathcal{X}_\nu$. In practice, these sets can for instance be regarded as operational constraints of the power units. Additionally, we assume the existence of *shared constraints* that link and restrain the variables a player with regards to the ones of the other players. If $\mathcal{Z} \subseteq \mathbb{R}^d$ denotes the set of simultaneous strategies that respect these shared constraints, then the strategy of a player ν is constrained to belong to the set $\boldsymbol{\mathcal{Z}}_\nu(\boldsymbol{y}_{-\nu}) : \mathbb{R}^{d-d_\nu} \rightrightarrows \mathbb{R}^{d_\nu}$ defined as

$$\boldsymbol{\mathcal{Z}}_\nu(\boldsymbol{y}_{-\nu}) \coloneqq \left\{ \boldsymbol{y}_\nu \in \mathbb{R}^{d_\nu} \,\middle|\, (\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu}) \in \mathcal{Z} \right\}. \tag{1.29}$$

Note that $\boldsymbol{\mathcal{Z}}_\nu(\boldsymbol{y}_{-\nu})$ is a set-valued map that takes as argument the strategies $\boldsymbol{y}_{-\nu}$ of the other players. We should mention that the set $\mathcal{X}_\nu$ of strategies for player ν respecting the private constraints could have been integrated in the set of shared constraints $\mathcal{Z}$. Nevertheless, using the formulation previously presented allows to clearly state that the sets $\mathcal{X}_\nu$ are private information held by each player $\nu \in [\![1, n]\!]$ and that $\mathcal{X}_\nu$ is independent of the strategies of the other players, whereas $\mathcal{Z}$ are constraints shared and known by all players. Figure 1.5 provides an example of these sets in a two-player game, where each player $\nu \in [\![1, 2]\!]$ has only one variable $y_\nu \in \mathbb{R}$.
The set of all feasible strategies, *i.e.* the set of the strategies that respect both the private and shared constraints, is denoted by $\mathcal{Y}$ and defined as follows

$$\mathcal{Y} \coloneqq \mathcal{X} \cap \mathcal{Z} = \left\{ \boldsymbol{y} \in \mathbb{R}^d \,\middle|\, \boldsymbol{y} \in \mathcal{Z} \text{ and } \boldsymbol{y}_\nu \in \mathcal{X}_\nu \text{ for all } \nu \in [\![1, n]\!] \right\}. \tag{1.30}$$

If the variables $\boldsymbol{y}_{-\nu}$ of the other players are fixed, then the objective of player ν is to find a feasible strategy $\boldsymbol{y}_\nu$ in the *reaction set* $\boldsymbol{\mathcal{Y}}_\nu(\boldsymbol{y}_{-\nu}) \coloneqq \mathcal{X}_\nu \cap \boldsymbol{\mathcal{Z}}_\nu(\boldsymbol{y}_{-\nu})$ that minimises their payoff function, *i.e.*

$$\min_{\boldsymbol{y}_\nu \in \mathbb{R}^{d_\nu}} \theta_\nu(\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu}) \text{ such that } \boldsymbol{y}_\nu \in \boldsymbol{\mathcal{Y}}_\nu(\boldsymbol{y}_{-\nu}). \tag{1.31}$$

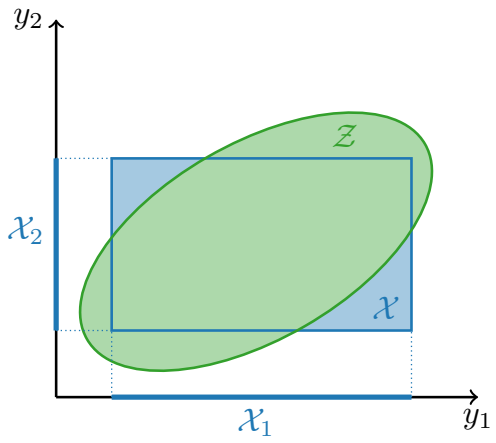


FIGURE 1.4: Illustration of private and shared constraints in a two-player game.

We define the set-valued map $\boldsymbol{\mathcal{Y}} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ representing the reaction set of all players for any feasible strategy $\boldsymbol{y} \in \mathcal{Y}$ as follows:

$$\boldsymbol{\mathcal{Y}}(\boldsymbol{y}) \coloneqq \prod_{\nu=1}^{n} \boldsymbol{\mathcal{Y}}_\nu(\boldsymbol{y}_{-\nu}) \,, \tag{1.32}$$

which corresponds to the Cartesian product of the reaction set of each player, *i.e.* the set of feasible strategies of each player if they assume that the other players stick with their current strategies $\boldsymbol{y}_{-\nu}$. Generally we have that $\mathcal{Y} \not\subseteq \boldsymbol{\mathcal{Y}}(\boldsymbol{y})$ and $\boldsymbol{\mathcal{Y}}(\boldsymbol{y}) \not\subseteq \mathcal{Y}$ (see Figure 1.5).

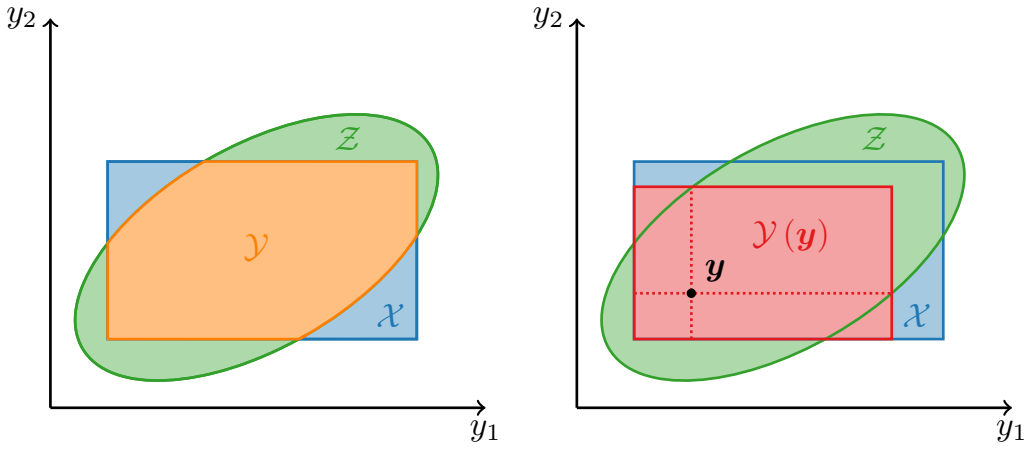


FIGURE 1.5: Illustration of the feasible set and the Cartesian product of the reaction sets in a two-player game.

PROPOSITION 1.1 (VON HEUSINGER AND KANZOW (2009, LEMMA 2.1)). *The following equivalences hold:*

$$\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}(\boldsymbol{y}) \quad \Leftrightarrow \quad \forall \nu \in [\![1, n]\!] : \; \boldsymbol{y}_\nu \in \boldsymbol{\mathcal{Y}}_\nu(\boldsymbol{y}_{-\nu}) \quad \Leftrightarrow \quad \boldsymbol{y} \in \mathcal{Y}. \tag{1.33}$$

DEFINITION 1.3 (GENERALISED NASH EQUILIBRIUM PROBLEM). *The* generalised Nash equilibrium problem (GNEP) *consists of finding a vector* $\boldsymbol{y}^\star \in \mathcal{Y}$ *such that for all* $\nu \in [\![1, n]\!]$*:*

$$\theta_\nu(\boldsymbol{y}_\nu^\star, \boldsymbol{y}_{-\nu}^\star) \leq \theta_\nu(\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu}^\star) \;\; \textit{for all} \;\; \boldsymbol{y}_\nu \in \boldsymbol{\mathcal{Y}}(\boldsymbol{y}_{-\nu}^\star) \,. \tag{1.34}$$

A vector $\boldsymbol{y}^\star$ *satisfying this condition is called* generalised Nash equilibrium (GNE).

In this work, we use the following notation $\boldsymbol{\Theta} \coloneqq \{\theta_\nu\}_{\nu\in[\![1,n]\!]}$ and denote the GNEP defined in Definition 1.3 by $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ or equivalently by $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$. Condition (1.34) is equivalent as imposing that

$$\boldsymbol{y}_\nu^\star \in \underset{\boldsymbol{y}_\nu\in\mathbb{R}^{d_\nu}}{\arg\min}\, \theta_\nu(\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu}^\star) \text{ such that } \boldsymbol{y}_\nu \in \boldsymbol{\mathcal{Y}}(\boldsymbol{y}_{-\nu}^\star)\,, \tag{1.35}$$

which reflects the idea that, at a generalised Nash equilibrium point, no player can decrease their payoff by unilaterally changing of strategy. The game is here assumed to be non-cooperative in the sense that no collusion between players is tolerated. In this thesis, we are interested in problems that satisfy the following assumption:

Assumption 1.1 (Jointly convex). *$\mathcal{Z} \subseteq \mathbb{R}^d$ is a non-empty closed convex set. For every player $\nu \in [\![1, n]\!]$, $\mathcal{X}_\nu \subseteq \mathbb{R}^{d_\nu}$ is a non-empty compact convex set, and the payoff function $\theta_\nu : \mathbb{R}^d \to \mathbb{R}; (\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu}) \mapsto \theta_\nu(\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu})$ is finite on $\mathcal{Y}$, continuous in $\boldsymbol{y}$, and convex in $\boldsymbol{y}_\nu$. A game $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ respecting these conditions is said to be* jointly-convex.

If Assumption 1.1 holds, then $\boldsymbol{\mathcal{Y}}_\nu(\boldsymbol{y}_{-\nu})$ is a convex set for any $\boldsymbol{y}_{-\nu} \in \mathcal{Y}_{-\nu}$. Finally, we should note that in this framework, if $\mathcal{Z} = \mathbb{R}^d$, then there are no shared constraints, and the GNEP reduces to a NEP. We refer to this case by saying that $\boldsymbol{\mathcal{Y}}$ is *orthogonal*: for each player $\nu \in [\![1, n]\!]$, $\boldsymbol{\mathcal{Y}}_\nu(\boldsymbol{y}_{-\nu}) = \mathcal{X}_\nu$. Formulated differently, for every player $\nu \in [\![1, n]\!]$, the orthogonal projection of the set of feasible strategies $\mathcal{Y}$ onto the Euclidean space $\mathbb{R}^{d_\nu}$ is equal to the private feasible set $\mathcal{X}_\nu$, *i.e.*

$$\forall \nu \in [\![1, n]\!]\,, \quad \Pi_{\mathbb{R}^{d_\nu}}(\mathcal{Y}) = \mathcal{X}_\nu. \tag{1.36}$$

1.3.2 Variational inequality

To analyse a GNEP $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$, it is handful to formulate the associated *generalised quasi-variational inequality* (GQVI). In all generalities, a GQVI is defined as follows: given a pair of set-valued maps $(\mathbf{K}, \boldsymbol{\Phi})$, the related generalised quasi-variational inequality denoted GQVI$(\mathbf{K}, \boldsymbol{\Phi})$ is the problem of finding a pair $(\boldsymbol{y}^\star, \boldsymbol{u}^\star)$ with $\boldsymbol{u}^\star \in \boldsymbol{\Phi}(\boldsymbol{y}^\star)$ such that

$$\langle \boldsymbol{u}^\star\,,\, \boldsymbol{y} - \boldsymbol{y}^\star \rangle \geq 0, \text{ for all } \boldsymbol{y} \in \mathbf{K}(\boldsymbol{y}^\star)\,. \tag{1.37}$$

We denote by SOL(GQVI$(\mathbf{K}, \boldsymbol{\Phi})$) the set of solutions of GQVI$(\mathbf{K}, \boldsymbol{\Phi})$, *i.e.*

$$\text{SOL}(\text{GQVI}(\mathbf{K}, \boldsymbol{\Phi})) \coloneqq \{\boldsymbol{y}^\star \,|\, \boldsymbol{y}^\star \in \mathbf{K}(\boldsymbol{y}^\star)\,,\, \exists \boldsymbol{u}^\star \in \boldsymbol{\Phi}(\boldsymbol{y}^\star) \text{ such that (1.37) holds}\}\,. \tag{1.38}$$

By abuse of language, we will also use the notation $(\boldsymbol{y}^\star, \boldsymbol{u}^\star) \in \text{SOL}(\text{GQVI}(\mathbf{K}, \boldsymbol{\Phi}))$ to specify that $\boldsymbol{u}^\star \in \boldsymbol{\Phi}(\boldsymbol{y}^\star)$ and that (1.37) holds for the pair $(\boldsymbol{y}^\star, \boldsymbol{u}^\star)$.

If for all $\boldsymbol{y}$ we have $\mathbf{K}(\boldsymbol{y}) = \mathbf{K}$, then we obtain a *generalised variational inequality* problem noted GVI$(\mathbf{K}, \boldsymbol{\Phi})$; its solution set is then noted SOL(GVI$(\mathbf{K}, \boldsymbol{\Phi})$). To establish the relationship

between a generalised Nash equilibrium problem $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ and a GVI problem, we consider the *pseudo-subdifferential set-valued map* $\boldsymbol{\partial\Theta} : \mathbb{R}^d \supseteq \mathcal{Y} \rightrightarrows \mathbb{R}^d$ defined as follows:

$$\boldsymbol{\partial\Theta}(\boldsymbol{y}) \coloneqq \left\{ \begin{pmatrix} \boldsymbol{g}_1 \\ \vdots \\ \boldsymbol{g}_n \end{pmatrix} \middle| \, \forall \nu \in [\![1, n]\!] : \boldsymbol{g}_\nu \in \boldsymbol{\partial}_\nu \theta_\nu(\boldsymbol{y}) \right\}, \tag{1.39}$$

where for a scalar function f we denote by $\boldsymbol{\partial}_\nu f(\boldsymbol{y})$ the subdifferential with respect to $\boldsymbol{y}_\nu$ of f at the point $\boldsymbol{y}$, *i.e.* the set of subgradients of f with respect to $\boldsymbol{y}_\nu$ at the point $\boldsymbol{y}$. If f is differentiable then $\boldsymbol{\partial}_\nu f(\boldsymbol{y})$ is the singleton $\{\nabla_\nu f(\boldsymbol{y})\}$, where $\nabla_\nu f(\boldsymbol{y})$ corresponds to the gradient of the function f with respect to $\boldsymbol{y}_\nu$ at the point $\boldsymbol{y}$.

With these notations, we can now formulate the following proposition that establish the link between the set of GNEs and the set of solutions to GQVI$(\mathbf{K}, \boldsymbol{\Phi})$ when Assumption 1.1 holds:

Proposition 1.2 (Facchinei and Pang (2009, Proposition 12.3)). *If Assumption 1.1 holds, a vector $\boldsymbol{y}^\star$ is a GNE of the game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$ if and only if there exists a couple $(\boldsymbol{y}^\star, \boldsymbol{u}^\star)$ with $\boldsymbol{u}^\star \in \boldsymbol{\partial\Theta}(\boldsymbol{y}^\star)$ that is a solution of* GQVI$(\boldsymbol{\mathcal{Y}}, \boldsymbol{\partial\Theta})$, *where $\boldsymbol{\mathcal{Y}}$ is the set-valued map defined in* (1.32).

If Assumption 1.1 holds, the condition that $(\boldsymbol{y}^\star, \boldsymbol{u}^\star) \in \mathrm{SOL}(\mathrm{GQVI}(\boldsymbol{\mathcal{Y}}, \boldsymbol{\partial\Theta}))$ is equivalent to

$$\boldsymbol{y}^\star = \Pi_{\boldsymbol{\mathcal{Y}}(\boldsymbol{y}^\star)}(\boldsymbol{y}^\star - \boldsymbol{u}^\star)\,. \tag{1.40}$$

In Proposition 1.2, we have defined the set-valued map $\mathbf{K}$ of the GQVI problem corresponding the GNE problem of the game $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ as the set-valued map $\boldsymbol{\mathcal{Y}}$, which corresponds to the reaction set. If, instead of $\boldsymbol{\mathcal{Y}}$, we use the set of jointly feasible strategies $\mathcal{Y} = \mathcal{X} \cap \mathcal{Z}$, we obtain a generalised variational inequality (GVI) since $\mathbf{K}$ does not depend any more on the point $\boldsymbol{y}$. Studying such a GVI problem is interesting as its solutions form an important subclass of GNEs to $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ called *variational equilibrium* or *normalised NE* (Rosen, 1965).

Definition 1.4 (Variational equilibrium). *A vector $\boldsymbol{y}^\star \in \mathcal{X} \cap \mathcal{Z}$, along with $\boldsymbol{u}^\star \in \boldsymbol{\partial\Theta}(\boldsymbol{y}^\star)$, is a* variational equilibrium *of $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$ if $(\boldsymbol{y}^\star, \boldsymbol{u}^\star) \in \mathrm{SOL}(\mathrm{GVI}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\partial\Theta}))$,* i.e.

$$\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y} - \boldsymbol{y}^\star \rangle \geq 0, \ \textit{for all } \boldsymbol{y} \in \mathcal{X} \cap \mathcal{Z}. \tag{1.41}$$

Based on Definition 1.4, we have that for a jointly-convex game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$, any variational equilibrium is a GNE, *i.e.*

$$\mathrm{SOL}(\mathrm{GVI}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\partial\Theta})) \subseteq \mathrm{GNE}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})\,, \tag{1.42}$$

where $\mathrm{GNE}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$ is the set of all GNEs of the game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$ (Facchinei and Pang, 2009, Proposition 12.4). The converse does however not hold. We prove a more general statement later on in Proposition 1.3.

If Assumption 1.1 holds, we have just seen that the set of variational equilibria of the generalised variational inequality GVI$(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\partial\Theta})$ is a subset of GNEs for the game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$.

We now show that if we slightly modify the GVI problem, we can potentially obtain another subset of GNEs for the game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$: To do so, we consider the variational equilibrium of the modified game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta}_{\boldsymbol{r}})$, where the objective function of each player $\nu \in [\![1, n]\!]$ is scaled by a positive parameter r_ν, *i.e.* $\boldsymbol{\Theta}_{\boldsymbol{r}} := \{r_\nu \theta_\nu\}_{\nu \in [\![1,n]\!]}$ for a given $\boldsymbol{r} \in \mathbb{R}^n_{++}$. The pseudo-subdifferential of this game is noted $\boldsymbol{\partial}\boldsymbol{\Theta}_{\boldsymbol{r}} : \mathbb{R}^d \supseteq \mathcal{Y} \rightrightarrows \mathbb{R}^d$ and is defined as follows:

$$\boldsymbol{\partial}\boldsymbol{\Theta}_{\boldsymbol{r}}(\boldsymbol{y}) := \left\{ \begin{pmatrix} r_1 \boldsymbol{g}_1 \\ \vdots \\ r_n \boldsymbol{g}_n \end{pmatrix} \middle| \, \forall \nu \in [\![1, n]\!] : \boldsymbol{g}_\nu \in \boldsymbol{\partial}_\nu \theta_\nu(\boldsymbol{y}) \right\}. \tag{1.43}$$

PROPOSITION 1.3. *If Assumption 1.1 holds and if $\boldsymbol{r} \in \mathbb{R}^n_{++}$, any variational equilibrium $\boldsymbol{y}^\star$ of $\mathrm{GVI}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\partial}\boldsymbol{\Theta}_{\boldsymbol{r}})$ is a GNE of $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$.*

PROOF. By definition, if $\boldsymbol{y}^\star \in \mathrm{SOL}(\mathrm{GVI}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\partial}\boldsymbol{\Theta}_{\boldsymbol{r}}))$ there exists a $\boldsymbol{u}^{\star,\boldsymbol{r}} := (r_\nu \boldsymbol{u}^\star_\nu)_{\nu=1}^n$ with $\boldsymbol{u}^\star = (\boldsymbol{u}^\star_\nu)_{\nu=1}^n \in \boldsymbol{\partial}\boldsymbol{\Theta}(\boldsymbol{y}^\star)$ such that (1.41) holds. Since for every player $\nu \in [\![1, n]\!]$, the following inclusion holds

$$\left\{ (\boldsymbol{y}_\nu, \boldsymbol{y}^\star_{-\nu}) \,|\, \boldsymbol{y}_\nu \in \mathcal{Y}_\nu(\boldsymbol{y}^\star_{-\nu}) \right\} \subseteq \mathcal{X} \cap \mathcal{Z}, \tag{1.44}$$

and since the couple $(\boldsymbol{y}^\star, \boldsymbol{u}^{\boldsymbol{r},\star})$ is such that (1.41) holds, then $\forall \nu \in [\![1, n]\!]$ and $\forall \boldsymbol{y}_\nu \in \mathcal{Y}_\nu(\boldsymbol{y}^\star_{-\nu})$

$$\langle \boldsymbol{u}^{\star,\boldsymbol{r}} \,,\, \boldsymbol{y} - \boldsymbol{y}^\star \rangle \geq 0 \Rightarrow \langle r_\nu \boldsymbol{u}^\star_\nu \,,\, \boldsymbol{y}_\nu - \boldsymbol{y}^\star_\nu \rangle \geq 0 \Rightarrow \langle \boldsymbol{u}^\star_\nu \,,\, \boldsymbol{y}_\nu - \boldsymbol{y}^\star_\nu \rangle \geq 0. \tag{1.45}$$

Equation (1.45) in turn implies that

$$\forall \boldsymbol{y} \in \mathcal{Y}(\boldsymbol{y}^\star) : \langle \boldsymbol{u}^\star \,,\, \boldsymbol{y} - \boldsymbol{y}^\star \rangle = \sum_{\nu=1}^n \langle \boldsymbol{u}^\star_\nu \,,\, \boldsymbol{y}_\nu - \boldsymbol{y}^\star_\nu \rangle \geq 0, \tag{1.46}$$

which means that $\boldsymbol{y}^\star \in \mathrm{SOL}(\mathrm{GQVI}(\boldsymbol{\mathcal{Y}}, \boldsymbol{\partial}\boldsymbol{\Theta}))$ by definition. Consequently $\boldsymbol{y}^\star$ is a GNE of $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$ with Proposition 1.2. □

With Proposition 1.3, the result of Fischer et al. (2014, Theorem 4.3) can then be generalised to the case of non-differentiable payoff functions as follows:

PROPOSITION 1.4. *If Assumption 1.1 holds, then*

$$\bigcup_{\boldsymbol{r} \in \mathbb{R}^n_{++}} \mathrm{SOL}(\mathrm{GVI}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\partial}\boldsymbol{\Theta}_{\boldsymbol{r}})) \subseteq \mathrm{GNE}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta}). \tag{1.47}$$

PROOF. The proof is straightforward based on Proposition 1.3. □

We have thus demonstrated a useful way to find subsets of GNEs, which consists in finding the variational equilibria of modified games of the form $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta}_{\boldsymbol{r}})$, where the objective function of each player $\nu \in [\![1, n]\!]$ is scaled by a positive parameter r_ν.

Comparison between GNE and VE. We conclude by highlighting the difference between GNEs and variational equilibria. For this purpose, we define the *inner cone* to $\mathcal{X} \cap \mathcal{Z}$ at a point $\boldsymbol{y}^\star$ denoted by $I_{\mathcal{X} \cap \mathcal{Z}}(\boldsymbol{y}^\star)$ as the smallest pointed, closed, convex cone that contains $\left\{ \boldsymbol{s} \in \mathbb{R}^d \,\middle|\, \boldsymbol{y}^\star + \boldsymbol{s} \in \boldsymbol{\mathcal{Y}}(\boldsymbol{y}^\star) \right\}$. If Assumption 1.1 holds, the condition that $(\boldsymbol{y}^\star, \boldsymbol{u}^\star)$ belongs to

SOL(GQVI($\boldsymbol{\mathcal{Y}}, \boldsymbol{\partial\Theta}$)) (and thus a GNE of the game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$ with Proposition 1.2) is equivalent to

$$\Pi_{I_{\mathcal{X}\cap\mathcal{Z}}(\boldsymbol{y}^\star)}(-\boldsymbol{u}^\star) = \mathbf{0}\,. \tag{1.48}$$

The difference between GNEs and variational equilibria arises from the fact that, at $\boldsymbol{y}^\star$, the inner cone $I_{\mathcal{X}\cap\mathcal{Z}}(\boldsymbol{y}^\star)$ is included in the tangent cone of $\mathcal{X} \cap \mathcal{Z}$ (see definition in Facchinei and Pang (2003a)), *i.e.* $I_{\mathcal{X}\cap\mathcal{Z}}(\boldsymbol{y}^\star) \subseteq T_{\mathcal{X}\cap\mathcal{Z}}(\boldsymbol{y}^\star)$. It means in words that, if the game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$ respects Assumption 1.1, $\boldsymbol{y}^\star$ is a GNE if no feasible deviation in the inner cone $I_{\mathcal{X}\cap\mathcal{Z}}(\boldsymbol{y}^\star)$ would be accepted by all players, whereas $\boldsymbol{y}^\star$ is a VE if no feasible deviation in the tangent cone $T_{\mathcal{X}\cap\mathcal{Z}}(\boldsymbol{y}^\star)$ would be accepted by all players. A VE has therefore greater social stability compared to a GNE. The next example presented in Cavazzuti et al. (2002) illustrates clearly this difference.

Example 1.1 (Post office location, Cavazzuti et al. (2002)). *Imagine a two-player game where each player $\nu \in \{1, 2\}$ is the mayor of a town ν located at $\mathbf{P}_\nu$, with $\mathbf{P}_1 = (1, 0)$ and $\mathbf{P}_2 = (0, 1)$. They must find an agreement on where to build a new post office that will be shared by both towns due to budget restrictions. Each mayor selects the coordinate $y_\nu \in \mathbb{R}$ of the post office location and wish to minimise the distance from their town to the post office,* i.e. $\theta_\nu(\boldsymbol{y}) := 0.5\,\|\boldsymbol{y} - \mathbf{P}_\nu\|_2^2$. *Moreover, the players are subject to the following constraints:*

- *Private constraint:* $\mathcal{X}_\nu = \{y_\nu \leq 0\}$.
- *Shared constraint:* $\mathcal{Z} = \left\{\boldsymbol{y} \in \mathbb{R}^2 \,\middle|\, y_1 + y_2 \leq -1\right\}$.

In such a situation, we have only one variational equilibrium but an infinite number of GNEs (see Figure 1.6).

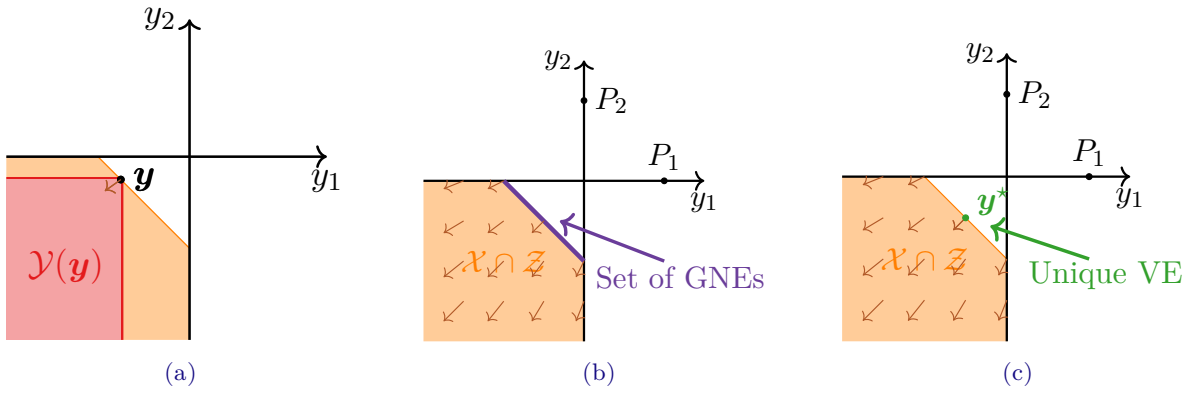


Figure 1.6: Comparison between GNE and VE

Remark 1.1. *If $\mathcal{Y}$ is orthogonal,* i.e. *in the absence of shared constraints, we have that*

$$I_{\mathcal{X}\cap\mathcal{Z}}(\boldsymbol{y}^\star) = T_{\mathcal{X}\cap\mathcal{Z}}(\boldsymbol{y}^\star)\,. \tag{1.49}$$

In this case the notions of NE and VE are equivalent.

1.3.3 Existence of a variational equilibrium

In the case of a jointly convex game, as shown in the next proposition, the existence of a variational equilibrium can be certified without any assumption on the boundedness of the joint feasible set $\mathcal{X} \cap \mathcal{Z}$ on which the existence result of Rosen (1965) relied on.

PROPOSITION 1.5 (FACCHINEI AND PANG (2009, THEOREM 12.3)). *If Assumption 1.1 holds, and if there exists a bounded open set $\mathcal{W}$ and a vector $\boldsymbol{y}^{ref} \in \mathcal{W} \cap \mathcal{X} \cap \mathcal{Z}$ such that $L_< \cap \boldsymbol{bd}(\mathcal{W}) = \emptyset$, where $\boldsymbol{bd}(\mathcal{W})$ is the boundary of $\mathcal{W}$ and*

$$L_< := \left\{ \boldsymbol{y} \in \mathcal{X} \cap \mathcal{Z} \,\middle|\, \exists \boldsymbol{u} \in \boldsymbol{\partial\Theta}(\boldsymbol{y}) \text{ such that } \left\langle \boldsymbol{u}\,,\, \boldsymbol{y} - \boldsymbol{y}^{ref} \right\rangle < 0 \right\}, \tag{1.50}$$

then the Nash game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$ has a variational equilibrium in $\mathcal{W}$.

We observe that if $\mathcal{X} \cap \mathcal{Z}$ is a bounded set, then any bounded set $\mathcal{W}$ that strictly contains $\mathcal{X} \cap \mathcal{Z}$ satisfies the assumptions of Proposition 1.5, where $\boldsymbol{y}^{ref}$ can be any point in $\mathcal{X} \cap \mathcal{Z}$. Therefore, we get back the existence result of Rosen (1965, Theorem 1) that states that if $\mathcal{X} \cap \mathcal{Z}$ is bounded, then there exists a variational equilibrium.

1.3.4 Uniqueness of a variational equilibrium

To prove the uniqueness of the variational equilibrium of the game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$, if one exists, Rosen (1965) introduced the property of *diagonal strict concavity*. This property is equivalent as saying that the pseudo-subdifferential set-valued map $\boldsymbol{\partial\Theta} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is *strictly monotone* on $\mathcal{X} \cap \mathcal{Z}$. The notion of monotonicity is central in the next chapters and is is defined as follows (Aubin and Cellina, 1984, Definition 3.1.1.):

DEFINITION 1.5 (MONOTICITY). *Let $\mathcal{X}$ be a Hilbert space, and let $\boldsymbol{\Psi}$ be a set-valued map from $\mathcal{X}$ into $\mathcal{X}$.*

- *$\boldsymbol{\Psi}$ is called* monotone *if and only if*

$$\forall \boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathbf{dom}(\boldsymbol{\Psi}), \forall \boldsymbol{v}_i \in \boldsymbol{\Psi}(\boldsymbol{x}_i), i \in \{1,2\} : \langle \boldsymbol{v}_1 - \boldsymbol{v}_2\,,\, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle \geq 0. \tag{1.51}$$

- *$\boldsymbol{\Psi}$ is called* strictly monotone *if and only if*

$$\forall \boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathbf{dom}(\boldsymbol{\Psi}), \forall \boldsymbol{v}_i \in \boldsymbol{\Psi}(\boldsymbol{x}_i), i \in \{1,2\} : \langle \boldsymbol{v}_1 - \boldsymbol{v}_2\,,\, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle > 0. \tag{1.52}$$

- *$\boldsymbol{\Psi}$ is called* uniformly monotone *with modulus ψ, where $\psi : \mathbb{R}_+ \to [0, +\infty]$ is an increasing function only vanishing at 0, if and only if*

$$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbf{dom}(\boldsymbol{\Psi}), \forall \boldsymbol{v}_i \in \boldsymbol{\Psi}(\boldsymbol{x}_i), i \in \{1,2\} : \langle \boldsymbol{v}_1 - \boldsymbol{v}_2\,,\, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle \geq \psi(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|). \tag{1.53}$$

- *$\boldsymbol{\Psi}$ is called* strongly monotone *if and only if there exists a constant $\mu > 0$ such that*

$$\forall \boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathbf{dom}(\boldsymbol{\Psi}), \forall \boldsymbol{v}_i \in \boldsymbol{\Psi}(\boldsymbol{x}_i), i \in \{1,2\} : \langle \boldsymbol{v}_1 - \boldsymbol{v}_2\,,\, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle \geq \mu \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2. \tag{1.54}$$

It is clear from these definitions that strong monotonicity implies uniform monotonicity, which itself implies (strict) monotonicity.

PROPOSITION 1.6. *If Assumption 1.1 holds, and if $\boldsymbol{\partial\Theta}$ is strictly monotone on $\mathcal{X} \cap \mathcal{Z}$. Then the game $\mathcal{G}(\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\Theta})$ has at most one variational equilibrium.*

PROOF. We use a similar proving strategy as Facchinei and Pang (2003a, Theorem 2.3.3). Let assume the existence of two distinct variational equilibria $\boldsymbol{y}^{\star,1}$ and $\boldsymbol{y}^{\star,2}$, which are solutions of GVI($\mathcal{X} \cap \mathcal{Z}, \boldsymbol{\partial\Theta}$). Then with (1.41), for $i \in \{1, 2\}$, there exist a $\boldsymbol{u}^{\star,i} \in \boldsymbol{\partial\Theta}(\boldsymbol{y}^{\star,i})$ such that

$$\forall \boldsymbol{y} \in \mathcal{X} \cap \mathcal{Z} : \left\langle \boldsymbol{u}^{\star,i}, \boldsymbol{y} - \boldsymbol{y}^{\star,i} \right\rangle \geq 0, \tag{1.55}$$

which implies $\langle \boldsymbol{u}^{\star,1}, \boldsymbol{y}^{\star,2} - \boldsymbol{y}^{\star,1} \rangle \geq 0$ and $\langle \boldsymbol{u}^{\star,2}, \boldsymbol{y}^{\star,1} - \boldsymbol{y}^{\star,2} \rangle \geq 0$. By combining these inequalities, we obtain $\langle \boldsymbol{u}^{\star,1} - \boldsymbol{u}^{\star,2}, \boldsymbol{y}^{\star,1} - \boldsymbol{y}^{\star,2} \rangle \leq 0$, which contradicts the strict monotonicity of $\boldsymbol{\partial\Theta}$ and completes the proof. □

COROLLARY 1.1. *If Assumption 1.1 holds and if the set $\mathcal{Y}$ is orthogonal (*i.e. *$\mathcal{Z} = \mathbb{R}^d$), then the strict monotonicity of $\boldsymbol{\partial\Theta}$ on $\mathcal{X} \cap \mathcal{Z}$ implies the uniqueness of the Nash equilibrium if it exists.*

PROOF. The proof follows from Corollary 1.1 and Remark 1.1. □

If the set $\mathcal{Y}$ is orthogonal (*i.e.* $\mathcal{Z} = \mathbb{R}^d$), then a weaker condition than the strict monotonicity of $\boldsymbol{\partial\Theta}$ is sufficient for the uniqueness of the NE, namely the P-property (see Facchinei and Pang (2003a)), which is equivalent as proving the monotonicity of $\boldsymbol{\partial\Theta}_{\boldsymbol{r}}$ for a $\boldsymbol{r} \in \mathbb{R}^n$ such that there is a unique index $\nu \in [\![1, n]\!]$ such that $r_\nu \to \infty$.

We conclude this chapter by stating the following proposition which provides a sufficient condition for $\boldsymbol{\partial\Theta}$ being monotone when every payoff function θ_ν, for $\nu \in [\![1, n]\!]$, is twice differentiable.

PROPOSITION 1.7 (ROSEN (1965, THEOREM 6)). *If we denote by $\mathbf{J}(\boldsymbol{y})$ the Jacobian of $\boldsymbol{\partial\Theta}$ at the point $\boldsymbol{y}$. Then, if the matrix $\boldsymbol{\Delta}(\boldsymbol{y}) \coloneqq \mathbf{J}(\boldsymbol{y}) + (\mathbf{J}(\boldsymbol{y}))^{\mathrm{T}}$ is positive definite (*resp. *positive semidefinite) for all $\boldsymbol{y} \in \mathcal{Y}$, then $\boldsymbol{\partial\Theta}$ is strictly monotone (*resp. *monotone) on $\mathcal{Y}$.*

Chapter 2

Optimal trade execution with uncertain volume target

The first question that arises when considering the uncertainty of end-user demand for electricity is how risk-averse energy wholesalers should optimally negotiate future contracts to meet this uncertain demand at the time of delivery. Electricity suppliers face a similar problem when they want to pre-sell the electricity they produce while being subject to generator outages and intermittent renewable energy production. In both situations, the risk comes from uncertain price dynamics on the one hand and uncertain demand/production of electricity on the other.

To answer to this question, this chapter introduces the uncertainty related to the volume target in the optimal execution problem introduced in the background material (Section 1.1). Most of the content of this chapter has been presented in the publication Vaes and Hauser (2022).

2.1 Introduction

In the classical formulation of the execution problem (see Section 1.1), the total volume of shares to be traded is deterministic and given as a parameter. However, in many situations, a trader only disposes of an estimate of this volume, which varies throughout the course of execution. There is therefore the need for a model that combines the risk associated with both the price dynamics and the uncertainty of the volume target. For example, the manager of an open ended investment fund may issue or redeem shares to avoid the asset-liability mismatch that could ensue if shares were only traded in the secondary market. Orders are usually collected and aggregated during the trading day, and the issue or redemption of the net share order creates the need to acquire or dispose of a certain volume of the fund's investment assets, typically over the next two trading days. Such a model would allow the fund manager to front load some of these trades and execute orders closer to the time when they are first placed, before all share orders are netted out at the end of the day, thus greatly increasing the market liquidity for the fund's investors. Another example is the clearing of power futures markets, which must result

in an outstanding volume close to the realised demand in every given delivery period, as the spot market has limited liquidity due to the physical constraints of the power plants used for generation. We focus in this chapter on the modelling of the execution problem in the situation where the trader is subject to uncertainty related to the price dynamics on the one hand, and related to the final volume to be traded by the end of the execution period on the other hand. For the latter, we assume that forecasts of increasing accuracy become available, so that volume uncertainty decreases over time, while price uncertainty increases.

To our knowledge, no previous model considers both of these two sources of uncertainty. Related articles are Cheng et al. (2017); Bulthuis et al. (2017); Cheng et al. (2019), where the optimal strategy with uncertain order fills is investigated, *i.e.* the risk for an order to be filled either incompletely or in excess (the latter being considered mainly for mathematical than practical reasons). In significant contrast to our situation, in their setting, the magnitude of the uncertainty of order fills is assumed to be either constant or proportional to the order size. In our case, the uncertainty is independent of the trader's decisions: the updates in the volume forecast depend exclusively on extraneous variables. Our model assumes consequently that both sources of uncertainty, *i.e.* the price dynamics and the forecast updates of the volume target, are inherent to the market and independent of the trader's trading strategy. Moreover, the model presented in this chapter allows for more flexibility compared to the approaches adopted by Cheng et al. (2017); Bulthuis et al. (2017); Cheng et al. (2019) as it does not rely on the assumption that the trading periods are homogeneous: for instance, it allows for the asset's liquidity to vary over the course of the execution period. Finally, unlike the alternative approaches, the strategies obtained with our model are static and their computation avoids relying on a computationally intensive dynamic programming approach.

This chapter is structured as follows: In Section 2.2, we propose a model that incorporates the volume uncertainty in the CVaR_α equivalent formulation of the risk-return trade-off execution model of Almgren and Chriss (2001). We establish the relationship between the viability of the market and the uniqueness of a convex set of optimal trading strategies. We conclude the section by adapting models from the literature to our problem and compare their relative advantages. Section 2.3 provides numerical evidence that our model achieves significantly better mean-CVaR_α trade-offs than when volume uncertainty is neglected, and it compares our approach against the current alternatives. We then illustrate how our model can be applied to power trading in Great Britain, where power wholesalers exchange power future contracts to hedge against the risk associated with the volatility of the power demand. Finally, we close this chapter on Sections 2.4 and 2.5, which discuss the results, propose avenues for further research on this subject, and conclude the chapter.

2.2 Model

We consider the model introduced in Section 1.1, where we assumed that the trading window spans the time interval $[\, t_s, t_f \,]$ with $t_s < t_f$, and that it is subdivided into m trading periods. We refer to t_f as the delivery time. The acquisition time is denoted by T and is equal to $t_f - t_s$.

Let S_0 be the initial price of the security, and let $\{S_k, k \in [\![0, m]\!]\}$ be the stochastic process that represents the fair price dynamics. The price process is assumed to be a random walk where the random step or *price shift* between two successive decision times t_{k-1} and t_k is modelled by a random variable ξ_k, *i.e.* $\forall k \in [\![1, m]\!] : S_k = S_{k-1} + \xi_k$. Furthermore, we assume that the market has limited liquidity and that trades induce linear temporary and permanent market impacts on the price dynamics. The price dynamics is then formulated as follows (see Section 1.1.2 for more details):

$$S_k = S_{k-1} + \xi_k + \gamma_k q_k, \tag{2.1a}$$

$$\widetilde{S}_k = S_{k-1} + \epsilon_k \operatorname{sgn}(q_k) + \frac{\eta_k}{\tau_k} q_k, \tag{2.1b}$$

where S_k is the fair price at decision time t_k and $\widetilde{S}_k$ is the effective security price for the trades executed during the trading period τ_k.

Unlike the classical optimal execution problem, we assume that the total volume to be traded by time t_f only becomes fully deterministically known during the execution period. As a consequence, Optimisation Problem (1.7) must be adapted in order to ensure that the trader acquires the right number of shares. From now on, we denote the volume target by a capital letter, *i.e.* D_{t_f}, to emphasise the fact that it is a random variable unlike in the mean-variance framework formulated in Equation (1.7).

In this work, we consider that the uncertainty associated with the total volume to be traded is defined as follows: for a given delivery time t_f, let D_k be the forecast at time t_k of the total volume to be traded by time $t_m = t_f$,

$$D_k \coloneqq \mathbb{E}\Big[\, D_{t_f} \;\Big|\; \mathcal{F}_k \Big]\,, \tag{2.2}$$

where $(\mathcal{F}_k)_{k=1}^m$ is the filtration of sigma algebras that represent the information available at time t_k. We assume that two successive forecasts differ by a continuous random variable δ_k for $k < m$. At decision time t_{m-1}, the volume target is assumed to be perfectly known. Hence, the forecasts of D_{t_f} evolve as follows over course of the entire execution period:

$$\begin{cases} D_k = D_{k-1} + \delta_k, & \text{if } k \in [\![1, m-1]\!] \\ D_m = D_{m-1}, \end{cases} \tag{2.3}$$

where $D_m = D_{t_f}$, the total volume to be traded by time t_f.

Obviously, in the presence of volume uncertainty, Constraint (1.7b) of Optimisation Problem (1.7) cannot be enforced since only a rough estimate of the total volume to be traded is known at time t_0, *i.e.* D_0. Nonetheless, the model of Almgren and Chriss (2001) may be useful

as a planning tool when deployed in conjunction with a systematic recourse whenever an update on the volume target becomes available. A trader would then, at each trading period, recompute and update their strategy based on the newest available volume forecast. The main flaw in this approach is that a trader is unable to define a static strategy at time t_0 that ensures the satisfaction of Constraint (1.7b). Indeed, the volumes to be traded in each trading period must constantly be updated. To get around this issue, we subsequently propose a new way of defining a trader's strategy. Furthermore, to avoid an explosion in the algorithmic complexity of numerical evaluations of the model, we want to avoid approaches based on dynamic programming and focus on models that account for the impact of recourse actions via the incorporation of a risk term.

First, unlike Almgren and Chriss (2001) who define a trading strategy as the volumes to be traded in each trading period, we define a trading strategy as the proportion y_k of the total volume D_{t_f} to be acquired over the course of each trading period τ_k. Naturally, the following constraint must hold to enforce the distribution of the entire volume target over all the trading periods:

$$\sum_{k=1}^{m} y_k = 1. \tag{2.4}$$

As a consequence, in the ideal situation where D_{t_f} is perfectly predictable, a trader would acquire $y_k D_{t_f}$ units of the security during trading period τ_k. Obviously, this definition must be adapted in the situation where D_{t_f} is uncertain as a trader should take recourse in order to satisfy Constraint (1.7b). As a starting point, we assume that a trader respects their initial strategy by trading during each trading period τ_k the planned proportion y_k of D_{k-1}, *i.e.* the best estimate of the volume target D_{t_f} available. This solves only partially the problem engendered by the forecast updates, since past decisions cannot be modified *a posteriori*, which means that $\sum_{k=1}^{m} y_k D_{k-1} \neq D_{t_f}$. Indeed, if we denote by δ_i, $i < m$, the *forecast update* reported at the decision time t_i, then the *forecast error* related to δ_i due to the past decisions amounts to $\delta_i^{\varepsilon} \coloneqq \delta_i \sum_{r=1}^{i} y_r$. This corresponds to the additional volume, which may be positive or negative, that needs to be traded over the remaining trading periods in order to match with D_{t_f}. We predicate that when an update on the volume target occurs, the trader adjusts their strategy by redistributing the forecast error over the remaining trading periods according to a fixed distribution independent of the decision variables. Considering these corrections upfront allows to partially account for future recourse actions without having to compute a costly dynamic programming simulation. In our approach, we redistribute this additional volume over the future trading periods according to fixed proportions. We treat these fixed proportions as model parameters and leave the learning of optimal parameter values to future work. We denote by $\beta_{i,k}$ the proportion of the forecast error δ_i^{ε} to correct for at trading period τ_k. Naturally, the following conditions must hold to ensure that the forecast error δ_i^{ε} is rectified in a non-anticipative manner: $\beta_{i,k} = 0$ if $k \leq i$ and $\sum_{k=1}^{m} \beta_{i,k} = 1$ for all $i \in [\![1, m-1]\!]$. We will refer to $\boldsymbol{\beta} \coloneqq (\beta_{i,k}) \in \mathbb{R}^{m-1 \times m}$ as the redistribution matrix.

In the case where the volume target is uncertain, we define the cost of a trading strategy

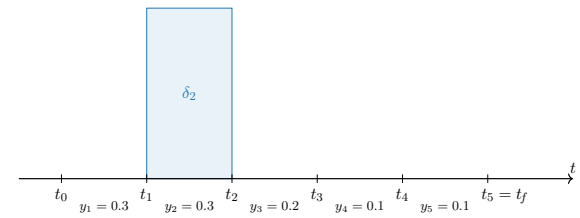
as the difference between *(i)* the trading cost incurred at the end of the execution period by following the initial trading strategy under the consideration of the predetermined recourse rule parametrised by the redistribution matrix $\boldsymbol{\beta}$ with *(ii)* the trading cost ideally obtained in an infinitely liquid market where the entire position D_{t_f} is traded at the start of the execution period. With these considerations, the trading costs associated to a strategy $\boldsymbol{y} = [y_1, \dots, y_m]^{\mathrm{T}}$ is the following random variable:

$$\begin{aligned}\mathcal{C}(\boldsymbol{y}) &:= \sum_{k=1}^{m} Q_k(\boldsymbol{y})\, \widetilde{S}_k - D_{t_f} S_0 \\ &= \sum_{k=1}^{m} \left(\xi_k + \gamma_k Q_k(\boldsymbol{y})\right)\left(D_0 + \sum_{i=1}^{m} \delta_i - \sum_{i=1}^{k} Q_i(\boldsymbol{y})\right) + \sum_{k=1}^{m} \epsilon \left|Q_k(\boldsymbol{y})\right| + \frac{\eta_k}{\tau_k}\left(Q_k(\boldsymbol{y})\right)^2, \end{aligned} \tag{2.5}$$

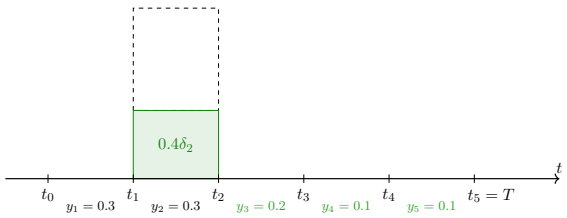
where $Q_k(\boldsymbol{y})$ is a random variable representing the volume to be traded during trading period τ_k:

$$Q_k(\boldsymbol{y}) = y_k D_{k-1} + \sum_{i=1}^{k-1} \beta_{i,k}\, \delta_i^{\varepsilon} = y_k D_0 + \sum_{i=1}^{k-1} \delta_i \left(y_k + \beta_{i,k} \sum_{r=1}^{i} y_r\right). \tag{2.6}$$

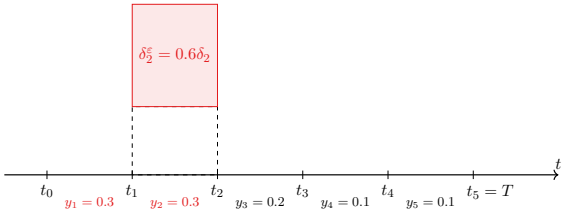
The number of positions $Q_k(\boldsymbol{y})$ to be traded at trading period τ_k is composed of *(i)* the proportion y_k of the best volume estimate D_{k-1}, and *(ii)* the position adjustments related to the previous volume forecast updates. An illustrative example is provided in Figure 2.1 when considering an even redistribution of the forecast error on the following trading periods. We now have that if $\forall k \in [\![1, m]\!]$, $Q_k(\boldsymbol{y})$ is defined as in (2.6), then $\sum_{k=1}^{m} Q_k(\boldsymbol{y}) = D_{t_f}$.



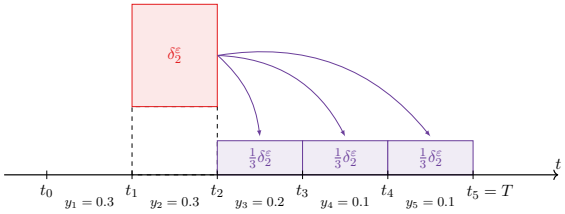
(a) δ_2: forecast update released at time t_2



(b) $0.4\delta_2$: part of the forecast update that should not yet have been traded according to $\boldsymbol{y}$



(c) $0.6\delta_2$: part of the forecast update that should have already been traded by time t_2 according to $\boldsymbol{y}$; it corresponds to the forecast error δ_2^{ε}



(d) Even distribution of the forecast error δ_2^{ε} on the following trading periods.

FIGURE 2.1: Illustration on a 5 periods trading window of the impact of $\boldsymbol{\beta}$ when it redistributes the forecast error evenly on the following trading periods for $\boldsymbol{y} = [0.3, 0.3, 0.2, 0.1, 0.1]^{\mathrm{T}}$.

From (2.6), we observe that, given any outcome ω of the sample space Ω and thus any realisation of the forecast updates, *i.e.* $\delta(\omega) := [\delta_1(\omega), \dots, \delta_{m-1}(\omega)]^{\mathrm{T}}$, the realisation of the volumes to be

traded $\mathbf{Q}(\boldsymbol{y}\,;\,\omega) \coloneqq [(Q_1(\boldsymbol{y}))(\omega)\,,\ldots,(Q_m(\boldsymbol{y}))(\omega)]^{\mathrm{T}}$ can be expressed as a linear combination of the decision variables $\boldsymbol{y}$. Formally this means that $\forall\omega \in \Omega$, we can find a (lower triangular) matrix $\mathbf{L}(\omega)$ such that

$$\mathbf{Q}(\boldsymbol{y}\,;\,\omega) = \mathbf{L}(\omega)\,\boldsymbol{y}. \tag{2.7}$$

Moreover, if $D_0 \neq 0$, $\mathbf{L}(\omega)$ is non-singular as all elements on the diagonal are different from zero, *i.e.* $\mathbf{L}_{kk}(\omega) = D_0$ for all $k \in [\![1, m]\!]$. In the rest of the chapter, $\mathcal{C}(\boldsymbol{y}\,;\,\omega)$ is used as a shorthand for $(\mathcal{C}(\boldsymbol{y}))(\omega)$, the trading cost of executing strategy $\boldsymbol{y}$ given the outcome ω. Finally, we consider that if a trader is guaranteed to not pay excessive prices in adverse times, the variance of their trading cost is not of foremost importance. We thus consider that a risk-averse trader is more interested in minimising their expected trading cost conditional on a quantile of worst case scenarios rather than minimising the variance of their trading cost over all scenarios. We therefore quantify the risk of a trading strategy with the $\mathbb{E}$-$\mathrm{CVaR}_{\alpha,\lambda}$ risk measure, which is defined as the trade-off parametrised by λ between the expectation and CVaR_α (see definition in Equation (1.25)). In the following, we denote λ by λ_{CVaR} in order to emphasise the difference with $\lambda_{\mathbb{V}\mathrm{ar}}$. Given a risk-aversion parameter $\lambda_{\mathrm{CVaR}} \in [0, 1]$, a trader thus tries to minimise the mean-CVaR_α trade-off of the total trading cost:

$$\underset{\boldsymbol{y}}{\text{minimise}} \quad \varphi_\alpha^{\lambda_{\mathrm{CVaR}}}(\boldsymbol{y}) \coloneqq (1 - \lambda_{\mathrm{CVaR}})\,\mathbb{E}[\mathcal{C}(\boldsymbol{y})] + \lambda_{\mathrm{CVaR}}\,\mathrm{CVaR}_\alpha[\mathcal{C}(\boldsymbol{y})] \tag{2.8a}$$

$$\text{subject to} \quad \sum_{k=1}^{m} y_k = 1. \tag{2.8b}$$

Note that in Optimisation Problem (2.8), a trader's recourse is estimated upfront rather than being simulated via dynamic programming to avoid the curse of dimensionality that plagues stochastic optimisation. Nevertheless, in contrast to the model of Almgren and Chriss (2001), Model (2.8) is well-defined since it adapts to the total volume variability and is thus guaranteed to satisfy the constraint that the total traded volume equals D_{t_f}.

2.2.1 Set of optimal strategies

Analogously than for Model (1.7) when the volume target is deterministic, we can prove that the objective function $\varphi_\alpha^{\lambda_{\mathrm{CVaR}}}$ of Optimisation Problem (2.8) is convex on its feasible set, *i.e.*

$$\mathcal{Y} \coloneqq \left\{ \boldsymbol{y} \in \mathbb{R}^m \,\middle|\, \sum_{k=1}^{m} y_k = 1 \right\}, \tag{2.9}$$

if the market free of any price manipulation. The convexity of the objective function $\varphi_\alpha^{\lambda_{\mathrm{CVaR}}}$ on $\mathcal{Y}$ for any value of $\lambda_{\mathrm{CVaR}} \in [0, 1]$ is equivalent to the simultaneous convexity of both functions $f_1 : \mathbb{R}^m \to \mathbb{R}; \boldsymbol{y} \mapsto \mathbb{E}[\mathcal{C}(\boldsymbol{y})]$ and $f_2 : \mathbb{R}^m \to \mathbb{R}; \boldsymbol{y} \mapsto \mathrm{CVaR}_\alpha[\mathcal{C}(\boldsymbol{y})]$ on $\mathcal{Y}$.

Lemma 2.1. *If the market is free of price manipulations, then function* $f_1 : \mathbb{R}^m \to \mathbb{R}; \boldsymbol{y} \mapsto \mathbb{E}[\mathcal{C}(\boldsymbol{y})]$ *is convex on* $\mathcal{Y}$.

Proof. By definition, we have $\mathbb{E}[\mathcal{C}(\boldsymbol{y})] = \int_\Omega \mathcal{C}(\boldsymbol{y}\,;\,\omega)\,\mathrm{d}\mathbb{P}(\omega)$. Hence, if for any outcome $\omega \in \Omega$, the function $\mathcal{C}(\bullet\,;\,\omega) : \mathbb{R}^m \to \mathbb{R}; \boldsymbol{y} \mapsto \mathcal{C}(\boldsymbol{y}\,;\,\omega)$ is (*resp.* strictly) convex in $\boldsymbol{y}$, then Lemma 2.1

holds since the expectation of $\mathcal{C}(\boldsymbol{y})$ can be seen as a non-negative weighted sum of the $\mathcal{C}(\boldsymbol{y}\,;\,\omega)$. Given (2.5), $\mathcal{C}(\boldsymbol{y}\,;\,\omega)$ is written as follows:

$$\mathcal{C}(\boldsymbol{y}\,;\,\omega) = \sum_{k=1}^{m}\left[(\xi_k(\omega) + \gamma_k Q_k(\boldsymbol{y}\,;\,\omega))\left(D_0 + \sum_{i=1}^{m}\delta_i(\omega) - \sum_{i=1}^{k} Q_i(\boldsymbol{y}\,;\,\omega)\right)\right] + \sum_{k=1}^{m}\left[\epsilon_k\,|Q_k(\boldsymbol{y}\,;\,\omega)| + \frac{\eta_k}{\tau_k}(Q_k(\boldsymbol{y}\,;\,\omega))^2\right]. \quad (2.10)$$

Since for any $\omega \in \Omega$, there exists a linear transformation between $\boldsymbol{y}$ and $\mathbf{Q}(\boldsymbol{y}\,;\,\omega)$ as shown in (2.7), proving the convexity of $\mathcal{C}(\bullet\,;\,\omega)$ in $\boldsymbol{y}$ is equivalent to proving the convexity of $\widetilde{\mathcal{C}}(\bullet\,;\,\omega)$ in $\mathbf{Q}(\boldsymbol{y}\,;\,\omega)$ (Boyd and Vandenberghe, 2004), where $\widetilde{\mathcal{C}}(\bullet\,;\,\omega)$ is defined as follows

$$\widetilde{\mathcal{C}}(\mathbf{Q}(\boldsymbol{y}\,;\,\omega)\,;\,\omega) \coloneqq \sum_{k=1}^{m}\xi_k(\omega)\left(D_{t_f}(\omega) - \sum_{i=1}^{k} Q_i(\boldsymbol{y}\,;\,\omega)\right) + \sum_{k=1}^{m}\gamma_k Q_k(\boldsymbol{y}\,;\,\omega)\,D_{t_f}(\omega) + \sum_{k=1}^{m}\epsilon_k\,|Q_k(\boldsymbol{y}\,;\,\omega)| - \sum_{k=1}^{m}\gamma_k Q_k(\boldsymbol{y}\,;\,\omega)\left(\sum_{i=1}^{k} Q_i(\boldsymbol{y}\,;\,\omega)\right) + \sum_{k=1}^{m}\frac{\eta_k}{\tau_k}(Q_k(\boldsymbol{y}\,;\,\omega))^2, \quad (2.11)$$

where $D_{t_f}(\omega) \coloneqq D_0 + \sum_{k=1}^{m}\delta_k(\omega)$. It is straightforward that the first three terms of (2.11) are convex in $\mathbf{Q}(\boldsymbol{y}\,;\,\omega)$, the last two can be combined in a similar manner as in the proof of Lemma 1.1:

$$-\sum_{k=1}^{m}\gamma_k Q_k(\boldsymbol{y}\,;\,\omega)\left(\sum_{i=1}^{k} Q_i(\boldsymbol{y}\,;\,\omega)\right) + \sum_{k=1}^{m}\frac{\eta_k}{\tau_k}(Q_k(\boldsymbol{y}\,;\,\omega))^2 = \frac{1}{2}\mathbf{Q}_{[1:m-1]}^{T}\,\mathbf{M}\,\mathbf{Q}_{[1:m-1]} - \gamma_m\left(D_{t_f}(\omega) - \sum_{k=1}^{m-1} Q_k(\boldsymbol{y}\,;\,\omega)\right)D_{t_f}(\omega) + \frac{\eta_m}{\tau_m}D_{t_f}(\omega)\left(D_{t_f}(\omega) - 2\sum_{k=1}^{m-1} Q_k(\boldsymbol{y}\,;\,\omega)\right), \quad (2.12)$$

where $\mathbf{Q}_{[1:m-1]} \coloneqq [Q_1(\boldsymbol{y}\,;\,\omega), \ldots, Q_{m-1}(\boldsymbol{y}\,;\,\omega)]^{\mathrm{T}}$ and $\mathbf{M}$ is defined as in Lemma 1.1. Since by assumption the market is free of price manipulations, we have with Lemma 1.1 that $\mathbf{M}$ is positive semi-definite, which in turn implies the convexity of the quadratic part of the trading costs. This completes the proof. □

Lemma 2.2. *If the market is free of price manipulations, then function* $f_2 : \mathbb{R}^m \to \mathbb{R}; \boldsymbol{y} \mapsto \mathrm{CVaR}_\alpha[\mathcal{C}(\boldsymbol{y})]$ *is convex on* $\mathcal{Y}$.

Proof. We have mentioned in Section 1.2 that any real valued coherent risk measure ϱ can be expressed in its dual representation, *i.e.*

$$\varrho[\mathcal{C}(\boldsymbol{y})] = \sup_{\mu\in\mathcal{D}} \mathbb{E}^{\mu}[\mathcal{C}(\boldsymbol{y})]\,, \quad (2.13)$$

where $\mathcal{D}$ is a convex subset of probability measures on Ω (Shapiro et al., 2009, Theorems 6.4 and 6.6). As a consequence, as CVaR_α is a coherent risk measure (Rockafellar and Uryasev, 2002, Corollary 12), the desired result is shown by combining the (*resp.* strict) convexity of $\mathcal{C}(\bullet\,;\,\omega)$ for any ω with the fact that a function defined as the pointwise supremum of (*resp.* strictly) convex functions is (*resp.* strictly) convex (Boyd and Vandenberghe, 2004). □

Theorem 2.1. *If the market is free of price manipulations, then the objective function of Optimisation Problem* (2.8) *is convex on its feasible set* $\mathcal{Y}$.

PROOF. The proof is straightforward based on Lemmas 2.1 and 2.2. □

Similarly to Model (1.7), the convexity and the non-emptiness of the feasible set $\mathcal{Y}$ and Theorem 2.1 imply, under the conditions that $\lambda_{\text{CVaR}} \in [0,1]$ and that $\mathbf{M} \succeq 0$, the existence and uniqueness of a convex set $\mathcal{Y}^\star \subseteq \mathcal{Y}$ of optimal execution strategies for the model defined in (2.8). If $\mathbf{M} \succ 0$, the set of optimal strategies is a singleton, *i.e.* $\mathcal{Y}^\star \coloneqq \{\boldsymbol{y}^\star\}$.

In other words, the necessary condition $\mathbf{M} \succeq 0$ for the absence of price manipulation, and thus for having a viable market where no trader has interest in impacting the price dynamics with artificial trades, is tantamount to the convexity of Optimisation Problem (2.8), and thus to a convex set of optimal strategies for a trader considering both price and volume uncertainty.

2.2.2 Comparison with other approaches

Mean-variance with recourse. In the case where the volume target is uncertain, the model of Almgren and Chriss (2001) can be deployed in conjunction with a systematic recourse whenever a forecast update on the volume target becomes available. We will show that such a strategy can be reproduced in our model with an appropriate choice of $\boldsymbol{y}$ and $\boldsymbol{\beta}$.

Similarly to Model (2.8), a trading strategy $\mathbf{q}$ in Model (1.7) that trades a fixed volume of d_{t_f} shares can be equivalently expressed as the proportions y_k of d_{t_f} to be traded during each trading period τ_k, $k \in [\![1, m]\!]$.

LEMMA 2.3. *If ξ_k, $k \in [\![1,m]\!]$, are draws from independent random variables each with zero mean,* i.e. $\mathbb{E}[\xi_k] = 0$, *if the permanent and temporary impact functions are given by* (1.3a)-(1.3b), *and if the temporary impact parameter ϵ_k is constant over all (*i.e. *independent of the) trading periods,* i.e. $\forall k \in [\![1,m]\!] : \epsilon_k = \epsilon$, *then the optimal strategy $\boldsymbol{y}^\star$ of Optimisation Problem* (1.7) *expressed in terms of the proportions of d_{t_f} is independent of d_{t_f}.*

PROOF. Based on (1.6), the expectation and variance of the trading cost are given by

$$\mathbb{E}[\mathcal{C}(\mathbf{q})] = \sum_{k=1}^{m} \gamma_k q_k \left(d_{t_f} - \sum_{i=1}^{k} q_i \right) + \sum_{k=1}^{m} \left(\epsilon \left| q_k \right| + \frac{\eta_k}{\tau_k} q_k^2 \right) \tag{2.14}$$

$$= \sum_{k=1}^{m} \gamma_k y_k \left(1 - \sum_{i=1}^{k} y_i \right) d_{t_f}^2 + \sum_{k=1}^{m} \left(\epsilon \left| y_k d_{t_f} \right| + \frac{\eta_k}{\tau_k} \left(y_k d_{t_f} \right)^2 \right), \tag{2.15}$$

$$\mathbb{V}\text{ar}[\mathcal{C}(\mathbf{q})] = \sum_{k=1}^{m} \mathbb{V}\text{ar}[\xi_k] \left(d_{t_f} - \sum_{i=1}^{k} q_i \right)^2 = \sum_{k=1}^{m} \mathbb{V}\text{ar}[\xi_k] \left(1 - \sum_{i=1}^{k} y_i \right)^2 d_{t_f}^2. \tag{2.16}$$

Moreover, if we assume w.l.o.g. that $d_{t_f} > 0$, then the optimal volumes $\mathbf{q}^\star$ of Optimisation Problem (1.7) are positive (Almgren and Chriss, 2001). Therefore, adding the non-negativeness constraint on the decision variables $\mathbf{q}$ to Optimisation Problem (1.7) does not alter its optimal solution. Hence, we can replace $\sum_{k=1}^{m} \epsilon \left| q_k \right|$ by $\sum_{k=1}^{m} \epsilon q_k = \epsilon d_{t_f}$. As a consequence, all the terms in the objective function of Optimisation Problem (1.7) that depend on the decision variables $\boldsymbol{y}$ are multiplied by $d_{t_f}^2$, which means that d_{t_f} has no impact on the optimal solution $\boldsymbol{y}^\star$. □

Lemma 2.3 implies that under price and volume uncertainty, the strategy of a trader that consists in deploying the model of Almgren and Chriss (2001) in conjunction with a systematic recourse

whenever an update on the volume target becomes available is time consistent. More precisely, given the optimal strategy $\boldsymbol{y}^{\star,t_0}$ of (1.7) computed at time t_0, if we recompute the optimal strategy $\boldsymbol{y}^{\star,t_{k-1}}$ of (1.7) at a subsequent decision time t_{k-1}, $k \in [\![2, m]\!]$, it would simply be the continuation of $\boldsymbol{y}^{\star,t_0}$ from time t_{k-1} to t_m, *i.e.*

$$\boldsymbol{y}^{\star,t_{k-1}} \coloneqq \left[y_k^{\star,t_{k-1}}, \dots, y_m^{\star,t_{k-1}}\right]^{\mathrm{T}} = \left(\frac{1}{\sum_{r=k}^{m} y_r^{\star,t_0}}\right) \boldsymbol{y}_{[k:m]}^{\star,t_0}, \tag{2.17}$$

where $\boldsymbol{y}_{[k:m]}^{\star,t_0} \coloneqq \left[y_k^{\star,t_0}, \dots, y_m^{\star,t_0}\right]^{\mathrm{T}}$. Note that the rescaled proportions $\boldsymbol{y}^{\star,t_{k-1}}$, $k \in [\![2, m]\!]$, are independent of past realisations (price moves and forecast updates) and can consequently be computed beforehand at the start of the execution period, *i.e.* at time t_0. If $\mathbf{Q}^{rec} \coloneqq [Q_1^{rec}, \dots, Q_m^{rec}]^{\mathrm{T}}$ denotes the traded volumes of a trader who updates their strategy by solving (1.7) whenever an update on the volume target occurs, then the volume Q_k^{rec} traded on trading period τ_k, $k \in [\![1, m]\!]$, is equal to the proportion $y_k^{\star,t_{k-1}}$ of the best volume target estimate D_{k-1} minus the volumes already traded, *i.e.* $Q_k^{rec} = y_k^{\star,t_{k-1}}\left(D_{k-1} - \sum_{i=1}^{k-1} Q_i^{rec}\right)$.

LEMMA 2.4. *If ξ_k, $k \in [\![1, m]\!]$, are draws from independent random variables each with zero mean,* i.e. $\mathbb{E}[\xi_k] = 0$, *if the permanent and temporary impact functions are given by* (1.3a)-(1.3b), *and if the temporary impact parameter ϵ_k is constant over all trading periods,* i.e. $\forall k \in [\![1, m]\!] : \epsilon_k = \epsilon$, *then there exist a strategy $\bar{\boldsymbol{y}}$ and a redistribution matrix $\bar{\boldsymbol{\beta}} \coloneqq (\beta_{i,k}) \in \mathbb{R}^{m-1 \times m}$ such that the trading volumes given by Equation* (2.6) *are equal to $\mathbf{Q}^{rec}$.*

PROOF. Let us define $\bar{\boldsymbol{y}}$ as $\boldsymbol{y}^{\star,t_0} > 0$, the optimal solution of Optimisation Problem (1.7) at time t_0, and $\bar{\boldsymbol{\beta}} = \left(\bar{\beta}_{i,k}\right) \in \mathbb{R}^{m-1 \times m}$ as the redistribution matrix implied by $\boldsymbol{y}^{\star,t_0}$, *i.e.* the matrix whose components are defined as follows

$$\bar{\beta}_{i,k}\left(\boldsymbol{y}^{\star,t_0}\right) \coloneqq \begin{cases} 0 & \text{if } k \leq i, \\ \frac{y_k^{\star,t_0}}{\sum_{r=i+1}^{m} y_r^{\star,t_0}} & \text{otherwise.} \end{cases} \tag{2.18}$$

Then, $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{\beta}}$ are such that the trading volumes given by Equation (2.6) are equal to $\mathbf{Q}^{rec}$. Indeed, using the fact that $\sum_{k=1}^{m} y_k^{\star,t_0} = 1$, we have that the volumes $\mathbf{Q}^{rec}$ are given by

$$Q_1^{rec} = y_1^{\star,t_0} D_0, \tag{2.19}$$

$$\begin{aligned} Q_2^{rec} &= \frac{y_2^{\star,t_0}}{\sum_{r=2}^{m} y_r^{\star,t_0}}\left(D_0 + \delta_1 - y_1^{\star,t_0} D_0\right) = \frac{y_2^{\star,t_0}}{1 - y_1^{\star,t_0}}\left(\left(1 - y_1^{\star,t_0}\right) D_0 + \delta_1\right) \\ &= y_2^{\star,t_0} D_0 + \frac{y_2^{\star,t_0}}{1 - y_1^{\star,t_0}} \delta_1 \end{aligned} \tag{2.20}$$

$$\begin{aligned} Q_3^{rec} &= \frac{y_3^{\star,t_0}}{\sum_{r=3}^{m} y_r^{\star,t_0}}\left(D_0 + \delta_1 + \delta_2 - y_1^{\star,t_0} D_0 - y_2^{\star,t_0} D_0 - \frac{y_2^{\star,t_0}}{1 - y_1^{\star,t_0}} \delta_1\right) \\ &= y_3^{\star,t_0} D_0 + \frac{y_3^{\star,t_0}}{1 - y_1^{\star,t_0}} \delta_1 + \frac{y_3^{\star,t_0}}{1 - y_1^{\star,t_0} - y_2^{\star,t_0}} \delta_2 \end{aligned} \tag{2.21}$$

$$\vdots$$

$$Q_k^{rec} = y_k^{\star,t_0} D_0 + \sum_{k=1}^{k-1} \left(\frac{y_k^{\star,t_0}}{1 - \sum_{r=1}^{k} y_r^{\star,t_0}} \right) \delta_k. \tag{2.22}$$

Using the definitions of $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{\beta}}$, we obtain

$$Q_k^{rec} = \bar{y}_k D_0 + \sum_{i=1}^{k-1} \left(\frac{\left(1 - \sum_{r=1}^{i} \bar{y}_r\right) \bar{y}_k + \bar{y}_k \left(\sum_{r=1}^{i} \bar{y}_r\right)}{1 - \sum_{r=1}^{i} \bar{y}_r} \right) \delta_i \tag{2.23}$$

$$= \bar{y}_k D_0 + \sum_{i=1}^{k-1} \left(\bar{y}_k + \bar{\beta}_{i,k} \sum_{r=1}^{i} \bar{y}_r \right) \delta_i, \tag{2.24}$$

which is exactly the expression of the traded volumes in (2.8) for the strategy $\bar{\boldsymbol{y}}$ and the redistribution matrix $\bar{\boldsymbol{\beta}}$ (see Equation (2.6)). □

Lemma 2.4 implies that, if we were to optimise (2.8) on the set of strategies $\mathcal{Y}$ and the set of redistribution matrices

$$\mathcal{B} := \left\{ \boldsymbol{\beta} \in \mathbb{R}^{m-1 \times m} \,\middle|\, \beta_{i,k} = 0 \;\text{ if }\; k \leq i \quad \text{and} \quad \forall i \in [\![1, m-1]\!] : \sum_{k=1}^{m} \beta_{i,k} = 1 \right\}, \tag{2.25}$$

then there exists a strategy $(\boldsymbol{y}, \boldsymbol{\beta}) \in \mathcal{Y} \times \mathcal{B}$ such that the volumes traded by following this strategy are the same as the ones of a trader recomputing the optimal strategy in the mean-variance framework whenever an update on the volume target is released. Hence, a wisely chosen redistribution matrix $\boldsymbol{\beta}$ can mimic the action of taking recourse. A sensible choice for the redistribution matrix $\boldsymbol{\beta}$ would for instance be the matrix implied by the optimal risk-neutral strategy $\boldsymbol{y}^{\star,t_0,\lambda_{\mathbb{V}\mathrm{ar}}=0}$ in the mean-variance framework (Optimisation Problem (1.7)) at time t_0, *i.e.* the matrix defined as in (2.18) but with $\boldsymbol{y}^{\star,t_0,\lambda_{\mathbb{V}\mathrm{ar}}=0}$.

Exponential utility and mean-QV. The uncertainty of the volume target has, to the best of our knowledge, not yet been considered in the literature. Some papers deal with a conceptually similar problem to ours but rely on assumptions that are discarded in our model, allowing for greater generality as explained hereinafter. In Cheng et al. (2017, 2019), the authors incorporate in the optimal execution problem the uncertainty related to order fills, which is the risk for an order to be under- or over-executed. Bulthuis et al. (2017) extend this model by including, additionally to the market orders, limit orders with uncertain fill rates. The authors also add penalties to tailor the orders' sign and magnitude. For comparison, we consider the model of Cheng et al. (2017) as we make similar assumptions on the market; the model is here formulated as a liquidation problem, *i.e.* from a seller's perspective rather than the buyer's perspective adopted earlier. Their model is as follows. Let x_t denote the number of shares that the trader holds at a time $t \in [0, t_f]$ during the execution process; x_0 denotes the initial amount of shares. Similarly to Almgren and Chriss (2001), the fair price S_t is driven by the SDE

$$\mathrm{d}S_t = \gamma \,\mathrm{d}x_t + \mu \,\mathrm{d}t + \sigma \,\mathrm{d}W_t \Leftrightarrow S_t = S_0 - \gamma x_0 + \gamma x_t + \mu t + \sigma W_t \tag{2.26}$$

and the transaction price $\tilde{S}_t$ is given by

$$\tilde{S}_t = S_t - \eta v_t, \tag{2.27}$$

where W_t is a Brownian motion, v_t denotes the rate at which the trader scheduled to trade at the instant t, μ parametrises the price drift, and γx_t, $\gamma \geq 0$, and $-\eta v_t$, $\eta > 0$, respectively refer to the permanent and temporary impacts of the trades. To take into account the uncertainty regarding the order fills, Cheng et al. (2017) add to the dynamics of the position x_t a noise driven by another Brownian motion Z_t (potentially correlated with W_t), *i.e.*

$$\mathrm{d}x_t = -v_t\,\mathrm{d}t + m\left(v_t\right)\mathrm{d}Z_t, \tag{2.28}$$

where the diffusion $m\left(v_t\right)$ characterises the magnitude of the uncertainty of order fills. At each time $t \in [0, t_f]$, the profit and loss (P&L) $\Pi_t(x)$ generated by the strategy x until time t is

$$\Pi_t(x) = x_t(S_t - S_0) - \int_0^t \left(\tilde{S}_u - S_0\right)\mathrm{d}x_u. \tag{2.29}$$

The first term on the right-hand side of (2.29) represents the change in fair value of the positions that still have to be transacted (it is beneficial for the trader if $S_t > S_0$), whereas the second term is the cost incurred due to the price risk and price impact of the shares traded up to time t. Using Equations (2.26) and (2.27), the P&L can be reformulated as

$$\Pi_t(x) = x_t(\gamma(x_t - x_0) + \mu t + \sigma W_t) - \int_0^t \left(\gamma(x_u - x_0) + \mu u + \sigma W_u - \eta v_u\right)\mathrm{d}x_u. \tag{2.30}$$

To discourage any residual positions at the terminal time t_f, a penalty term $f : x_{t_f} \to -\beta x_{t_f}^2$, $\beta \geq \eta > \frac{\gamma}{2}$, is added to $\Pi_{t_f}(x)$, which gives the final P&L $\widetilde{\Pi}_{t_f}(x) \coloneqq \Pi_{t_f}(x) + f\left(x_{t_f}\right)$. In the case where the magnitude of the uncertainty of order fills is constant ($m\left(v_t\right) = m_0$) and where there is no drift ($\mu = 0$), Cheng et al. (2017) provide closed form expressions for the optimal trading rate $v_t^\star$ and liquidation trajectory $x_t^\star$ of a risk-averse trader that assesses risk in two distinct ways. On the one hand, by maximising the expected utility of the final P&L, *i.e.* $\sup_{v\in\mathcal{A}} \mathbb{E}\left[U\left(\widetilde{\Pi}_{t_f}(x)\right)\right]$ with U the exponential utility function: $U(x) = \frac{1-e^{-\theta_{\mathrm{EU}}x}}{\theta_{\mathrm{EU}}}$. On the other hand, by maximising the expected P&L penalised by its quadratic variation (Forsyth et al., 2012), *i.e.* $\sup_{v\in\mathcal{A}} \mathbb{E}\left[\widetilde{\Pi}_{t_f}(x) - \lambda_{\mathrm{QV}}\mathrm{QV}\left[\Pi_{t_f}(x)\right]\right]$. In both cases, $\mathcal{A}$ denotes the set of admissible controls. The risk-aversion parameter are respectively given by $\theta_{\mathrm{EU}} > 0$ and $\lambda_{\mathrm{QV}} > 0$; the larger θ_{EU} (*resp.* λ_{QV}), the more risk-averse the trader.

Based on the optimal trading rate $v_t^\star$ given by Cheng et al. (2017), it is straightforward to obtain the optimal trading rate of the execution problem formulated from the perspective of a buyer desiring to acquire an uncertain amount of shares D_{t_f}. Indeed, the fluctuations in the number of positions of the trader due to the uncertain order fills can be seen as fluctuations in the uncertain volume target. Hence, if the dynamics of the volume target forecast $D_t \coloneqq \mathbb{E}\left[D_{t_f} \,\middle|\, \mathcal{F}_t\right]$ is given by $\mathrm{d}D_t = m_0\,\mathrm{d}X_t$, where X_t is a Brownian motion and $(\mathcal{F}_t)_{t\in[0,t_f]}$ is the filtration of sigma algebras that represent the information available at time t, if w_t denotes the trading rate at the instant t, and if z_t denotes the number of shares left to be traded in order to match with the current forecast D_t on the volume target, then the dynamics of z_t is given by

$$\mathrm{d}z_t = -w_t\,\mathrm{d}t + m_0\,\mathrm{d}X_t, \tag{2.31}$$

where the initial amount of shares to be traded z_0 is equal to $D_0 \coloneqq \mathbb{E}\Big[D_{t_f} \,\Big|\, \mathcal{F}_0\Big]$. Note that, in contrast to v_t, a positive value for w_t corresponds to the acquisition and not the liquidation of shares. Consequently, the impact of the trades on the price dynamics and the transaction price coincides with that of Cheng et al. (2017) up to a change of sign, and similarly for the price drift, the rationale being that a positive price drift in a liquidation program has a similar beneficial effect in terms of costs as a negative price drift in an acquisition program. The fair price S_t dynamics, the transaction price $\tilde{S}_t$, and the P&L $\Pi_t(z)$ are then given by

$$\mathrm{d}S_t = -\gamma\,\mathrm{d}z_t - \mu\,\mathrm{d}t - \sigma\,\mathrm{d}W_t \Leftrightarrow S_t = S_0 + \gamma(z_0 - z_t) - \mu t - \sigma W_t, \tag{2.32}$$

$$\tilde{S}_t = S_t + \eta w_t, \tag{2.33}$$

$$\Pi_t(z) = z_t(S_0 - S_t) - \int_0^t \Big(S_0 - \tilde{S}_u\Big)\,\mathrm{d}z_u. \tag{2.34}$$

Similarly to (2.29), the first term on the right-hand side of (2.34) represents the change in fair value of the positions that still have to be acquired (it is beneficial for the trader if $S_t < S_0$), whereas the second term is the cost incurred due to the price risk and price impact of the shares traded up to time t. Using Equations (2.32) and (2.33), the P&L can be reformulated as

$$\Pi_t(z) = z_t(\gamma(z_t - z_0) + \mu t + \sigma W_t) - \int_0^t (\gamma(z_u - z_0) + \mu u + \sigma W_u - \eta w_u)\,\mathrm{d}z_u. \tag{2.35}$$

We observe that Equations (2.30) and (2.35) are similar. Therefore, if a risk-averse trader either maximises the expected utility of the final P&L, *i.e.* $\widetilde{\Pi}_{t_f}(z) = \Pi_{t_f}(z) + f\Big(z_{t_f}\Big)$, or the expected P&L penalised by its quadratic variation, then the optimal trading rate $w_t^\star$ in an acquisition program subjected to an uncertain volume target, will be the same as the optimal trading rate $v_t^\star$ in a liquidation program subjected to uncertain order fills. If additionally, we assume the absence of price drift and that any remaining positions is prohibited, *i.e.* $\mu = 0$ and $\beta \to \infty$, then the continuous execution problem formulated here above tackles essentially the same problem as Model (2.8) in the absence of fixed cost, *i.e.* $\forall k \in [\![1, m]\!] : \epsilon_k = 0$, when the price shifts ξ_k and forecast updates δ_k, $k \in [\![1, m]\!]$, follow respectively a zero mean normal distribution of variance $\tau_k\sigma^2$ and $\tau_k m_0^2$, *i.e.* $\xi_k \sim \mathcal{N}(0, \tau_k\sigma^2)$ and $\delta_k \sim \mathcal{N}(0, \tau_k m_0^2)$.

To conclude, there are multiple advantages of our approach. Firstly, we showed in Lemma 2.4 that the traded volumes resulting from the strategy of recomputing the optimal strategy in the mean-variance framework whenever a volume target update becomes available can be reproduced in our model with an appropriate updating rule $\boldsymbol{\beta}$. Secondly, our model is significantly more versatile compared to the model of Cheng et al. (2017). Indeed, to provide closed form expressions of the optimal trading rate, Cheng et al. (2017) rely on the assumptions that both the price dynamics and the uncertain order fills are modelled with Brownian motions, as well as on the assumption that the price volatility (σ), the volatility of the updates on the volume target (m_0), and the price impact parameters (γ and η) are constant over the entire execution period. All these considerations are not needed in our model, which allows us to consider a wider range of situations, such as the case where the liquidity profile is not constant over the course of the execution period or the situation where the forecast updates on the volume target are correlated.

In these situations, finding closed form expressions of the optimal trading rate for the model proposed by Cheng et al. (2017) would be significantly more challenging. Alternatively, the optimal rate could be obtained using a dynamic programming approach. However, this would result in a significantly higher computational cost compared to our approach. Finally, as we will show in Section 2.3, we will observe that, despite the fact that the CVaR_α risk measure is not time-consistent which may lead to sub-optimal strategies (Rudloff et al., 2014), the strategies obtained with our model have competitive performance compared to the optimal solution provided by the dynamical programming approach suggested in Cheng et al. (2017). This suggests that integrating the uncertainty related to the volume target in the estimate of the trading cost of a strategy (see Equation (2.5)) and taking partial recourse via a predetermined rule $\boldsymbol{\beta}$ lead to intuitive and competitive trading strategies while avoiding the computational cost of more convoluted approaches.

2.3 Numerical results

In this section, we analyse the influence of incorporating the uncertainty related to the volume target into the trade execution problem. We provide some numerical evidence that, when the uncertainty related to the volume target is considered, a trader should adapt their execution program by delaying their trades. To illustrate the results, we apply our model in two distinct test cases. In both cases, it is assumed that the market has a single stock with initial price $S_0 = 50$, and with median daily trading volume V of 5 million shares:

(a) **Constant bid-ask spread.** We assume that the market liquidity remains constant over all the trading periods; we assume a bid-ask spread b_k equal to 0.25% of the initial asset price S_0 for each trading period τ_k, *i.e.* $\forall k \in [\![1, m]\!] : b_k = 0.125$.

(b) **U-shaped intra-day bid-ask spread profile.** Previous papers, *e.g.* Chan et al. (1995), suggest that the traded volumes and bid-ask spreads of NYSE stocks often follow a U-shaped pattern over the trading day; the trading activity being concentrated at the market opening and closure. Consequently, the price-impact slopes, *i.e.* η_k, are larger in the first and last periods than in the middle periods (Huberman and Stanzl, 2005). To capture such behaviour, we assume the following profile for the bid-ask spread over the 5 trading periods: $\boldsymbol{b} := [b_1, \ldots, b_5]^{\mathrm{T}} = [0.15, 0.125, 0.125, 0.2, 0.25]^{\mathrm{T}}$, which correspond to 0.3%, 0.25%, 0.25%, 0.4% and 0.5% of the initial asset price S_0.

For both cases, we assume that the volatility σ of the asset is constant during the entire execution period and is equal to 0.95, and that the price shift ξ_k, $k \in [\![1, m]\!]$, follows a zero mean Gaussian distribution of variance $\tau_k \sigma^2$, *i.e.* $\xi_k \sim \mathcal{N}(0, \tau_k \sigma^2)$. For the temporary impact, the fixed cost factor is assumed to amount to one half of the bid-ask spread, *i.e.* $\forall k \in [\![1, m]\!] : \epsilon_k = 0.5 b_k$. Additionally, Almgren and Chriss (2001) suggest that a trader incurs a price impact equal to one bid-ask spread for every percent of the daily volume, *i.e.* $\forall k \in [\![1, m]\!] : \eta_k = b_k/(0.01V)$, which

we adopt in this work as well. As for the permanent price impact, Almgren and Chriss (2001) assume that the price effects become significant when 10% of the daily volume is traded. In this application we suggest that the effect is significant if it corresponds to five times the bid-ask spread of the trading period, whereas Almgren and Chriss (2001) consider a price shift equal to one bid-ask spread as significant. This increase of scale allows us to illustrate more clearly the contribution of volume uncertainty in the risk adopted by the trader, but it is also not completely unrealistic for illiquid assets. We therefore set $\gamma_k = (5b_k)/(0.1V)$ for all $k \in [\![1, m]\!]$. For smaller values of γ_k, the impact of volume uncertainty is slightly smaller but, due to the cumulative effect of sequential trades, even small savings in the trading cost make a large difference over time.

Finally, we assume for both cases that the initial volume target D_0 amounts to 0.5 million shares, and that for each trading period τ_k except the last one, the forecast update follows a zero mean Gaussian distribution of variance $\tau_k \nu^2$ with ν being equal to 10% of the initial volume target ($\nu := 0.1 D_0$), *i.e.* $\forall k \in [\![1, m-1]\!] : \delta_k \sim \mathcal{N}(0, \tau_k \nu^2)$. As suggested in Section 2.2.2, we assume that the residual volumes to be traded due to the forecast updates are redistributed over the future trading periods based on the matrix $\boldsymbol{\beta}$ implied by the risk-neutral optimal trading strategy $\boldsymbol{y}^{\star, t_0, \lambda_{\mathbb{V}\mathrm{ar}}=0}$ in the mean-variance framework (Optimisation Problem (1.7)) computed at time t_0, *i.e.* $\boldsymbol{\beta} = \left(\beta_{i,k}\left(\boldsymbol{y}^{\star, t_0, \lambda_{\mathbb{V}\mathrm{ar}}=0}\right)\right) \in \mathbb{R}^{m-1 \times m}$ with

$$\beta_{i,k}\left(\boldsymbol{y}^{\star, t_0, \lambda_{\mathbb{V}\mathrm{ar}}=0}\right) = \begin{cases} 0 & \text{if } k \leq i, \\ \dfrac{y_k^{\star, t_0, \lambda_{\mathbb{V}\mathrm{ar}}=0}}{\sum_{r=i+1}^{m} y_r^{\star, t_0, \lambda_{\mathbb{V}\mathrm{ar}}=0}} & \text{otherwise.} \end{cases} \tag{2.36}$$

In practice, the exact distributions of the price moves and the forecast updates are unknown and are estimated from historical data (*e.g.* see Section 2.3.4). In the following, we denote by $\hat{\cdot}$ any trader's estimation of the ground truth parameter. For example, $\tau_k \hat{\nu_k}^2$ is the trader estimated value of $\tau_k \nu_k^2$, the variance of the forecast update released after trading period τ_k. Table 2.1 summarises the market details of both cases.

2.3.1 Optimal strategies and impact of the liquidity profile

Given the market parameters of Table 2.1, the condition $\mathbf{M} \succ 0$ of Theorem 2.1 is verified, which means that for any risk-aversion parameter $\lambda_{\mathrm{CVaR}} \in [0, 1]$, the optimal trading strategy in the Mean-CVaR framework is uniquely defined. Figures 2.2a and 2.2b depict, for both liquidity profiles, the optimal strategies of Model (2.8) when only price uncertainty (PU) is considered, whereas Figures 2.2c and 2.2d depict, the optimal trading strategies when both sources of uncertainty are taken into account, *i.e.* price and volume uncertainty (PVU). The difference between the optimal strategies in these two situations is represented for both cases respectively in Figures 2.2e and 2.2f. This difference illustrates how the optimal strategies under price uncertainty only should be adjusted when a trader additionally takes into account the uncertainty related to the volume target.

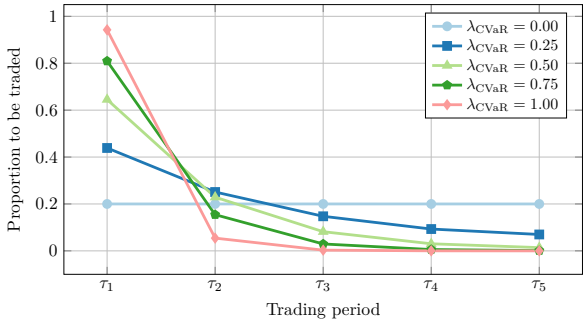
TABLE 2.1: Market parameters

Acquisition time:	$T = t_f - t_s = 5$ [day]
Number of trading periods:	$m = 5$
Trading period length:	$\forall k \in [\![1, m]\!] : \tau_k = T/m = 1$ [day]
Initial asset price:	$S_0 = 50$ [\$/share]
Asset's volatility:	$\sigma = 0.95$ [(\$/share)/day$^{1/2}$]
Price shifts distribution:	$\forall k \in [\![1, m]\!] : \xi_k \sim \mathcal{N}(0, \tau_k \sigma^2)$
Daily volume:	$V = 5 \cdot 10^6$ [share]
Initial volume target forecast:	$D_0 = 0.5 \cdot 10^6$ [share]
Volume forecast updates distribution:	$\forall k \in [\![1, m-1]\!] : \delta_k \sim \mathcal{N}(0, \tau_k \nu^2)$ with $\nu = 0.1 D_0$
Bid-ask spread:	Case (a): $\forall k \in [\![1, m]\!] : b_k = 0.125$ [\$/share]
	Case (b): $\boldsymbol{b} = [0.15, 0.125, 0.125, 0.2, 0.25]$ [\$/share]
Fixed cost:	$\forall k \in [\![1, m]\!] : \epsilon_k = 0.5 b_k$ [\$/share]
Impact at 1% of market:	$\forall k \in [\![1, m]\!] : \eta_k = b_k/(0.01V)$ [(\$/share)/(share/day)]
Permanent impact parameter:	$\forall k \in [\![1, m]\!] : \gamma_k = (5 b_k)/(0.1V)$ [\$/share^2]
CVaR$_\alpha$ parameter:	$\alpha = 0.1$

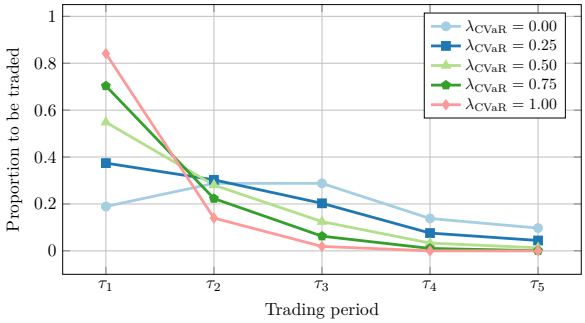
The impact of considering the uncertainty related to the volume to be traded on the optimal strategies is consistent with intuition: if the volume to be traded is uncertain, waiting to receive a more precise estimate of this volume is advantageous in terms of risk reduction. Indeed, we observe that a trader should mitigate their traded volume of the first trading period and spread it over the next periods. Moreover, the more risk-averse the trader, the bigger the impact on their optimal strategy.

These plots also illustrate that the impact of considering volume uncertainty on the optimal execution strategy is non-monotonic due to two different subsets of extreme events: those in which volume forecasts are increased in late trading periods, and those where volume forecasts are downsized. The first situation requires carrying out enough early trades to avoid having to upsize the traded volume in the latest stages to a level that is strongly affected by liquidity constraints. The second situation makes it necessary to avoid over-trading in the early trading periods and having to trade out of these positions later. The two effects tend to counterbalance each other, with the result of reallocating some of the early trades to the middle section of the trading period. This effect is picked up due to the incorporation of recourse estimates and the use of the CVaR$_\alpha$ measure that focuses on extreme events.

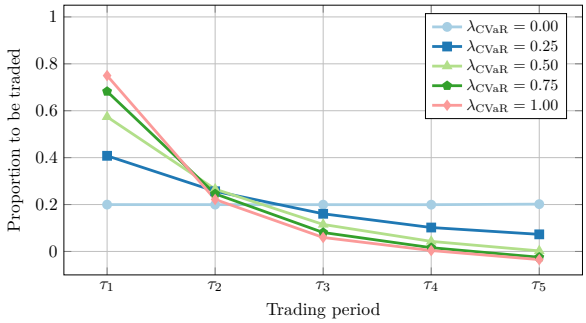
The liquidity profile has an impact both on the optimal strategies and on the adjustments to make to the strategies when considering the uncertainty related to the volume target additionally to the price uncertainty. First, compared to the situation where liquidity remains constant, a risk-neutral trader should shift their trades from less liquid to more liquid periods. This is not as simple for risk-averse traders, since the more risk-averse a trader is, the larger the impact of adversarial scenarios on their objective function, and consequently, the more the trader factors



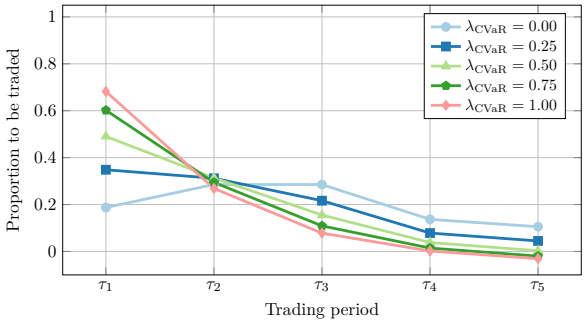
(a) Case (a): optimal trading strategies under PU only, *i.e.* $\forall k \in [\![1, m-1]\!] : \hat{\nu}_k = 0$.



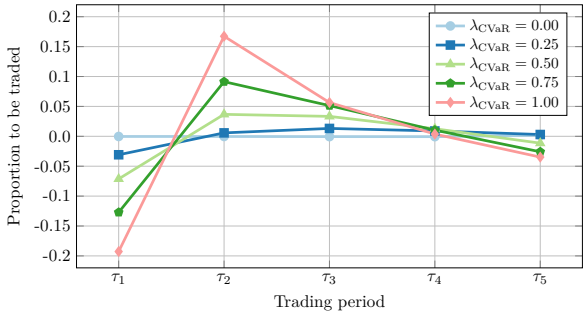
(b) Case (b): optimal trading strategies under PU only, *i.e.* $\forall k \in [\![1, m-1]\!] : \hat{\nu}_k = 0$.



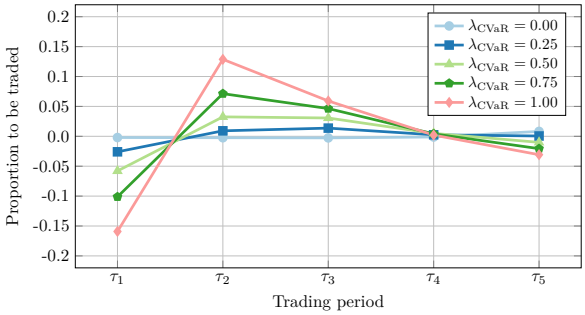
(c) Case (a): optimal trading strategies under PVU, *i.e.* $\forall k \in [\![1, m-1]\!] : \hat{\nu}_k = \nu_k$.



(d) Case (b): optimal trading strategies under PVU, *i.e.* $\forall k \in [\![1, m-1]\!] : \hat{\nu}_k = \nu_k$.



(e) Case (a): difference between the strategies represented in Figure 2.2c with the ones represented in Figure 2.2a.



(f) Case (b): difference between the strategies represented in Figure 2.2d with the ones represented in Figure 2.2b.

FIGURE 2.2: Optimal trading strategies under different assumptions for both test cases.

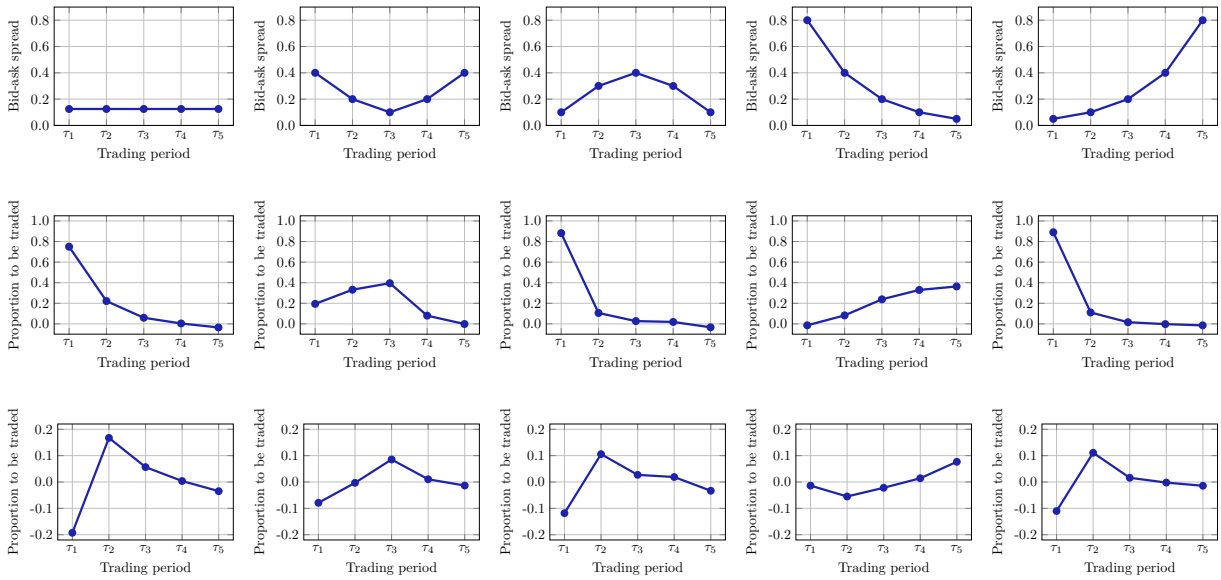


FIGURE 2.3: Analysis of the impact of the bid-ask spread profile on the strategy adjustments of an exclusively risk-focused trader ($\lambda_{\text{CVaR}} = 1$) when considering the uncertainty related to the volume target (row 1: bid-ask spread profile, row 2: optimal strategy in the mean-$\text{CVaR}_{0.1}$ (PVU) framework, row 3: strategy adjustments between the strategies under PU and PVU).

in the impact of market liquidity in conjunction with uncertainty. The decreasing influence of the market liquidity is observed in Figures 2.2b and 2.2d as the shape of the risk-neutral strategy gradually vanishes with increasing risk-aversion parameter, in favour of front loading to cope with price uncertainty. Second, the liquidity profile also impacts the adjustments to make to the strategies when introducing volume uncertainty into the model as illustrated in Figures 2.2e, 2.2f and 2.3; Figure 2.3 depicts the adjustments to make for several other liquidity profiles as well as the optimal strategies under PVU.

To cope with volume uncertainty, a trader has to consider the two different subsets of extreme events previously described. On the one hand, the less liquid the last trading periods, the more sensitive the trader to the risk of having to upsize the traded volume of the last trading periods, and therefore the smaller the planned proportions of the volume to be traded over these periods. On the other hand, a risk-averse trader considers the risk of over-trading by mitigating their trading quantity of the first trading period and redistributing it over the next periods. The less liquid the market during these next trading periods, the greater is the risk of over-trading, and therefore the more beneficial in terms of risk reduction is the decrease of the trading quantity of the first trading period. Figure 2.3 also shows that the adjustment in the trading proportion in the first trading period is redistributed onto the next trading periods according to the liquidity profile: the more liquid a trading period relatively to another, the favourable it is to increase its planned trade volume allocation rather than the one of this other period.

2.3.2 Model performance

In this section, we analyse the impact of considering volume uncertainty as part of the model on the strategy performance in case (a); similar observations can be made for case (b). We compare the performance of the strategies obtained when minimising the mean-$\text{CVaR}_{0.1}$ trade-off in two distinct cases. The first case corresponds to a trader that wrongly assumes that the total volume to be traded is fixed and will not change along the course of the execution of their strategy, *i.e.* $\forall k \in [\![1, m-1]\!] : \hat{\nu}_k = 0$. As a consequence, the corresponding strategies are the ones obtained if the trader considers that uncertainty only arises from the price dynamics (see Figure 2.2a and Table 2.2a). The second case corresponds to the situation where the trader estimates correctly the variability of the volume target forecasts, *i.e.* $\forall k \in [\![1, m-1]\!] : \hat{\nu}_k = \nu_k$, and takes it into account to define their trading strategy (see Figure 2.2c and Table 2.2b). Table 2.2 reports for both cases the expectation, the α-Conditional Value-at-Risk for $\alpha = 0.1$, 0.05 and 0.01, and the variance of the trading cost incurred when following the strategies obtained by solving Model (2.8).

Figure 2.4 compares the probability density functions (pdfs) of the trading cost of both situations. The more risk-averse a trader, the more pronounced the difference between the densities. Table 2.2 and Figure 2.4 clearly illustrate the benefit of including the uncertainty related to the volume target in the CVaR_α equivalent formulation of the risk-return trade-off model of Almgren and Chriss (2001): a trader assessing correctly the uncertainty of the volume target achieves better mean-$\text{CVaR}_{0.1}$ trade-offs.

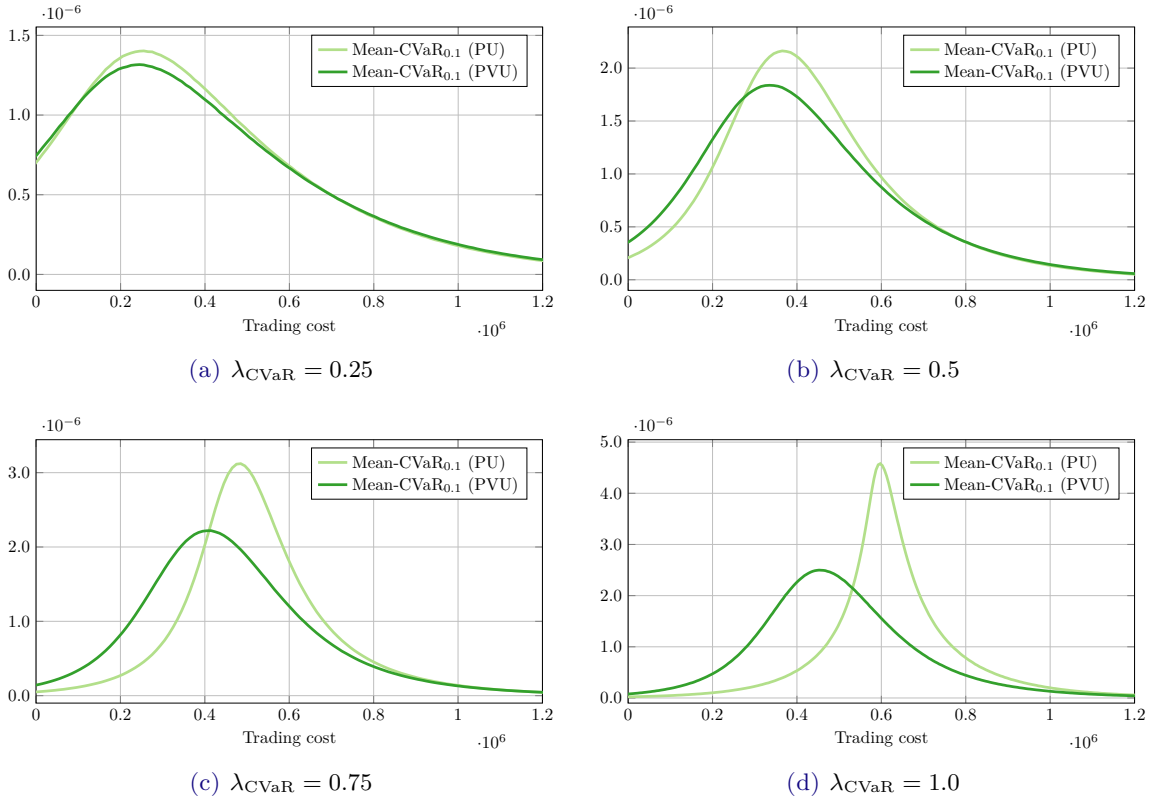
TABLE 2.2: Performance of the strategies obtained by solving Optimisation Problem (2.8) for different risk-aversion parameter values λ_{CVaR}.

(a) A trader considers that the only source of uncertainty is the price dynamics (strategies under PU only depicted in Figure 2.2a).

λ_{CVaR} [1]	Expectation [$]	$\text{CVaR}_{0.1}$ [$]	$\text{CVaR}_{0.05}$ [$]	$\text{CVaR}_{0.01}$ [$]	Variance [$^2]	$\varphi_{0.1}^{\lambda_{\text{CVaR}}}$ [$]
0.00	2.973e5	1.351e6	1.586e6	2.093e6	3.085e11	2.973e5
0.25	3.412e5	1.062e6	1.242e6	1.641e6	1.334e11	5.216e5
0.50	4.303e5	9.445e5	1.088e6	1.419e6	6.393e10	6.874e5
0.75	5.268e5	9.320e5	1.055e6	1.349e6	3.930e10	8.307e5
1.00	6.273e5	9.833e5	1.096e6	1.368e6	3.143e10	9.833e5

(b) A trader considers that the uncertainty arises from both the price dynamics and the total volume to be traded (strategies under PVU depicted in Figure 2.2c).

λ_{CVaR} [1]	Expectation [$]	$\text{CVaR}_{0.1}$ [$]	$\text{CVaR}_{0.05}$ [$]	$\text{CVaR}_{0.01}$ [$]	Variance [$^2]	$\varphi_{0.1}^{\lambda_{\text{CVaR}}}$ [$]
0.00	2.973e5	1.353e6	1.589e6	2.096e6	3.096e11	2.973e5
0.25	3.331e5	1.084e6	1.268e6	1.676e6	1.464e11	5.210e5
0.50	4.015e5	9.638e5	1.113e6	1.456e6	7.877e10	6.826e5
0.75	4.586e5	9.274e5	1.058e6	1.366e6	5.474e10	8.102e5
1.00	4.982e5	9.211e5	1.041e6	1.330e6	4.502e10	9.211e5



(a) $\lambda_{\text{CVaR}} = 0.25$ (b) $\lambda_{\text{CVaR}} = 0.5$

(c) $\lambda_{\text{CVaR}} = 0.75$ (d) $\lambda_{\text{CVaR}} = 1.0$

FIGURE 2.4: Comparison of the pdfs of the trading cost when the uncertainty related to the total volume to be traded is or is not taken into account in Model (2.8) (the pdfs are estimated based on 10^8 realisations of the underlying processes).

2.3.3 Comparison of the performance of the different frameworks

In the previous section, we noted that a trader who correctly assesses the volume uncertainty achieves better mean-CVaR$_{0.1}$ trade-offs at the expense of a greater variance in their trading costs. In this section, we compare our model with the ones in the literature. To evaluate the relative performance of each framework, we decided to use as comparison criteria the expectation and the α-Conditional Value-at-Risk of the trading costs; the variance being not of utmost importance as justified hereafter. Additionally, we also compare the probability density function (pdf) and cumulative distribution function (cdf) of the trading costs of the different frameworks.

Mean-variance with recourse. First, as mentioned previously, the mean-variance framework of Almgren and Chriss (2001) is in practice not directly applicable in the presence of an uncertain volume target due to the constraint of Optimisation Problem (1.7). However, a natural solution to circumvent this issue is to use it in conjunction with a systematic recourse at every trading period. Since the objective function in the mean-variance and mean-CVaR$_\alpha$ frameworks differs, it is not straightforward to compare the performance of the two models as we do not have a mapping between $\lambda_{\mathbb{V}\text{ar}}$ and λ_{CVaR}. However, the optimal strategy in the mean-variance framework converges when the risk-aversion increases. Hence, the parameter values

$\lambda_{\mathrm{CVaR}} = 1$ and $\lambda_{\mathbb{V}\mathrm{ar}} \to +\infty$ both correspond to a trader that is exclusively risk focused; these values enable us to compare the performance of both models. The strategy of such a risk focused trader in the mean-variance framework with recourse consists of trading the entire initial forecast of the total volume to be traded during the first trading period and then, during the subsequent trading periods τ_k, $k \in [\![2, m]\!]$, trading the amount δ_{k-1} of shares, which corresponds to the forecast update released at time t_{k-1}. This strategy can be reproduced in our model by choosing $\boldsymbol{y} = [1, 0, \ldots, 0]^{\mathrm{T}}$ and $\boldsymbol{\beta} = (\beta_{i,k})$ with $\beta_{i,k} = 1$ if $k = i + 1$ and 0 otherwise. The last line of Tables 2.2a and 2.3 reports the performance of an exclusively risk focused trader in both models; note that in both situations the trader integrates only price uncertainty into their estimate of the trading cost but the actual trading costs are evaluated based on a market subjected to both sources of uncertainty. The last line of Table 2.2b, which uses the same value of λ_{CVaR}, illustrates the benefit of factoring in the trade volume uncertainty: both the expectation and the $\mathrm{CVaR}_{0.1}$ of the trading cost are significantly reduced. The standard deviation of the trading cost slightly increases from 1.773e5 to 2.122e5, but this is irrelevant: Indeed, Figure 2.5, which depicts the probability densities of the trading cost for both models, shows empirically that the random variable that represents the trading cost in the mean-variance framework with recourse is first order stochastically dominant over the trading cost of the strategy that results from Model (2.8). As a consequence, a trader who is driven by profit should adopt the strategy proposed by Model (2.8) as the increased variance is due to increased downward risk, which is beneficial.

TABLE 2.3: Performance of the strategies derived from the mean-variance framework with recourse for different risk-aversion parameter values $\lambda_{\mathbb{V}\mathrm{ar}}$.

$\lambda_{\mathbb{V}\mathrm{ar}}$ [1/\$]	Expectation [\$]	$\mathrm{CVaR}_{0.1}$ [\$]	$\mathrm{CVaR}_{0.05}$ [\$]	$\mathrm{CVaR}_{0.01}$ [\$]	Variance [\$^{2}]
0.0	2.974e5	1.351e6	1.587e6	2.094e6	3.087e11
10^{-6}	3.626e5	1.016e6	1.182e6	1.554e6	1.076e11
10^{-5}	5.616e5	9.336e5	1.042e6	1.302e6	3.335e10
10^{-4}	6.716e5	9.935e5	1.088e6	1.319e6	2.722e10
$\to +\infty$	6.912e5	1.008e6	1.102e6	1.328e6	2.698e10

Exponential utility and mean-QV. We have previously seen that the model proposed by Cheng et al. (2017) could be slightly transformed in order to tackle a similar problem to ours but in continuous time. In their model, the authors assume that a risk-averse trader maximises either the exponential utility of their final P&L (profit-and-loss) $\widetilde{\Pi}_{t_f}(z)$, or the expectation of the final P&L penalised by its quadratic variation. In both situations, the authors provide closed form expressions for the optimal trading trajectory $z_t^\star$ when no residual volume is tolerated, *i.e.* when $\beta \to \infty$. When allowing enough trading periods, *e.g.* $m = 10^2$, we observe numerically that the discrete approximation of the strategy derived from the model of Cheng et al. (2017) has nearly exactly the same final P&L compared to the one of the continuous strategy. Hence, to compare

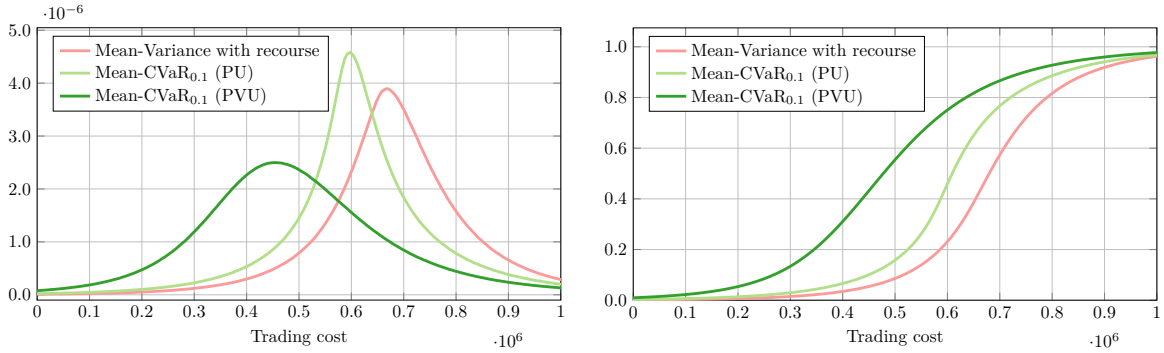


FIGURE 2.5: Comparison of the pdfs and cdfs of the trading cost of a risk-averse trader computing their strategy with the mean-variance framework with recourse for $\lambda_{\mathbb{V}\text{ar}} \to \infty$ and with the mean-CVaR$_{0.1}$ (PU and PVU) framework (Model (2.8)) for $\lambda_{\text{CVaR}} = 1$ (the pdfs are estimated based on 10^8 realisations of the underlying processes).

their model with ours, we consider the case (a) described in Table 2.1 with the difference that we assume one trading day ($T = 1$) divided in one hundred trading periods ($m = 10^2$). Note that the final P&L $\widetilde{\Pi}_{t_f}(z)$ of a strategy z is equivalent to minus its trading cost.

Cheng et al. (2017) first argue that the strategy resulting from the exponential utility (EU) maximisation can be regarded as a version of an adaptive Almgren-Chriss strategy. When the uncertainty of the volume target disappears, *i.e.* $m_0 \to 0$, the optimal strategy given the risk-aversion parameter θ_{EU} obtained via the exponential utility framework recovers the classical Almgren-Chriss strategy with parameter value $\lambda_{\mathbb{V}\text{ar}} = \frac{\theta_{\text{EU}}}{2}$. Hence, we can consider that a risk-averse trader with risk-aversion parameter $\lambda_{\mathbb{V}\text{ar}}$ in the mean-variance framework would have a risk-aversion parameter θ_{EU} equal to $2\lambda_{\mathbb{V}\text{ar}}$ in the exponential utility framework; this provides a mapping between $\lambda_{\mathbb{V}\text{ar}}$ and θ_{EU} that allows the comparison of both frameworks across the range of risk-aversion levels. By comparing Tables 2.4a and 2.4b, we observe that the exponential utility framework produces results of the same order as the mean-variance framework with recourse.

While Cheng et al. (2017) employ the exponential utility for the sake of tractability, the authors also provide, in the special case where the uncertainty of order fills is modelled with a diffusion process, a closed form solution of the optimal trading rate in the mean-QV framework by solving analytically the dynamic program. They motivate the use of the quadratic variation since, as pointed out by Almgren (2012), it captures the bulk of uncertainty. The optimal trading rate in this model can be regarded as an adaptive Almgren-Chriss strategy with the difference that the characteristic time scales of liquidation are slightly different due to the uncertain order fills (Cheng et al., 2017). When $m_0 \to 0$, the optimal strategy given the risk-aversion parameter λ_{QV} recovers the classical Almgren-Chriss strategy with parameter value $\lambda_{\mathbb{V}\text{ar}} = \lambda_{\text{QV}}$. This provides a mapping between $\lambda_{\mathbb{V}\text{ar}}$ and λ_{QV}. The performance of the strategies derived with this model given different risk-aversion parameter values λ_{QV} is given in Table 2.4c: the more risk-averse the trader, the smaller the variance of their trading cost. Compared to the two previous models (mean-variance with recourse and exponential utility maximisation),

TABLE 2.4: Trading cost comparison of the trading strategies for the different frameworks studied with an execution period subdivided in $m = 10^2$ trading periods.

(a) Mean-variance with recourse

$\lambda_{\mathbb{V}\mathrm{ar}}$ [1/$]	Expectation [$]	$\mathrm{CVaR}_{0.1}$ [$]	$\mathrm{CVaR}_{0.05}$ [$]	$\mathrm{CVaR}_{0.01}$ [$]	Variance [$^2]
0	8.133e5	1.411e6	1.532e6	1.784e6	1.006e11
10^{-6}	8.154e5	1.399e6	1.519e6	1.765e6	9.557e10
10^{-5}	9.155e5	1.409e6	1.510e6	1.722e6	6.792e10
10^{-4}	2.071e6	2.424e6	2.496e6	2.647e6	3.547e10
$\to +\infty$	6.311e7	6.330e7	6.334e7	6.341e7	9.803e9

(b) Exponential utility

θ_{EU} [1]	Expectation [$]	$\mathrm{CVaR}_{0.1}$ [$]	$\mathrm{CVaR}_{0.05}$ [$]	$\mathrm{CVaR}_{0.01}$ [$]	Variance [$^2]
$\to 0$	7.861e5	1.264e6	1.348e6	1.512e6	7.423e10
$2 \cdot 10^{-6}$	7.883e5	1.252e6	1.334e6	1.493e6	7.002e10
$2 \cdot 10^{-5}$	9.354e5	1.313e6	1.379e6	1.510e6	4.631e10
$2 \cdot 10^{-4}$	3.466e6	3.640e6	3.670e6	3.730e6	9.814e9
$\to +\infty$	2.780e7	2.794e7	2.797e7	2.802e7	7.081e9

(c) Mean-QV

λ_{QV}[1/$]	Expectation [$]	$\mathrm{CVaR}_{0.1}$ [$]	$\mathrm{CVaR}_{0.05}$ [$]	$\mathrm{CVaR}_{0.01}$ [$]	Variance [$^2]
$\to 0$	7.860e5	1.263e6	1.346e6	1.508e6	7.389e10
10^{-6}	7.885e5	1.255e6	1.336e6	1.495e6	7.075e10
10^{-5}	8.791e5	1.276e6	1.346e6	1.482e6	5.133e10
10^{-4}	1.637e6	1.902e6	1.948e6	2.038e6	2.282e10
$\to +\infty$	2.534e6	2.741e6	2.777e6	2.848e6	1.393e10

(d) Mean-$\mathrm{CVaR}_{0.1}$ (PVU)

λ_{CVaR} [1]	Expectation [$]	$\mathrm{CVaR}_{0.1}$ [$]	$\mathrm{CVaR}_{0.05}$ [$]	$\mathrm{CVaR}_{0.01}$ [$]	Variance [$^2]
0.00	8.132e5	1.409e6	1.530e6	1.780e6	1.001e11
0.25	8.153e5	1.398e6	1.517e6	1.764e6	9.529e10
0.50	8.206e5	1.387e6	1.503e6	1.742e6	9.035e10
0.75	8.286e5	1.382e6	1.495e6	1.729e6	8.597e10
1.00	8.396e5	1.380e6	1.491e6	1.722e6	8.186e10

for large risk-aversion parameters we observe a gain of one order of magnitude for both the expectation and $\text{CVaR}_{0.1}$ of the trading costs, whereas the variance remains (approximately) of the same order.

Finally, we consider the model presented in this chapter where a trader considers the uncertainty arising from both the price dynamics and the updates on the volume target, *i.e.* the mean-$\text{CVaR}_{0.1}$ (PVU) framework. The performance of the strategies obtained with this model is shown in Table 2.4d.

In a first step, as a sanity check, let us consider a risk-neutral trader, *i.e.* $\lambda_{\mathbb{V}\text{ar}} = 0$, $\theta_{\text{EU}} \to 0$, $\lambda_{\text{QV}} = 0$, and $\lambda_{\text{CVaR}} = 0$. In this case all the frameworks minimise the same objective function, *i.e.* the expectation of the trading costs. The first line of Tables 2.4a and 2.4d show that the mean-variance with recourse and the mean-$\text{CVaR}_{0.1}$ (PVU) frameworks produce nearly identical results; this is coherent with Lemma 2.4. Indeed, since the redistribution matrix $\boldsymbol{\beta}$ was chosen as the one implied by the risk-neutral strategy of the mean-variance framework computed at time t_0 (see Equation (2.36)), the optimal strategy in the mean-variance framework with recourse belongs to the feasible set of strategies of the mean-$\text{CVaR}_{0.1}$ (PVU) framework. Besides, we observe that the exponential utility and mean-QV frameworks are performing better than the two previous frameworks. Since both frameworks solve the same dynamical program, the discrepancy between the values reported in Tables 2.4b and 2.4c for a risk-neutral trader can be attributed to numerical errors. This better performance arises from the fact that these frameworks solve the dynamic program and reach consequently the optimal trading strategy. We can therefore estimate the sub-optimality gap of our model to $\frac{(8.129e5-7.861e5)}{7.855e5} \simeq 3.5\%$. This can be seen as the price to pay for taking only partial recourse via a predetermined rule. In a second step, we compare our model with the mean-QV framework for an exclusively risk-focused trader, which can be modelled by taking $\lambda_{\text{QV}} \to +\infty$ and $\lambda_{\text{CVaR}} = 1$. Tables 2.4c and 2.4d illustrate that, if a trader allows for a large variance of their trading cost, then a significant reduction in the expectation and α-Conditional Value-at-Risk of the trading costs can be achieved by using our new model compared to the mean-QV framework. Similarly as for the comparison with the mean-variance framework with recourse, we observe empirically in Figure 2.6 that the random variable representing the trading cost in the mean-QV framework is first order stochastically dominant over the trading cost of the strategy resulting from Model (2.8). Again, letting the variance increase enables our model to shift the distribution of the trading cost to smaller values, which is beneficial. In fact, the increase in variance is due to a larger spread in lower trading costs.

Finally, Table 2.4 reports that a risk-neutral trader in the mean-QV framework achieves a smaller value for the $\text{CVaR}_{0.1}$ of the trading costs compared to a trader that exclusively aims to minimise $\text{CVaR}_{0.1}$ in the mean-$\text{CVaR}_{0.1}$ (PVU) framework (*i.e.* $\lambda_{\text{CVaR}} = 1$). This may call into question the advantage of our model, however two important points should be brought to attention. Firstly, despite the apparent better performance of the mean-QV framework, we should keep in mind that the analytical solution to the dynamic program to solve to obtain the

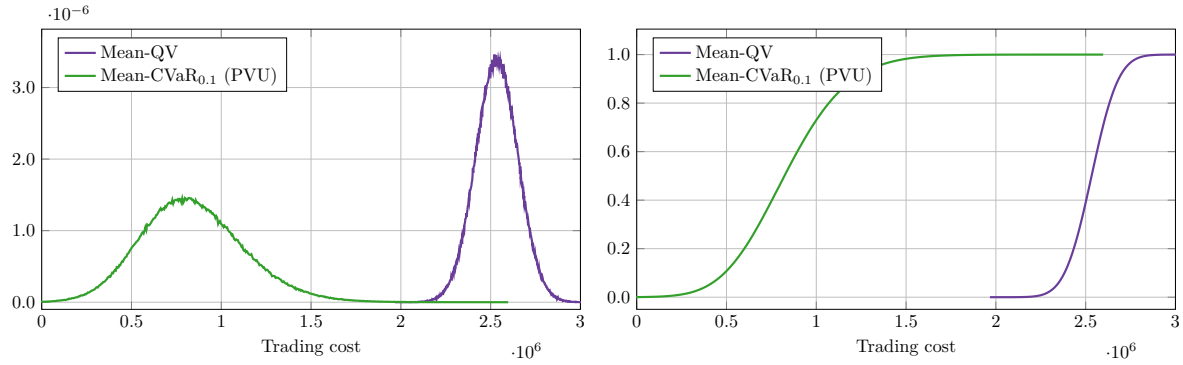


FIGURE 2.6: Comparison of the pdfs and cdfs of the trading cost of a risk-averse trader computing their strategy with the mean-QV framework proposed by Cheng et al. (2017) for $\lambda_{\text{QV}} \to \infty$ and with the mean-CVaR$_{0.1}$ (PVU) framework (Model (2.8)) for $\lambda_{\text{CVaR}} = 1$ (the pdfs are estimated based on 10^6 realisations of the underlying processes).

optimal mean-QV trading rate is available only for special well behaved cases. In the general case, obtaining an analytical solution of the dynamic program would be more challenging, if not impossible, and obtaining a numerical solution would be significantly more computationally expensive compared to our approach. Secondly, the quadratic variation fails to correctly take into account the risk-aversion of a trader. Indeed, as represented in Figure 2.7 and reported in Table 2.4c, we generally observe that, the greater the risk-aversion level of the trader (*i.e.* λ_{QV}), the greater the expectation and CVaR$_{0.1}$ of their trading costs. Hence, the decrease of the variance of the trading costs is coupled with a shift of the distribution of the trading costs to greater values, which is a rationale that would not be encountered in practice.

In conclusion, our model is significantly more versatile as it does not rely on any assumptions on the underlying distributions modelling the uncertainty related to the price dynamics and the forecast updates of the volume target. It is also computationally cheaper than the alternative approaches, and it coherently models the risk-aversion level of a trader: the expectation of the worst-cases trading costs decreases as the risk-aversion level of the trader increases.

2.3.4 Real world application: power trading

One of the main advantages of our model is that it does not rely on any particular assumptions on the distributions of the price moves and forecast updates. It can consequently be applied to a wider range of applications compared to the models of Almgren and Chriss (2001) and Cheng et al. (2017). Here, we show how our model can be applied to power trading, where for each half hourly settlement period, power wholesalers (*resp.* producers) trade power future contracts with the objective of achieving a contractual net position that matches the expected power demand of their end-consumers (*resp.* power generation).

In Great Britain, for every given settlement period, the market participants can trade future contracts from several years ahead down until the market closure, which occurs one hour before

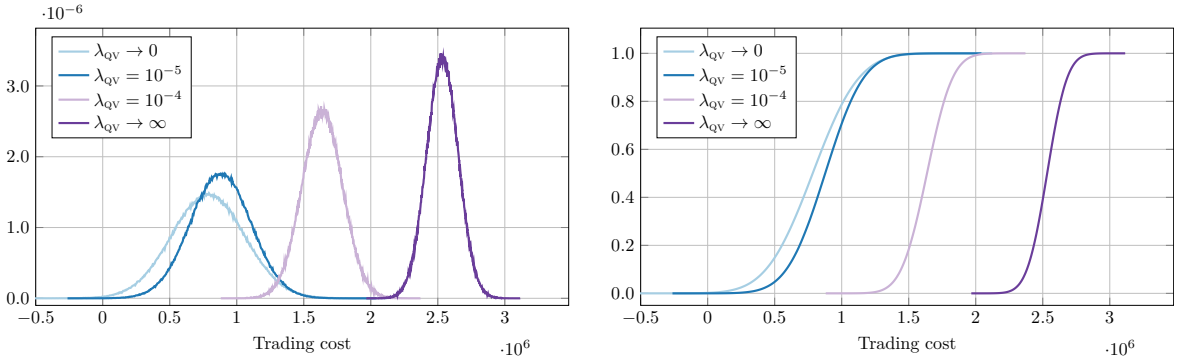


FIGURE 2.7: Comparison of the pdfs and cdfs of the trading cost of a risk-averse trader computing their strategy with the mean-QV framework proposed by Cheng et al. (2017) for different risk-aversion parameter values (the pdfs are estimated based on 10^6 realisations of the underlying processes).

the start of delivery, *i.e.* the start of the settlement period. For any given settlement period starting at time t_f, a wholesaler aims to achieve a final net position close to the realised power demand D_{t_f} of their end-consumers. If they supply power to a fraction ρ of the population, the power demand of their customers is here assumed to be given by $D_{t_f} = \rho V_{t_f}$, where V_{t_f} denotes the national power demand during the settlement period. Any residual volume resulting from the disparity between their contractual position against the realised demand D_{t_f} is charged at the imbalance price, which results from the balancing mechanism operated by the System Operator after the market closure. For any half-hour settlement period, there is consequently no obligation for a market participant to balance their contractual positions against the expected demand (or *resp.* generation). However, the exposure to this imbalance price drives the incentive to avoid any residual volume.

We investigate here how a risk-averse power wholesaler in Great Britain should acquire power positions ahead of a given settlement period in order to minimise their costs while factoring in risk as suggested by our Model (2.8). The market is modelled as follows: The settlement period is one half-hour long and starts at the delivery time t_f. The trading window spans from 48 to 2 half-hours ahead of delivery time; the initial trading time is denoted t_s. The trading window is subdivided into 46 half-hour long trading periods. To be in line with the notation introduced in Section 1.1, the k-th trading period is denoted by τ_k and corresponds to the half-hour period that starts $48 - k + 1$ half-hours ahead of the settlement period (see Figure 2.8).

To make their trading decisions, the market participants use the operational data relating to the electricity balancing and settlement arrangements in Great Britain that is published by the *Balancing Mechanism Reporting Service* (BMRS). Using the data of the past five years, Figure 2.9 represents the trend of the forecast updates of the national power demand in Great Britain: Figure 2.9a illustrates the mean of the relative forecast error, which is computed as follows $\left|V_k - V_{t_f}\right| / V_{t_f}$, where V_{t_f} denotes the realised national demand and V_k the demand forecast released after the i-th trading period τ_k, *i.e.* at decision time t_k. Figure 2.9b represents

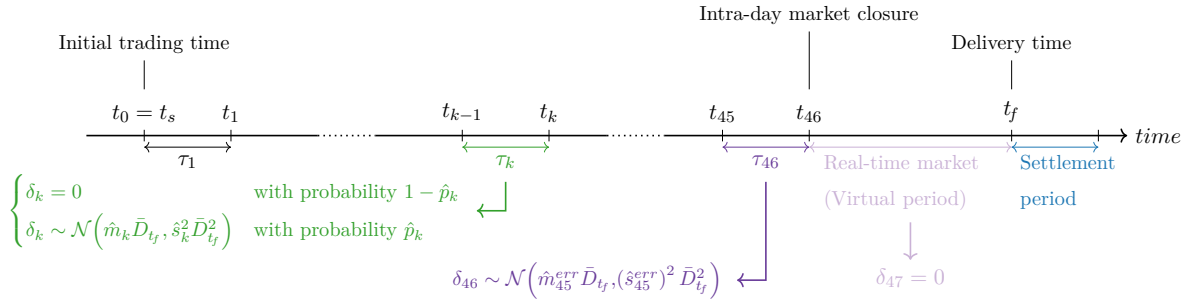


FIGURE 2.8: Power trading time line and modelling assumptions.

the probability p_k of having a forecast update of the power demand at time t_k. Conditional on there being a forecast update, Figures 2.9c and 2.9d represent respectively the mean m_k and standard deviation s_k of the relative forecast update, which is given by $(V_k - V_{i-1})/V_{t_f}$. For all graphs, we have fitted a polynomial approximation of the data points; let $\hat{p}_k$, $\hat{m}_k$, and $\hat{s}_k$ denote respectively the value of the polynomial approximating the probability, the mean and the standard deviation of the forecast update that comes to the trader's knowledge at time t_k. Based on these values, we assume that, for each trading period τ_k with $k \leq 45$, a wholesaler estimates the distribution of the related forecast update on the power demand of their customers as follows: Firstly, there is a probability $\hat{p}_k$ of having a forecast update. Secondly, conditional on a forecast update occurring, the forecast update δ_k is assumed to follow the normal distribution $\mathcal{N}\left(\hat{m}_k \bar{D}_{t_f}, \hat{s}_k^2 \bar{D}_{t_f}^2\right)$, where $\bar{D}_{t_f} = \rho \bar{V}_{t_f}$ and $\bar{V}_{t_f}$ denotes the average realised national power demand. Finally, because the market closure happens one hour before the start of delivery, a trader cannot take into consideration in his trading strategy any forecast update released after the decision time t_{45}. Therefore, the residual volume that should have been traded and that will be exchanged at the imbalance price is given by the random variable $\rho\left(V_{t_f} - V_{45}\right)$. This is taken into consideration in our model by assuming that the forecast update δ_{46} of the last trading period follows the normal distribution $\mathcal{N}\left(\hat{m}_{45}^{err} \bar{D}_{t_f}, (\hat{s}_{45}^{err})^2 \bar{D}_{t_f}^2\right)$, where $\hat{m}_{45}^{err}$ and $\hat{s}_{45}^{err}$ are respectively the estimates of the average and standard deviation of the relative forecast error at decision time t_{45}, *i.e.* $\left(V_{t_f} - V_{45}\right)/V_{t_f}$. Additionally to these 46 trading periods, we add a virtual 47-th trading period, which corresponds to the balancing period after the market closure. Note that during this balancing period, we artificially assume that there is no forecast update on the demand as it has already been taken into consideration in the forecast update distribution of τ_{46}. To incentivise the trader to trade the entirety of their position before the balancing period τ_{47}, we penalise any outstanding volume by fixing a larger temporary impact parameter η_{47} for the balancing period τ_{47} compared to the other trading periods; a value three times larger is here used (see Table 2.5). In a similar vein as the penalty term parametrised by β in the model of Cheng et al. (2017), the larger η_{47}, the more any outstanding volume is penalised. As suggested in Section 2.2.2, we assume that the residual volumes to be traded due to the forecast updates are redistributed over the future trading periods based on the redistribution matrix $\boldsymbol{\beta}$ implied by the risk-neutral optimal trading strategy $\boldsymbol{y}^{\star,t_0,\lambda_{\mathbb{V}\mathrm{ar}}=0}$ in the mean-variance framework.

From the perspective of the price distributions, we consider an initial price S_0 of 90. For the sake of simplicity (the data not being publicly available), we assume that the volatility σ of the asset is constant during the entire execution period and is equal to 0.9, and that for each trading period τ_k the price shift ξ_k follows the normal distribution $\mathcal{N}(0, \tau_k\sigma^2)$. The rest of the parameters are fixed as follows: the average realised national power demand V_{t_f} equals $3{\cdot}10^4$MW. The wholesaler provides power to $\rho = \frac{1}{3}$ of the population, has an initial position of 0, and has an initial estimate of the power demand of the end-consumers equal to $D_0 = \bar{D}_{t_f} = \rho\bar{V}_{t_f} = 10^4$. Finally, the bid-ask spread of each trading period is assumed to be equal to 0.25% of the initial asset price S_0, *i.e.* $\forall k \in [\![1, m]\!] : b_k = 0.225$; the values used for temporary and permanent impact parameters can be found in Table 2.5.

TABLE 2.5: Power market parameters

Number of trading periods:	$m = 47$
Trading period length:	$\forall k \in [\![1, m]\!] : \tau_k = \frac{1}{48}$ [day]
Initial future contract price:	$S_0 = 90$ [£/MW]
Asset's volatility:	$\sigma = 0.9$ [(£/MW)/day$^{1/2}$]
Price shifts distribution:	$\forall k \in [\![1, m]\!] : \xi_k \sim \mathcal{N}(0, \tau_k\sigma^2)$
Average realised national power demand:	$\bar{V}_{t_f} = 3 \cdot 10^4$ [MW]
Proportion of the population supplying:	$\rho = \frac{1}{3}$
Power demand initial estimate:	$D_0 = \bar{D}_{t_f} = \rho\bar{V}_{t_f} = 10^4$ [MW]
Volume forecast updates distribution:	$\forall k \in [\![1, 45]\!] : \delta_k \sim \mathcal{N}\left(\hat{m}_k \bar{D}_{t_f}, \hat{s}_k^2 \bar{D}_{t_f}^2\right)$
	$\delta_{46} \sim \mathcal{N}\left(\hat{m}_{45}^{err} \bar{D}_{t_f}, (\hat{s}_{45}^{err})^2 \bar{D}_{t_f}^2\right)$
	$\delta_{47} = 0$
Bid-ask spread:	$\forall k \in [\![1, m]\!] : b_k = 0.25\%\, S_0 = 0.225$ [£/MW]
Fixed cost:	$\forall k \in [\![1, m]\!] : \epsilon_k = 0.5 b_k = 0.1125$ [£/MW]
Impact at 2% of market:	$\forall k \in [\![1, 46]\!] : \eta_k = b_k/(0.02\bar{V}_{t_f}) = 3.75 \cdot 10^{-5}$ [(£/MW)/(MW/day)]
	$\eta_{47} = 3b_{47}/(0.02\bar{V}_{t_f}) = 11.25 \cdot 10^{-5}$ [(£/MW)/(MW/day)]
Permanent impact parameter:	$\forall k \in [\![1, m]\!] : \gamma_k = (5b_k)/(0.1\bar{V}_{t_f}) = 3.75 \cdot 10^{-5}$ [£/MW^2]
CVaR$_\alpha$ parameter:	$\alpha = 0.1$

Similarly to Figure 2.2, Figure 2.10a depicts the trading strategies returned by our model for a risk-averse wholesaler in Great Britain when they consider the uncertainty related to both the price dynamics and the power demand of the end-consumers. For different risk-aversion levels, Figure 2.10b depicts how the trading strategy is modified compared to the situation where the wholesalers wrongly assumes that the power demand of the end-consumers is perfectly known at the start of the execution period. The difference in performance of the trading strategies obtained in both situations is listed in Table 2.6. When we consider the uncertainty related to the power demand, better mean-CVaR$_{0.1}$ trade-offs are achieved for any risk-level. We do not further analyse the results as similar comments can be made compared with the previous test cases (see Sections 2.3.1 and 2.3.2).

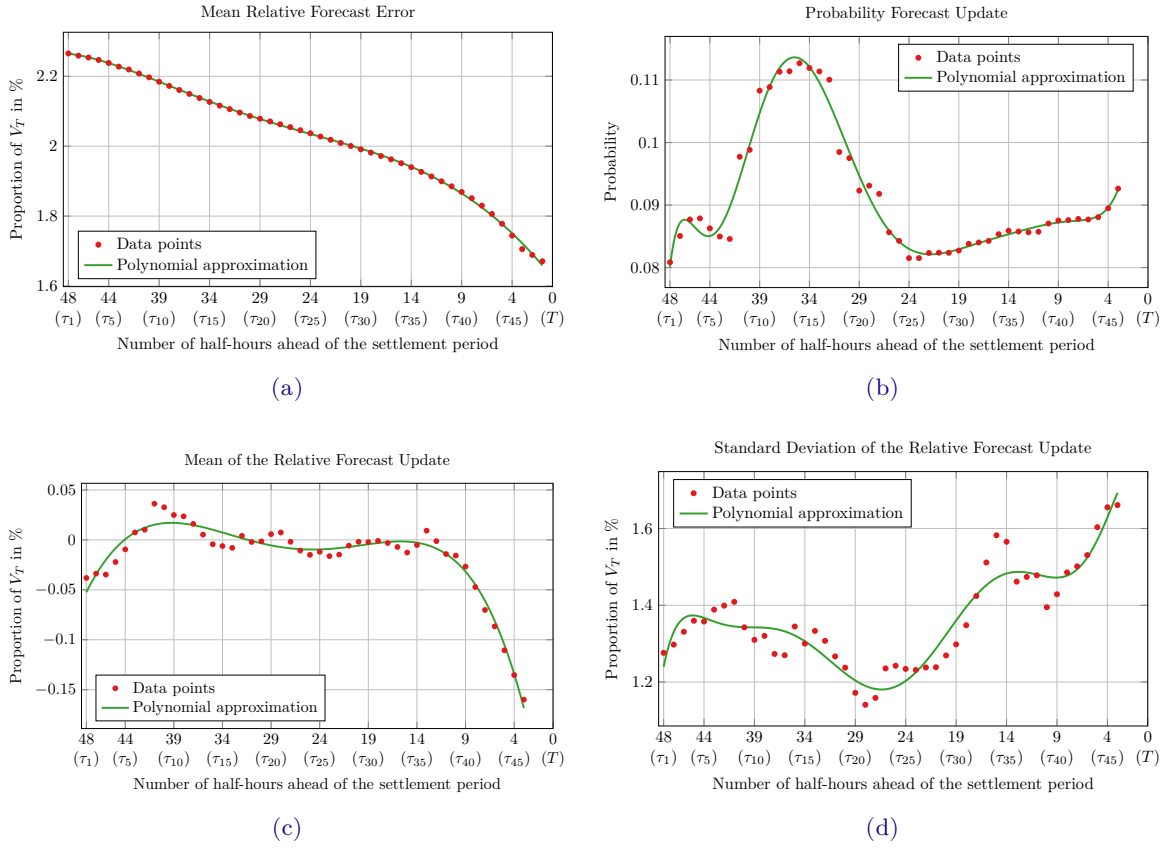
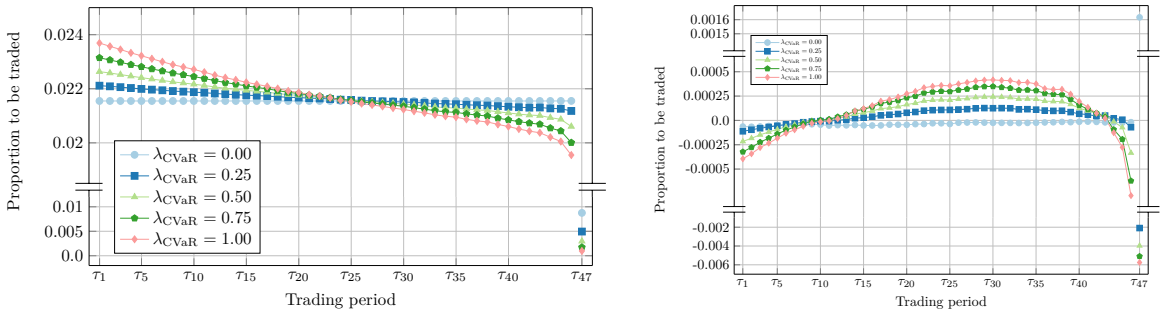


FIGURE 2.9: Analysis of the forecast updates of the national power demand in Great Britain.

2.4 Discussion

Our choice to use the α-Conditional Value-at-Risk as risk measure in Model (2.8) was guided by its widespread use and ease of interpretation, despite the controversy that this risk measure may lead to time inconsistent optimal policies: the policy obtained by optimising the model at an intermediate stage of the execution period is not necessarily equivalent to the continuation of the optimal policy computed at the initial time t_0 (Shapiro, 2009; Rudloff et al., 2014). In Section 2.3 we saw however that due to the forward planning of recourse with a predetermined trade volume re-allocation rule, our model competes favourably with established alternatives, yielding lower trading cost in expectation and α-Conditional Value-at-Risk. Our merit function accounts for the trade-off between the partially conflicting goals of lowering the expectation and the α-Conditional Value-at-Risk of trading costs and is computationally cheaper than either taking recourse under the mean-variance framework whenever forecast updates on the volume target become available or solving a dynamic program in order to optimise the mean-QV trade-off. These advantages of our model outweigh the disadvantage that the time inconsistency of the Mean-CVaR$_\alpha$ risk measure may lead to suboptimal strategies (Rudloff et al., 2014).

The model proposed in this chapter should thus be considered as an alternative to the traditional methods for its competitive performance, its ease of interpretation, and for its relatively



(a) (b)

FIGURE 2.10: Trading power in Great Britain: Figure 2.10a: optimal trading strategies under PVU, Figure 2.10b: difference between the strategies under PVU and PU.

TABLE 2.6: Trading power in Great Britain: performance of the strategies obtained with (2.8) for different risk-aversion parameter values λ_{CVaR}.

(a) A trader considers that the only source of uncertainty is the price dynamics.

λ_{CVaR} [1]	Expectation [£]	$\text{CVaR}_{0.1}$ [£]	$\text{CVaR}_{0.05}$ [£]	$\text{CVaR}_{0.01}$ [£]	Variance [£^{2}]	$\varphi_{0.1}^{\lambda_{\text{CVaR}}}$ [£]
0.00	6.102e4	7.572e4	7.945e4	8.841e4	5.558e7	6.102e4
0.25	6.103e4	7.566e4	7.938e4	8.832e4	5.503e7	6.469e4
0.50	6.105e4	7.562e4	7.933e4	8.825e4	5.449e7	6.834e4
0.75	6.109e4	7.559e4	7.929e4	8.819e4	5.396e7	7.196e4
1.00	6.114e4	7.559e4	7.928e4	8.818e4	5.345e7	7.559e4

(b) A trader considers that the uncertainty arises from both the price dynamics and the power demand.

λ_{CVaR} [1]	Expectation [£]	$\text{CVaR}_{0.1}$ [£]	$\text{CVaR}_{0.05}$ [£]	$\text{CVaR}_{0.01}$ [£]	Variance [£^{2}]	$\varphi_{0.1}^{\lambda_{\text{CVaR}}}$ [£]
0.00	6.101e4	7.606e4	7.993e4	8.922e4	5.751e7	6.101e4
0.25	6.110e4	7.531e4	7.885e4	8.737e4	5.282e7	6.465e4
0.50	6.122e4	7.510e4	7.850e4	8.668e4	5.094e7	6.816e4
0.75	6.134e4	7.503e4	7.837e4	8.638e4	4.988e7	7.161e4
1.00	6.144e4	7.502e4	7.832e4	8.623e4	4.914e7	7.502e4

cheap computational cost. If we desire to take full recourse in order to further improve the performance of our model, we could replace the CVaR_α term by the dynamic coherent risk measure (Riedel, 2004) induced by the CVaR_α risk measure. Lin et al. (2015) explored such an approach in the context of the traditional trade execution problem under price uncertainty only. An extension to the case with volume uncertainty would imply the applicability of the dynamic programming principle, which guarantees the time consistency of the optimal strategies but at the cost of a greater computational time. This is the subject of the next chapter.

Among the parameters for our model we need to specify a redistribution matrix to decide how any residual trading volumes that become apparent after forecast updates are redistributed over future trading periods. In the numerical section of this chapter we chose this redistribution matrix according to the trade volume proportions implied by the risk-neutral optimal trading

strategy $\boldsymbol{y}^{\star,t_0,\lambda_{\mathbb{V}\mathrm{ar}}=0}$ under the mean-variance framework of Almgren and Chriss (2001). This choice is not guaranteed to be optimal, and we could optimize over the redistribution matrix as well. Although we did not address this issue in the present work, we believe it is of interest for future research. Numerical simulations we carried out suggest that a mild additional cost savings can be achieved.

Finally, in this work we assumed that the random variables that model forecast updates are identically distributed. In practice this assumption is often not justified, such as in power markets, where empirical data show that the standard deviation of forecast updates is time dependent and follows a "$\cap$" pattern over the trading window: Forecast updates at long time horizons are relatively small, since they are mainly based on the seasonal trend. At shorter time scales, *i.e.* for time horizons ranging from a couple of weeks to several hours ahead of the delivery time, the forecast updates are more significant, since weather forecasts are available and steadily become more precise. Finally, the last forecast updates represent slight adjustments to ensure that the market clearing constraint is satisfied. Another example where inhomogeneous forecast updates occur is in open ended funds, where the fund manager – if notified as and when share remissions and purchases are ordered during the trading day – could pro-actively trade in the underlying assets on the trading day itself rather than on the next, so as to increase the liquidity of the fund. In this case the standard deviation of the updates on the volume target often follows a "$\cup$" profile instead. Further investigation is thus needed to analyse the impact of the forecasts uncertainty profile on the performance of our approach relative to the model of Almgren and Chriss (2001). Since the optimal redistribution coefficients depend on the profile of the standard deviation of the forecast updates, these two points should be jointly addressed in further investigations.

2.5 Conclusion

In this chapter we considered the optimal trade execution problem in a setting where price uncertainty grows in time, but, unlike what is assumed in the existing literature, the required trade volume is also uncertain and becomes only known at the end of the execution period. We assume that forecasts of increasing accuracy become available, so that volume uncertainty decreases over time, while price uncertainty increases. The model presented in this work is designed to manage both uncertainties via risk terms, so as to be pre-computable and avoid the combinatorial explosion of dynamic programming approaches. We have demonstrated that the model has desirable convexity properties, guaranteeing the existence and uniqueness of a convex set of optimal strategies. We show that it produces significantly lower transaction costs in comparison to classical trade execution approaches, even if the latter allow for recourse whenever a new forecast becomes available.

Chapter 3

Trade execution with uncertain volume target under a dynamic coherent risk measure

This chapter investigates how the uncertainty related to the volume target can be introduced via a dynamic programming approach. This chapter builds on the background material on the optimal execution problem (Section 1.1). The content of this chapter is the object of a publication currently in preparation and soon available on *arXiv* (Vaes and Hauser, b).

3.1 Introduction

On a daily basis, traders face the problem of trading a large volume of a given asset. When the liquidity of the corresponding market is limited, trading large volumes all at once is paired with a large trading cost due to the impact of the trades on the price process (Almgren and Chriss, 2001). Therefore, in order to reduce this trading cost, it is of common usage to spread this trade over time. However, by doing so, traders expose themselves to the risk associated with the randomness of the price process. To balance cost and risk, Almgren and Chriss modelled the trade execution problem from the perspective of a risk-averse trader by applying the return-risk trade-off introduced by Markowitz (1952) to the trade execution problem formulated by Bianchi and Lo (1998).

In the current literature, there exists two main approaches to integrate risk in the trade execution model. First, there are the *one-shot* models, where it is assumed that traders decide on their trading strategy at the beginning of the trading window without considering the possibility of taking any recourse actions. Different works have explored various ways to measure risk such as, for instance, the variance in the seminal work of Almgren and Chriss, the α-*Conditional Value-at-Risk* (CVaR_α) (Rockafellar and Uryasev, 2000, 2002) in Butenko et al. (2005); Feng et al. (2012); Krokhmal and Uryasev (2007); Vaes and Hauser (2022), and the *expected utility* in Schied and Schöneborn (2009); Schöneborn (2016). The second approach is to derive stochas-

tic strategies that can depend on market information as it arrives. For the continuous-time framework, Gatheral and Schied (2011) derived the closed-form formula of the unique optimal stochastic strategy when quantifying the risk of a strategy as the time-average Value-at-Risk (VaR) associated with the P&L of the position. In Schied et al. (2010), the authors show that, under specific assumptions, the optimal execution strategy is deterministic when maximising the expected-utility for exponential, or constant absolute risk aversion (CARA), utility functions. In the discrete-time setting, Lin et al. (2015) showed that the optimal strategy of the trade execution problem formulated by Almgren and Chriss is deterministic when using a dynamic coherent risk measure. Indeed, they prove that, at any stage, the optimal quantity to be traded on the subsequent trading period is only depending on the volume left to be traded. While the one-shot approach is attractive because of its simplicity of use, the solution obtained by this approach is often not time-consistent, in the sense that mid-term action could lead to better results (see *e.g.* Lorenz and Almgren (2011)). The second approach leads to time-consistent strategies which are often also more cost-effective. However, except under specific modelling assumptions, finding such strategies is computationally prohibitive.

In this work, we focus on the trade execution problem introduced by Vaes and Hauser (2022), which considers that the volume to be traded is a random variable that varies during the execution period. More precisely, we assume that a trader has access to forecasts of the volume to be traded, which become progressively more accurate throughout the trading period. A trader is thus subject to two sources of uncertainty that evolve differently with time: volume uncertainty decreases, while price uncertainty increases. In Vaes and Hauser (2022), the trade execution problem is considered in its discrete-time version and modelled as one-shot optimisation problem for which risk is assessed using the $\mathbb{E}$-CVaR$_{\alpha,\lambda}$ risk-measure. In the continuous-time framework, recent works (Bulthuis et al., 2017; Cheng et al., 2017, 2019; Mariani and Fatone, 2022) have approached the liquidation problem under execution risk. This problem can be related to the liquidation problem with uncertain volume target as discussed in Vaes and Hauser (2022). Both approaches have advantages and drawbacks: In Vaes and Hauser (2022), finding the optimal strategy is computationally tractable and easily implementable in practice. The choice of using $\mathbb{E}$-CVaR$_{\alpha,\lambda}$ as the risk measure leads to strategies that are interpretable all the while being competitive in terms of cost compared to the approaches in the continuous-time framework. However, the solution suggested in Vaes and Hauser (2022) is deterministic and incorporates only partial recourse actions via a predetermined rule. This solution not is time-consistent, and so a mid-recourse action could lead to better results. The second approach, on the other hand, guarantees the time consistency of the optimal solution but relies on specific and potentially restrictive modelling assumptions in order to be solvable or numerically tractable. These assumptions include, for example, the choice of the risk measure used, the distribution of the random process modelling the volume target updates, and the homogeneity of the trading periods. The choice of the risk measure influences greatly the trading cost. Hence, even though it leads to time-consistent strategies, the use of certain risk measures, such as the exponential util-

ity, is questionable as the cost obtained can be significantly higher than when using a one-shot approach; see Vaes and Hauser (2022, Section 3.3).

In this work, we consider the discrete-time version of the trade execution problem when subject to uncertainty in the price process and uncertainty in the volume to be traded. Our main contribution is to show that, under commonly accepted assumptions, a stochastic and time-consistent strategy can be found via the dynamic programming approach. We use as risk measure the dynamic coherent risk measure induced by the one-step conditional risk measure $\mathbb{E}$-CVaR$_{\alpha,\lambda}$, which is the coherent risk measure that corresponds to the weighted sum of the expectation and CVaR$_\alpha$. Similarly to Lin et al. (2015) for the trade execution problem under exclusively price uncertainty, we show that, at any stage, the optimal quantity to be traded on the next trading period is only function of the volume left to be traded and is thus independent of the price process. The size of the scenario tree can therefore be significantly reduced, and the explosion in computational complexity when the size of the problem increases (*e.g.* number of trading periods), which often plagues dynamic programming approaches, is avoided. Our model circumvents the shortcomings previously raised: it is significantly more versatile compared to the continuous-time models which relies on convenient modelling assumptions, is tractable and provides time-consistent optimal strategies.

This rest of this chapter is structured as follows. In Section 3.2, we introduce the set-up in which we are working. In Section 3.3, we derive the dynamic programming equations related to our problem and show that the optimal solution is time-consistent. Section 3.4 contains the major contribution of our work: we show that the size of the scenario tree can be reduced. To this end, we first show that the decision to take at any state is only function of the best estimate on the volume left to be traded, and then we prove that a trader has not interest in over-trading, which enables a reduction of the action space at each state. In Section 3.5, we explain how the problem specificities can be leveraged in order to efficiently solve the state optimisation problem by calculating the value function using a recursive approach. Finally, Sections 3.6 and 3.7 are reserved for numerical results and conclusions.

3.2 Framework

We consider the model introduced in Section 1.1, where we assumed that the trading window spans the time interval $[\,t_s, t_f\,]$ with $t_s < t_f$, and that it is subdivided into m trading periods. As mentioned in the introduction, the market is subject to the uncertainty related to the price dynamics and to the volume to be traded, for which the estimate becomes increasingly precise over time. We will make assumptions similar to those in Chapter 2 regarding these two sources of uncertainty.

This section is divided as follows: in Sections 3.2.1 to 3.2.3, we present the assumptions related to the market and the uncertainty structure related to the problem. In Section 3.2.4, we define what is a trading strategy and its associated cost. Finally, in Section 3.2.5, we construct

the dynamic risk measure used to quantify the risk associated to a strategy.

3.2.1 Price dynamics

Let S_0 be the initial price of the security, and let $\{S_k, k \in [\![0, m]\!]\}$ be the stochastic process that represents the fair price dynamics. The price process is assumed to be a random walk where the random step or *price shift* between two successive decision times t_{k-1} and t_k is modelled by a random variable ξ_k, *i.e.* $\forall k \in [\![1, m]\!] : S_k = S_{k-1} + \xi_k$. Furthermore, we assume that the market has limited liquidity and that trades induce linear temporary and permanent market impacts on the price dynamics. The price dynamics is then formulated as follows (see Section 1.1.2 for more details):

$$S_k = S_{k-1} + \xi_k + \gamma_k q_k, \tag{3.1a}$$

$$\widetilde{S}_k = S_{k-1} + \epsilon_k \operatorname{sgn}\left(\frac{q_k}{\tau_k}\right) + \frac{\eta_k}{\tau_k} q_k, \tag{3.1b}$$

where S_k is the fair price at decision time t_k and $\widetilde{S}_k$ is the effective security price for the trades executed during the trading period τ_k. In addition to these modelling assumptions, we make the following assumptions:

Assumption 3.1. *The market is free of price manipulation.*

Assumption 3.2. *For all* $k \in [\![1, m-1]\!]$, *it holds that* $0 < \epsilon_k \leq \epsilon_{k+1}$ *and* $0 < \frac{\eta_k}{\tau_k} \leq \frac{\eta_{k+1}}{\tau_{k+1}}$.

Assumption 3.3. *Trades do not have a permanent impact on the price process,* i.e. $\forall k \in [\![1, m]\!] : \gamma_k = 0$.

Assumption 3.1 is natural, otherwise the market would not be viable as traders would take advantage of the possible price manipulations; see the implications of this assumption according to Lemma 1.1. Assumption 3.2 means that either the liquidity is stable or it dries up over time. As a consequence, a risk-averse trader who must trade one position will prefer to trade it as soon as possible rather than delaying the trade. Assumption 3.3 is more restrictive but can be justified in some situations. One of these situations is when trading on the power market. Indeed, for every given delivery period, the market participants can trade power contracts from several years to a few seconds ahead of delivery. These trades are made on markets having different granularities: bilateral trades (years/months ahead), day-ahead market (day ahead), intra-day market (hours ahead) and real-time market (minutes/seconds ahead). Assumption 3.2 means that the closer we are to delivery time, the less liquidity in the market there is. For instance, power negotiated by future contracts several months ahead of maturity will be produced by nuclear generators, which typically have small marginal cost. It is however realistic to assume that these trades are independent of the day-ahead market for which the generation of power is mainly taken in charge by more flexible technologies such as natural gas turbines; modelling a permanent impact of trades is thus not necessarily justified when trading on such different granularities.

3.2.2 Volume target dynamics

We denote by D_0 the initial estimate of the volume D_{t_f} to be acquired by time t_f. The volume target at time t_k, *i.e.* the best estimate of the volume to be traded by time t_f, is theoretically given by $D_k \coloneqq \mathbb{E}\Big[D_{t_f} \;\Big|\; \mathcal{F}_k\Big]$, where $(\mathcal{F}_k)_{k=0}^{m-1}$ is the filtration of sigma algebras that represent the information available at time t_k. Here, we assume that the volume target dynamics can be represented as a stochastic process $\{D_k, k \in [\![0, m]\!]\}$ that follows a random walk, where for $k \in [\![1, m-1]\!]$ the random step between decision times t_{k-1} and t_k is modelled by the random variable δ_k, which can be interpreted as the volume forecast update. Moreover, like in Chapter 2, we assume that the volume target is perfectly known at decision time t_{m-1}, *i.e.* $\delta_m = 0$. The volume target dynamics is therefore given by

$$D_k = D_{k-1} + \delta_k, \text{ if } k \in [\![1, m]\!] \tag{3.2}$$

where $D_m = D_{t_f}$. Since $\delta_m = 0$, it implies that $D_{m-1} = D_m = D_{t_f}$. To allow for the application of the dynamic programming approach, we make the following assumption:

Assumption 3.4. *For all $k \in [\![1, m]\!]$, the random variables representing both the price shift ξ_k and the forecast update δ_k are independent of $\xi_{k'}$ and $\delta_{k'}$ for all $k' \neq k$.*

3.2.3 Uncertainty structure

We adopt here the notation and the structure suggested in Shapiro (2009): We assume that the random variables representing the price shifts and the forecast updates on the volume to be traded follow a discrete distribution with a finite number of outcomes. Therefore, the evolution of the corresponding random processes can be represented as an m-stage scenario tree. At the initial trading time t_s, and thus at the initial stage $k = 0$, there is a single node, which is the *root node* that is denoted by ζ_0. At the second stage $k = 1$, the number of nodes corresponds to the number of possible realisations, each one of them is connected to the parent node (here the root node) by an arc; and so forth for the subsequent stages. We denote by ζ_k a generic node at stage k (*i.e.* at decision time t_k), and by Ω_k the set of all possible nodes at stage $k \in [\![0, m-1]\!]$. A *scenario* is a particular realisation or *path* of the data process and corresponds to a sequence $\zeta_0, \ldots, \zeta_{m-1}$, where $\zeta_{k-1} \in \Omega_{k-1}$, $\zeta_k \in \Omega_k$ and ζ_k is a child node of ζ_{k-1}. For any $0 \leq i \leq j \leq m-1$, $\boldsymbol{\zeta}_{[\![i,j]\!]} \coloneqq \{\zeta_i, \ldots, \zeta_j\}$ denotes the realisation path from ζ_i to ζ_j. In particular, $\boldsymbol{\zeta}_{[\![0,k]\!]}$ represents the history of the processes from the root node ζ_0 to ζ_k.

As mentioned previously, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ associated to this random process, and a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_{m-1} \subset \mathcal{F}$, where $\mathcal{F}_k$ is the sigma algebra that represents the information available at time t_k, or stage k. More precisely, $\mathcal{F}_{m-1}$ is the sigma algebra formed of all subsets of Ω_{m-1}, for any $k \in [\![0, m-2]\!]$, $\mathcal{F}_k$ is the sigma subalgebra of $\mathcal{F}_{k+1}$ generated by sets of Ω_{k+1} which are children nodes of nodes at stage k, and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. We also define $\mathcal{Z}_k \coloneqq \mathcal{L}_p(\Omega_{m-1}, \mathcal{F}_k, \mathbb{P})$ as the spaces of random variables on $(\Omega_{m-1}, \mathcal{F}_k, \mathbb{P})$ with finite p-th moment, $p \in [0, \infty]$, $k \in [\![0, m-1]\!]$, and $\mathcal{Z}_{[\![k,m-1]\!]} \coloneqq \mathcal{Z}_k \times \cdots \times \mathcal{Z}_{m-1}$.

3.2.4 Trading decisions and trading cost

At each stage $k \in [\![0, m-1]\!]$, the decision variables are denoted by a_k; in our case these correspond to the volume to be traded on the next trading period τ_{k+1}, *i.e.* $a_k \coloneqq q_{k+1}$. The lower case of q_{k+1} emphasises the fact that the quantity to be traded during the $(k+1)$-th trading period is deterministic at stage k; it is however a random variable at the initial state ζ_0. To apply dynamic programming, it is required that the decision variables satisfy the following two criteria:

- *nonanticipativity*: the decision variables a_k can only depend on the current available information at ζ_k. The decision variables at stage k can consequently only depend on the history of the data process and the previous decisions but must be independent of future observations. Mathematically, we consequently have that a_k should only be a function of $\boldsymbol{\zeta}_{[\![0,k]\!]}$ and $\boldsymbol{a}_{[\![0,k-1]\!]} \coloneqq \{a_0, \dots, a_{k-1}\}$.
- *time-consistency*: at each state $\zeta_k \in \Omega_k$, $k \in [\![0, m-1]\!]$, the optimal decisions should not depend on states that do not descend from the current node ζ_k.

For a node ζ_k at any stage $k \in [\![0, m-1]\!]$, we denote by $\mathcal{A}_k\left(\zeta_k, \boldsymbol{a}_{[\![0,k-1]\!]}\right)$ the action space, which corresponds to the set of possible actions or decisions; to lighten the notations, $\mathcal{A}_k$ is used as a shorthand for $\mathcal{A}_k\left(\zeta_k, \boldsymbol{a}_{[\![0,k-1]\!]}\right)$ when there is no risk of confusion. This set potentially depends on the current state ζ_k as well as on the past decisions, *i.e.* $\boldsymbol{a}_{[\![0,k-1]\!]}$. In the case of the trade execution problem of interest here, the only restriction is the market clearing constraint, which requires the cumulative volume traded to be equal to D_{t_f}. Therefore, we have that $\forall k \in [\![0, m-2]\!] : \mathcal{A}_k = \mathbb{R}$ and $\mathcal{A}_{m-1} = \left\{D_{t_f} - \sum_{k=1}^{m-1} q_k\right\}$. Hence, $\mathcal{A}_k$ is closed and convex for all $k \in [\![0, m-1]\!]$. A *trading strategy* starting at stage k is denoted by $\boldsymbol{\Delta}_{[\![k,m]\!]}$ and corresponds to a set of decision variables for each stage $i \in [\![k, m-1]\!]$, *i.e.* $\boldsymbol{\Delta}_{[\![k,m]\!]} \coloneqq (a_k, a_{k+2}, \dots, a_{m-1}) = (q_{k+1}, Q_{k+1}, \dots, Q_m)$. Note that, at a state ζ_k, the action taken is deterministic whereas the actions of stages posterior to stage k, *i.e.* any stage i with $i \in [\![k+1, m-1]\!]$, depend on the realisations of the data process and must be considered as random variables at stage k. The decision process can be modelled as in Figure 3.1.

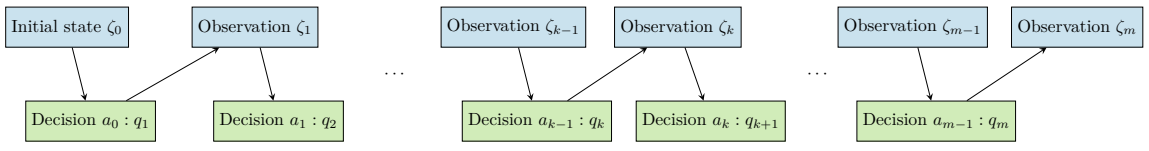


FIGURE 3.1: Information and decision processes

For any stage $k \in [\![0, m-1]\!]$, the *stage cost* is a function $\mathcal{C}_k : \Omega_k \times \mathcal{A}_k \to \mathcal{Z}_{k+1}; (\zeta_k, a_k) \mapsto \mathcal{C}_k(\zeta_k, a_k)$ which we define for $k \in [\![0, m-1]\!]$ as follows:

$$\begin{aligned}\mathcal{C}_k(\zeta_k, a_k) &\coloneqq \widetilde{S}_{k+1}\, q_{k+1} - \delta_{k+1} S_{k+1} \\ &= \epsilon_{k+1}\, |q_{k+1}| + \left(S_k + \frac{\eta_{k+1}}{\tau_{k+1}}\, q_{k+1}\right) q_{k+1} - \delta_{k+1}(S_k + \xi_{k+1})\,. \end{aligned} \tag{3.3}$$

Since $\delta_m = 0$, we have that the stage cost at stage $m-1$ is given by

$$\mathcal{C}_{m-1}(\zeta_{m-1}, a_{m-1}) \coloneqq \widetilde{S}_m \, q_m = \epsilon_m \, |q_m| + \left(S_{m-1} + \frac{\eta_m}{\tau_m} \, q_m \right) q_m. \tag{3.4}$$

In the following, we use $\mathcal{C}_k$ as a shorthand for $\mathcal{C}_k(\zeta_k, a_k)$ with $k \in [\![0, m-1]\!]$. The cumulative cost $\mathcal{C}_k^{cum}$ at a state ζ_k (and thus given S_k) of a strategy $\mathbf{\Delta}_{[\![k,m]\!]}$ is defined as the sum of future stage costs, *i.e.*

$$\begin{aligned} \mathcal{C}_k^{cum}\Big(\mathbf{\Delta}_{[\![k,m]\!]} \,\Big|\, \zeta_k \Big) &\coloneqq \sum_{i=k}^{m-1} \mathcal{C}_i = \sum_{i=k}^{m-1} \widetilde{S}_{i+1} \, Q_{i+1} - \sum_{i=k}^{m-2} \delta_{i+1} S_{i+1} \\ &= \sum_{i=k}^{m-1} \epsilon_{i+1} \, |Q_{i+1}| + \sum_{i=k}^{m-1} \left(S_i + \frac{\eta_{i+1}}{\tau_{i+1}} \, Q_{i+1} \right) Q_{i+1} - \sum_{i=k}^{m-2} \delta_{i+1} (S_i + \xi_{i+1}) \, . \end{aligned} \tag{3.5}$$

Consequently, the cumulative trading cost at the initial stage given a strategy $\mathbf{\Delta}_{[\![0,m]\!]}$ amounts to $\mathcal{C}_0^{cum}\Big(\mathbf{\Delta}_{[\![0,m]\!]} \Big) = \sum_{k=1}^{m} \widetilde{S}_k \, Q_k - \delta_k S_k$; we have omitted to specify ζ_0 as it is the only state possible at the initial stage.

REMARK 3.1. *Since there is no update on the volume target during the last trading period, we have that* $\mathcal{C}_0^{cum}\Big(\mathbf{\Delta}_{[\![0,m]\!]} \Big) \in \mathcal{Z}_{m-1}$. *Moreover,* $\forall k \in [\![0, m-2]\!] : \mathcal{C}_k \in \mathcal{Z}_{k+1}$ *and* $\mathcal{C}_{m-1} \in \mathcal{Z}_{m-1}$.

REMARK 3.2. *If we denote by* $\mathcal{C}_0^{eff}\Big(\mathbf{\Delta}_{[\![0,m]\!]} \Big) \coloneqq \sum_{k=1}^{m} \widetilde{S}_k Q_k$ *the effective cost of a strategy* $\mathbf{\Delta}_{[\![0,m]\!]}$, *which corresponds to the actual amount of money disbursed by the trader to acquire the volume* D_{t_f}, *we observe a discrepancy between the cumulative cost and the effective cost. Indeed,*

$$\mathcal{C}_0^{cum}\Big(\mathbf{\Delta}_{[\![0,m]\!]} \Big) = \mathcal{C}_0^{eff}\Big(\mathbf{\Delta}_{[\![0,m]\!]} \Big) - \sum_{k=1}^{m} \delta_k S_k. \tag{3.6}$$

In other words, the cumulative cost is equal to the effective cost up to the random variable $\sum_{k=1}^{m} \delta_k S_k$. *Therefore, minimising the risk associated with the cumulative cost is not equivalent to minimising the risk for the effective cost. The latter problem can be seen as minimising the risk associated with the part of the effective cost over which the trader has an influence. Our motivation for addressing the former problem is that, as shown in Section 3.4 and Section 3.6, using the cumulative cost enables a significant reduction in the size of the dynamic programming scenario tree (see Section 3.4), while providing strategies that are competitive compared to other approaches (see Section 3.6).*

3.2.5 Risk minimisation

In order to define the optimal decision at every node $\zeta_k \in \Omega_k$, $k \in [\![0, m-1]\!]$, we need to specify the optimisation problem that has to be solved by the trader. For generality, we assume that, at every stage $k \in [\![0, m-1]\!]$, the objective function for future optimisation starting from ζ_k is given by $\mathbb{F}_k : \mathcal{Z}_{[\![k,m-1]\!]} \to \mathcal{Z}_k$. More precisely, the optimisation problem to be solved at stage k is given by

$$V_k\Big(\zeta_k, \boldsymbol{a}_{[\![0,k-1]\!]}\Big) \coloneqq \underset{a_k,\dots,a_{m-1}}{\text{minimise}} \quad \mathbb{F}_k\left[0, \mathcal{C}_k, \mathcal{C}_{k+1}, \dots, \mathcal{C}_{m-3}, \mathcal{C}_{m-2} + \mathcal{C}_{m-1}\right], \tag{3.7a}$$

$$\text{subject to} \quad \forall j \in [\![k, m-1]\!] : a_j \in \mathcal{A}_j\Big(\zeta_j, \boldsymbol{a}_{[\![0,j-1]\!]}\Big), \tag{3.7b}$$

where at $j = 0$, $\mathcal{A}_j\Big(\zeta_j, \boldsymbol{a}_{[\![0,j-1]\!]}\Big)$ must be understood as $\mathcal{A}_0$, the initial feasible set of the first stage decision variables a_0. The first argument corresponds to the part of the cost that is $\mathcal{F}_k$-measurable. We mention this argument to be consistent with the definition of a dynamic risk measure provided later on and set it to 0 since the stage cost $\mathcal{C}_k(\zeta_k, a_k)$ is not $\mathcal{F}_k$-measurable but rather $\mathcal{F}_{k+1}$-measurable. The optimal value of (3.7) given the past decisions $\boldsymbol{a}_{[\![0,k-1]\!]}$ and current state ζ_k is called *value function* and is denoted by $V_k\Big(\zeta_k, \boldsymbol{a}_{[\![0,k-1]\!]}\Big)$. At the initial stage ζ_0, the trader solves the following optimisation problem:

$$V_0(\zeta_0) \coloneqq \underset{a_0,\dots,a_{m-1}}{\text{minimise}} \quad \mathbb{F}_0\left[0, \mathcal{C}_0 \dots, \mathcal{C}_{m-3}, \mathcal{C}_{m-2} + \mathcal{C}_{m-1}\right], \tag{3.8a}$$

$$\text{subject to} \quad \forall j \in [\![0, m-1]\!] : a_j \in \mathcal{A}_j\Big(\zeta_j, \boldsymbol{a}_{[\![0,j-1]\!]}\Big). \tag{3.8b}$$

As previously mentioned, we assume that the trader is averse to risk and therefore seeks the strategy that minimises the risk related to their cumulative cost; the functions $\mathbb{F}_k$, $k \in [\![0, m-1]\!]$, should therefore translate their preference over uncertain sequence of future stage costs. To do so, we assume that the trader measures risk with a *time consistent dynamic coherent risk measure*. The notion of *dynamic risk measure* has been introduced in Ruszczyński (2010) and is built on the notion of *conditional risk measure*:

Definition 3.1 (Conditional Risk Measure (Ruszczyński, 2010, Definition 1)). *A mapping $\varrho_{[\![k,m-1]\!]} : \mathcal{Z}_{[\![k,m-1]\!]} \to \mathcal{Z}_k$, with $k \in [\![0, m-1]\!]$, is called a* conditional risk measure *if it has the following monotonicity property:*

$$\varrho_{[\![k,m-1]\!]}[Z] \leq \varrho_{[\![k,m-1]\!]}[W] \;\; \textit{for all } Z, W \in \mathcal{Z}_{[\![k,m-1]\!]} \textit{ such that } \; Z \leq W, \tag{3.9}$$

where the inequality $Z \leq W$ should be interpreted componentwise and in the almost sure sense.

The conditional risk measure $\varrho_{[\![k,m-1]\!]}[Z_k, \dots, Z_{m-1}]$ can be interpreted as a fair $\mathcal{F}_k$-measurable cost the trader would be willing to pay to avoid the uncertain future stage costs $Z_k, \dots, Z_{m-1}$.

Definition 3.2 (Dynamic risk measure Ruszczyński (2010, Definition 2)). *A* dynamic risk measure *is a sequence of conditional risk measures $\varrho_{[\![k,m-1]\!]} : \mathcal{Z}_{[\![k,m-1]\!]} \to \mathcal{Z}_k$, for any $k \in [\![0, m-1]\!]$.*

The objective of a *dynamic risk measure* (DRM) is to be able to assess the risk associated to a random process over time. In Definition 3.2, no link exists between two successive conditional risk measures, which could potentially lead to incoherent risk evaluations over time. To solve this issue, Ruszczyński (2010) suggested based on the work of Cheridito et al. (2006, Proposition 4.4) that a dynamic risk measure should satisfy an additional property called *time-consistency*:

Definition 3.3 (Time consistency Ruszczyński (2010, Definition 3)). *A dynamic risk measure is* time consistent *if for all* $0 \leq \tau < \theta \leq m-1$ *and all sequences* $Z, W \in \mathcal{Z}_{[\![\tau, m-1]\!]}$*, the conditions*

$$Z_k = W_k,\ k = \tau, \ldots, \theta - 1 \tag{3.10}$$

and

$$\varrho_{[\![\theta, m-1]\!]}[Z_\theta, \ldots, Z_{m-1}] \leq \varrho_{[\![\theta, m-1]\!]}[W_\theta, \ldots, W_{m-1}] \tag{3.11}$$

imply that

$$\varrho_{[\![\tau, m-1]\!]}[Z_\tau, \ldots, Z_{m-1}] \leq \varrho_{[\![\tau, m-1]\!]}[W_\tau, \ldots, W_{m-1}]\,. \tag{3.12}$$

In words, the time-consistency of a dynamic risk measure $\left\{\varrho_{[\![k, m-1]\!]}\right\}_{k=0}^{m-1}$ implies that, if a trader prefers Z to W at a future stage θ, and if Z and W are equivalent from τ until θ, then Z should either be preferred or be equivalent to W at time τ.

Ruszczyński (2010) showed that a time-consistent dynamic risk measure can be completely defined by a sequence $\{\varrho_k\}_{k=0}^{m-1}$ of one-step conditional risk measures, where, for all $k \in [\![0, m-1]\!]$, ϱ_k is a function from $\mathcal{Z}_{k+1}$ to $\mathcal{Z}_k$. In this vein, similarly to Lin et al. (2015), we construct a dynamic risk measure based on a sequence $\{\varrho_k\}_{k=0}^{m-1}$ of coherent risk measures. For any $k \in [\![0, m-1]\!]$, a risk measure $\varrho_k : \mathcal{Z}_{k+1} \to \mathcal{Z}_k$ is said to be *coherent* if it satisfies the following four axioms for all $Z_{k+1}, W_{k+1} \in \mathcal{Z}_{k+1}$ (Artzner et al., 1999):

Axiom 3.1 (Convexity).

$$\forall \lambda \in [0,1]: \quad \varrho_k[\lambda Z_{k+1} + (1-\lambda) W_{k+1}] \leq \lambda\, \varrho_k[Z_{k+1}] + (1-\lambda)\, \varrho_k[W_{k+1}]\,. \tag{3.13}$$

Axiom 3.2 (Translation invariance).

$$\forall X_k \in \mathcal{Z}_k: \quad \varrho_k[Z_{k+1} + X_k] = \varrho_k[Z_{k+1}] + X_k. \tag{3.14}$$

Axiom 3.3 (Positive homogeneity).

$$\forall \lambda \geq 0: \quad \varrho_k[\lambda Z_{k+1}] = \lambda\, \varrho_k[Z_{k+1}]\,. \tag{3.15}$$

Axiom 3.4 (Monotocity).

$$Z_{k+1} \leq W_{k+1} \Rightarrow \varrho_k[Z_{k+1}] \leq \varrho_k[W_{k+1}]\,. \tag{3.16}$$

Given a sequence $\{\varrho_k\}_{k=0}^{m-1}$ of coherent risk measures, for any $k \in [\![0, m-1]\!]$, the induced *conditional risk measures* $\varrho_{[\![k, m-1]\!]} : \mathcal{Z}_{[\![k, m-1]\!]} \to \mathcal{Z}_k$ is defined as the nested composition of the coherent risk measures $\{\varrho_k\}_{k=0}^{m-1}$, *i.e.*

$$\begin{aligned} &\varrho_{[\![k, m-1]\!]}[Z_k, \ldots, Z_{m-1}] \\ &\qquad = \varrho_k\big[\varrho_{k+1}\big[\cdots \varrho_{m-2}\big[\varrho_{m-1}[Z_k + Z_{k+1} + \cdots + Z_{m-2} + Z_{m-1}]\big] \cdots\big]\big]\,. \end{aligned} \tag{3.17}$$

For any $k \in [\![0, m-2]\!]$, $\varrho_{[\![k, m-1]\!]}$ defined in (3.17) is indeed a conditional risk measure: for any $Z, W \in \mathcal{Z}_{[\![k, m-1]\!]}$ such that $Z \leq W$, we have that $\sum_{k=i}^{m-1} Z_k \leq \sum_{k=i}^{m-1} W_k$, hence (3.9) holds true

based on (3.17) and Axiom 3.4. Consequently, the sequence $\left\{\varrho_{[\![k,m-1]\!]}\right\}_{k=0}^{m-1}$ with $\varrho_{[\![k,m-1]\!]}$ defined as in (3.17) is indeed a dynamic risk measure. By a slight abuse of notation, we also denote by $\varrho_{[\![k,m-1]\!]} : \mathcal{Z}_{m-1} \to \mathcal{Z}_k$ the function defined as follows:

$$\varrho_{[\![k,m-1]\!]}[Z_{m-1}] = \varrho_k\big[\varrho_{k+1}\big[\cdots \varrho_{m-2}\big[\varrho_{m-1}[Z_{m-1}]\big]\cdots\big]\big]\,. \tag{3.18}$$

The benefit of using coherent risk measures as one-step conditional risk measures is that the induced dynamic risk measure is time-consistent as shown in Proposition 3.1.

PROPOSITION 3.1. *Let $\{\varrho_k\}_{k=0}^{m-1}$ be a sequence of coherent risk measures. Then the dynamic risk measure $\left\{\varrho_{[\![k,m-1]\!]}\right\}_{k=0}^{m-1}$ induced by $\{\varrho_k\}_{k=0}^{m-1}$, where $\varrho_{[\![k,m-1]\!]}$ is defined as in (3.17), is time-consistent.*

PROOF. Using the definition (3.17) and the axioms that a coherent risk measure must satisfy, we have that the following equations hold $\forall k \in [\![0, m-1]\!]$, $0 \le \tau \le \theta \le m-1$ and $Z \in \mathcal{Z}_{[\![0,m-1]\!]}$:

$$\begin{aligned}
&\varrho_{[\![k,m-1]\!]}[Z_k, \ldots, Z_{m-1}] \\
&\quad= \varrho_k\big[\varrho_{k+1}\big[\cdots \varrho_{m-2}\big[\varrho_{m-1}[Z_k + Z_{k+1} + \cdots + Z_{m-2} + Z_{m-1}]\big]\cdots\big]\big] \\
&\quad= \varrho_k\big[\varrho_{k+1}\big[\cdots \varrho_{m-2}\big[Z_k + \varrho_{m-1}[Z_{k+1} + \cdots + Z_{m-2} + Z_{m-1}]\big]\cdots\big]\big] \\
&\quad= \ldots \\
&\quad= Z_k + \varrho_k\big[\varrho_{k+1}\big[\cdots \varrho_{m-2}\big[\varrho_{m-1}[Z_{k+1} + \cdots + Z_{m-2} + Z_{m-1}]\big]\cdots\big]\big] \\
&\quad= Z_k + \varrho_{[\![k,m-1]\!]}[0, Z_{k+1}, \ldots, Z_{m-1}]\,,
\end{aligned} \tag{3.19}$$

$$\begin{aligned}
&\varrho_{[\![k,m-1]\!]}[0, 0, \ldots, 0] \\
&\quad= \varrho_k\big[\varrho_{k+1}\big[\cdots \varrho_{m-2}\big[\varrho_{m-1}[0 + 0 + \cdots + 0]\big]\cdots\big]\big] = 0,
\end{aligned} \tag{3.20}$$

$$\begin{aligned}
&\varrho_{[\![\tau,m-1]\!]}[Z_\tau, \ldots, Z_\theta, \ldots, Z_{m-1}] \\
&\quad= \varrho_\tau\big[\cdots \varrho_{\theta-1}\big[\varrho_\theta\big[\cdots \varrho_{m-1}[Z_\tau + \cdots + Z_{\theta-1} + Z_\theta + \cdots + Z_{m-1}]\cdots\big]\big]\cdots\big] \\
&\quad= \varrho_\tau\big[\cdots \varrho_{\theta-1}\big[Z_\tau + \cdots + Z_{\theta-1} + \varrho_\theta\big[\ldots \varrho_{m-1}[Z_\theta + \cdots + Z_{m-1}]\cdots\big]\big]\cdots\big] \\
&\quad= \varrho_\tau\Big[\cdots \varrho_{\theta-1}\Big[Z_\tau + \cdots + Z_{\theta-1} + \varrho_{[\![\theta,m-1]\!]}[Z_\theta, \ldots, Z_{m-2}]\Big]\cdots\Big] \\
&\quad= \varrho_{[\![\tau,\theta]\!]}\Big[Z_\tau, \ldots, Z_{\theta-1}, \varrho_{[\![\theta,m-1]\!]}[Z_\theta, \ldots, Z_{m-1}]\Big]\,.
\end{aligned} \tag{3.21}$$

By Ruszczyński (2010, Theorem 1), Equations (3.19) to (3.21) are sufficient conditions for the dynamic risk measure $\left\{\varrho_{[\![k,m-1]\!]}\right\}_{k=0}^{m-1}$ induced by $\{\varrho_k\}_{k=0}^{m-1}$ to be time-consistent. □

To exploit this property, we assume in this work that the trader evaluates risk of future costs with the dynamic risk measure induced by $\{\mathbb{E}\text{-CVaR}_{\alpha_k,\lambda_k}\}_{k=0}^{m-1}$. For any $k \in [\![0, m-1]\!]$, $\mathbb{E}\text{-CVaR}_{\alpha_k,\lambda_k} : \mathcal{Z}_{k+1} \to \mathcal{Z}_k; Z_{k+1} \mapsto \mathbb{E}\text{-CVaR}_{\alpha_k,\lambda_k}[Z_{k+1}]$ is a coherent risk measure that corresponds to the weighted sum of the expectation and the so-called α_k-*Conditional Value-at-Risk* (CVaR_{α_k}) and is defined as follows

$$\mathbb{E}\text{-CVaR}_{\alpha_k,\lambda_k}[Z_{k+1}] := (1-\lambda_k)\,\mathbb{E}[\,Z_{k+1} \mid \mathcal{F}_k] + \lambda_k\,\text{CVaR}_{\alpha_k}[\,Z_{k+1} \mid \mathcal{F}_k]\,, \tag{3.22}$$

where $\mathrm{CVaR}_{\alpha_k} : \mathcal{Z}_{k+1} \to \mathcal{Z}_k$ is the coherent risk measure defined as follows (Rockafellar and Uryasev, 2000, 2002) :

$$\mathrm{CVaR}_{\alpha_k}[Z_{k+1} \mid \mathcal{F}_k] = \inf_{W \in \mathcal{Z}_k} \left\{ W + \frac{1}{\alpha_k} \mathbb{E}\left[(Z_{k+1} - W)_+ \,\middle|\, \mathcal{F}_k \right] \right\}. \tag{3.23}$$

For each stage $k \in [\![0, m-1]\!]$, the risk measure $\mathbb{E}\text{-}\mathrm{CVaR}_{\alpha_k,\lambda_k}$ relies on two parameters: First $\lambda_k \in [0,1]$, which represents the risk-aversion of the trader at stage k: the greater λ_k, the more risk-averse the trader. Then α_k parametrises the risk-level. Indeed, CVaR_{α_k} measures risk by focusing on the proportion α_k of extreme costs and can be interpreted as the expectation of the costs conditional on them exceeding the $1-\alpha_k$ quantile. In the following, we use ϱ_k as a shorthand for $\mathbb{E}\text{-}\mathrm{CVaR}_{\alpha_k,\lambda_k}$, and define $\varrho_{[\![k,m-1]\!]}$ as in Equation (3.17). If for each $k = 0, \dots m-1$, we define the objective function $\mathbb{F}_k : \mathcal{Z}_{[\![k,m-1]\!]} \to \mathcal{Z}_k$ of Optimisation Problem (3.7) as follows:

$$\begin{aligned} \mathbb{F}_k\left[Z_k, Z_{k+1}, \dots, Z_{m-1}\right] &\coloneqq \varrho_{[\![k,m-1]\!]}[Z_k, Z_{k+1}, \dots, Z_{m-1}] && (3.24)\\ &= \varrho_k\left[\varrho_{k+1}\left[\cdots \varrho_{m-2}\left[\varrho_{m-1}[Z_k + Z_{k+1} + \cdots + Z_{m-1}]\right]\cdots\right]\right], && (3.25) \end{aligned}$$

then $\mathbb{F}_k$ is a conditional risk measure and $\{\mathbb{F}_k\}_{k=0}^{m-1}$ defines a time consistent dynamic risk measure. Consequently, at any stage ζ_k with $k \in [\![0, m-1]\!]$, a trader willing to minimise the risk associated to the cumulative cost would solve the following optimisation problem:

$$\begin{aligned} V_k\left(\zeta_k, \boldsymbol{a}_{[\![0,k-1]\!]}\right) \coloneqq \underset{a_k, \dots, a_{m-1}}{\text{minimise}} \quad & \varrho_k\left[\mathcal{C}_k + \varrho_{k+1}\left[\mathcal{C}_{k+1} + \dots \varrho_{m-1}[\mathcal{C}_{m-1}]\right]\right], && (3.26a)\\ \text{subject to} \quad & \forall j \in [\![0, m-1]\!] : a_j \in \mathcal{A}_j\left(\zeta_j, \boldsymbol{a}_{[\![0,j-1]\!]}\right), && (3.26b) \end{aligned}$$

where $\mathcal{C}_i$ should be understood as depending on the current stage and the decision variables, *i.e.* $\mathcal{C}_i(\zeta_i, a_i)$.

3.3 Dynamic programming equations

Based on the special structure of the optimisation problem (3.26), we can simplify its resolution by using the value iteration algorithm, which relies on the following proposition.

Proposition 3.2. *At each stage $k \in [\![0, m-2]\!]$, the optimal value of (3.26) given the past decisions $\boldsymbol{a}_{[\![0,k-1]\!]}$ and the current node ζ_k can be recursively obtained as follows: for any stage $k \in [\![0, m-2]\!]$:*

$$V_k\left(\zeta_k, \boldsymbol{a}_{[\![0,k-1]\!]}\right) = \min_{a_k \in \mathcal{A}_k\left(\zeta_k, \boldsymbol{a}_{[\![0,k-1]\!]}\right)} \varrho_k\left[\mathcal{C}_k(\zeta_k, a_k) + V_{k+1}\left(\zeta_{k+1}, \boldsymbol{a}_{[\![0,k]\!]}\right)\right], \tag{3.27}$$

and

$$V_{m-1}\left(\zeta_{m-1}, \boldsymbol{a}_{[\![0,m-2]\!]}\right) = \min_{a_{m-1} \in \mathcal{A}_{m-1}\left(\zeta_{m-1}, \boldsymbol{a}_{[\![0,m-2]\!]}\right)} \mathcal{C}_{m-1}(\zeta_{m-1}, a_{m-1}). \tag{3.28}$$

Proof. First, $\varrho_{m-1}[\mathcal{C}_{m-1}] = \mathcal{C}_{m-1}$ using Remark 3.1 and Axiom 3.2. Hence, Equation (3.28) is equivalent to (3.26) for $k = m-1$. Second, for all $k \in [\![0, m-2]\!]$, $\mathcal{C}_k(\zeta_k, a_k)$ is independent of the decision variables $a_{k+1}, \dots, a_{m-1}$, therefore the following equations hold true:

$$V_k\left(\zeta_k, \boldsymbol{a}_{[\![0,k-1]\!]}\right) = \min_{a_k, a_{k+1}, \dots, a_{m-1}} \left\{\varrho_k\left[\mathcal{C}_k(\zeta_k, a_k) + \varrho_{k+1}\left[\mathcal{C}_{k+1} + \dots \varrho_{m-1}[\mathcal{C}_{m-1}]\right]\right]\right\}$$

$$= \min_{a_k} \left\{ \varrho_k \left[\mathcal{C}_k(\zeta_k, a_k) + \min_{a_{k+1},\dots,a_{m-1}} \left\{ \varrho_{k+1} \left[\mathcal{C}_{k+1} + \dots \varrho_{m-1}[\mathcal{C}_{m-1}] \right] \right\} \right] \right\}$$
$$= \min_{a_k} \left\{ \varrho_k \left[\mathcal{C}_k(\zeta_k, a_k) + V_{k+1}\left(\zeta_{k+1}, \boldsymbol{a}_{[\![0,k]\!]}\right) \right] \right\}.$$

This concludes the proof. □

As a consequence of Proposition 3.2, Optimisation problem (3.26) can be solved by iteratively computing the dynamic programming Equations (3.27) and (3.28). As shown in Shapiro (2009), Proposition 3.2 implies the time consistency of the optimal solution, which is defined as follows:

Definition 3.4 (Time consistency (Shapiro, 2009, A1)). *Let* $0 \leq i_1 \leq i_2 \leq m-1$. *If* $a_j^*\left(\boldsymbol{\zeta}_{[\![i_1,j]\!]}\right)$, $j \in [\![i_1, m-1]\!]$, *is an optimal solution of* (3.7) *for* $k = i_1$ *conditional to the sample path* $\boldsymbol{\zeta}_{[\![0,i_1]\!]}$, *then* $a_j^*\left(\boldsymbol{\zeta}_{[\![i_1,j]\!]}\right)$, $j \in [\![i_2, m-1]\!]$, *is an optimal solution of* (3.7) *for* $k = i_2$ *conditional to the sample path* $\boldsymbol{\zeta}_{[\![0,i_2]\!]}$.

This definition intuitively means that, given the optimal policy at a stage k_1, if we solve (3.7) at a subsequent stage $k_2 \in [\![k_1+1, m-1]\!]$, then the optimal solution would simply be the continuation from stage k_2 to $m-1$ of the optimal policy computed at stage k_1.

In the present case of the trade execution problem with an uncertain volume target, we have assumed that, at each stage $k \in [\![0, m-1]\!]$, the trader determines q_{k+1} based on the current state ζ_k. The useful information available to the trader for making a decision at stage k can be summarised by the latest estimate on the volume target D_k, the current asset price S_k, and the trading history $(Q_i)_{i\in[\![1,k]\!]}$. Let $(X_k)_{k\geq 0}$ be the stochastic process, where X_k is the volume left to be traded (buy if positive and sell if negative) at decision stage k in order to match with the current estimate of the volume target D_k:

$$X_0 \coloneqq D_0 \quad \text{and} \quad \forall k \in [\![1, m]\!] : X_k \coloneqq D_k - \sum_{i=1}^{k} Q_i. \tag{3.29}$$

Based on this definition, we observe that the state at stage k can be fully described by the current price S_k and the current estimate on the volume left to be traded X_k. We consequently adopt the following notation $\zeta_k \coloneqq (S_k, X_k)$. Also, as the trading history figures in the definition of X_k, we will omit to mention it when it does not provide any additional information. For instance, we will denote the value function by $V_k(\zeta_k)$ or equivalently by $V_k(S_k, X_k)$ instead of $V_k\left(\zeta_k, \boldsymbol{a}_{[\![0,k-1]\!]}\right)$. Also, for any stage $k \in [\![0, m-1]\!]$, the feasible set is independent of the price process and will be denoted by $\mathcal{A}_k(X_k)$ instead of $\mathcal{A}_k(S_k, X_k)$. With this notation, Proposition 3.2 can be formulated as follows when applied to the case of the optimal trade execution with an uncertain volume target:

Proposition 3.3. *Let* $V_k : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ *for* $k \in [\![0, m-1]\!]$ *be defined by* $V_{m-1}(s,x) \coloneqq \epsilon_m |x| + \left(s + \frac{\eta_m x}{\tau_m}\right) x$ *and*

$$V_k(s,x) \coloneqq \min_{q\in\mathcal{A}_k(x)} \left\{ \varrho_k \left[\epsilon_{k+1} |q| + \left(s + \frac{\eta_{k+1} q}{\tau_{k+1}}\right) q - \delta_{k+1}(s + \xi_{k+1}) \right.\right.$$

$$+ V_{k+1}(s + \xi_{k+1}, x + \delta_{k+1} - q) \Big] \Big\} \quad (3.30)$$

for $k \in [\![0, m-2]\!]$. *Then the optimal value of* (3.26) *for* $k = 0$ *equals* $V_0(S_0, D_0)$. *Moreover, the optimal strategy* $\mathbf{\Delta}^*_{[\![0,m]\!]} := (Q_1^*, \dots, Q_m^*)$ *of* (3.26) *for* $k = 0$ *is Markovian, where* $Q_m^*(S_{m-1}, X_{m-1}) = X_{m-1}$ *and, for* $k \in [\![0, m-2]\!]$, $Q_{k+1}^*(S_k, X_k)$ *is the optimal solution of* (3.30) *given* $s = S_k$ *and* $x = X_k$.

Proof. This proof combines elements of Proposition 3.2 and Lin et al. (2015, Proposition 1). First, at stage $m-1$, the volume target D_{t_f} is known. Consequently, the optimal volume Q_m^* to be traded during the last trading period τ_m must be equal to X_{m-1} in order to satisfy the market clearing constraint that

$$\sum_{k=1}^{m} Q_k = D_{t_f}. \quad (3.31)$$

Indeed, given the volumes previously traded, *i.e.* $Q_1^*, \dots, Q_{m-1}^*$, we have that the following equation must hold:

$$Q_m^* = D_{t_f} - \sum_{k=1}^{m-1} Q_k^* \overset{(3.2)}{=} D_{m-1} - \sum_{k=1}^{m-1} Q_k^* \overset{(3.29)}{=} X_{m-1}. \quad (3.32)$$

Thus $a_{m-1}^* = Q_m^* = X_{m-1}$ is the unique possible action, *i.e.* $\mathcal{A}_{m-1}(\zeta_{m-1}) = \{X_{m-1}\}$, and

$$\mathcal{C}_{m-1}(\zeta_{m-1}, a_{m-1}^*) \overset{(3.4)}{=} \widetilde{S}_m Q_m^* = \epsilon_m |X_{m-1}| + \left(S_{m-1} + \frac{\eta_m X_{m-1}}{\tau_m}\right) X_{m-1} \in \mathcal{Z}_{m-1}, \quad (3.33)$$

which implies that

$$V_{m-1}(s, x) \overset{(3.26a)}{=} \varrho_{m-1}[\mathcal{C}_{m-1}(\zeta_{m-1}, a_{m-1}^*)] \quad (3.34)$$

$$\overset{Axiom\ 3.2}{=} \mathcal{C}_{m-1}(\zeta_{m-1}, a_{m-1}^*) = \epsilon_m |x| + \left(s + \frac{\eta_m x}{\tau_m}\right) x. \quad (3.35)$$

At the initial stage, the trader desires to minimise $\varrho_{[\![0,m-1]\!]}\left[\mathcal{C}_0^{cum}\left(\mathbf{\Delta}_{[\![0,m]\!]}\right)\right]$ and, to this end, solves the following optimisation problem

$$\min_{\mathbf{Q}} \ \varrho_{[\![0,m-1]\!]}[\mathcal{C}_0^{cum}(\mathbf{\Delta}_{[\![0,m]\!]})] \quad (3.36)$$

$$= \min_{Q_1,\dots,Q_m} \varrho_0 \left[\mathcal{C}_0 + \dots + \varrho_{m-3}\left[\mathcal{C}_{m-3} + \varrho_{m-2}[\mathcal{C}_{m-2} + \varrho_{m-1}[\mathcal{C}_{m-1}]]\right] \cdots\right] \quad (3.37)$$

$$\overset{(3.35)}{=} \min_{Q_1,\dots,Q_{m-1}} \varrho_0 \left[\mathcal{C}_0 + \dots + \varrho_{m-3}\left[\mathcal{C}_{m-3} + \varrho_{m-2}[\mathcal{C}_{m-2} + V_{m-1}(S_{m-1}, X_{m-1})]\right] \cdots\right]. \quad (3.38)$$

By Bellman's principle of optimality, or equivalently, the time-consistency of the optimal solution, we obtain that (3.38) equals to

$$\min_{Q_1,\dots,Q_{m-2}} \varrho_0 \Big[\mathcal{C}_0 + \dots + \varrho_{m-3}\Big[\mathcal{C}_{m-3} + \min_{Q_{m-1} \in \mathcal{A}_{m-2}(X_{m-2})} \Big\{ \varrho_{m-2}[\mathcal{C}_{m-2} + V_{m-1}(S_{m-1}, X_{m-1})] \Big\}\Big] \cdots\Big] \quad (3.39)$$

which in turn can be rewritten with Equation (3.3) and Axiom 3.2 as

$$\min_{Q_1,\dots,Q_{m-2}} \varrho_0 \Bigg[\mathcal{C}_0 + \cdots + \varrho_{m-3}\Bigg[\mathcal{C}_{m-3}+ \min_{Q_{m-1}\in\mathcal{A}_{m-2}(X_{m-2})} \Big\{\widetilde{S}_{m-1}Q_{m-1} + \varrho_{m-2}[-\delta_{m-1}S_{m-1} + V_{m-1}(S_{m-1}, X_{m-1})]\Big\}\Bigg]\cdots\Bigg]. \quad (3.40)$$

With the definition of X_k and with the temporary impact function (3.1b), the objective function of the innermost optimisation problem of (3.40) can be reformulated as

$$\epsilon_{m-1}\left|Q_{m-1}\right| + \left(S_{m-2} + \frac{\eta_{m-1}Q_{m-1}}{\tau_{m-1}}\right)Q_{m-1} + \varrho_{m-2}\Big[-\delta_{m-1}(S_{m-2}+\xi_{m-1}) + V_{m-1}(S_{m-2}+\xi_{m-1}, X_{m-2}+\delta_{m-1}-Q_{m-1})\Big]. \quad (3.41)$$

If, for $k \in [\![0, m-2]\!]$, we define the function

$$\phi_k(s,x,q) \coloneqq \epsilon_{k+1}\left|q\right| + \left(s + \frac{\eta_{k+1}q}{\tau_{k+1}}\right)q + \varrho_k[V_{k+1}(s+\xi_{k+1}, x+\delta_{k+1}-q) - \delta_{k+1}(s+\xi_{k+1})], \quad (3.42)$$

then the value function defined in (3.30) can be expressed as

$$V_k(S_k, X_k) = \min_{q\in\mathcal{A}_k(X_k)} \phi_k(S_k, X_k, q), \quad (3.43)$$

and in particular

$$V_{m-2}(S_{m-2}, X_{m-2}) = \min_{q\in\mathcal{A}_{m-2}(X_{m-2})} \phi_{m-2}(S_{m-2}, X_{m-2}, q). \quad (3.44)$$

Consequently, since S_{m-2}, X_{m-2} and Q_{m-1} are independent of the realisations of ξ_{m-1} and δ_{m-1}, we can replace the objective function of the inner most optimisation problem of (3.40) by $\phi_{m-2}(S_{m-2}, X_{m-2}, Q_{m-1})$, which is a function in terms of S_{m-2}, X_{m-2} and Q_{m-1}. As the dependency of the trading quantity Q_{m-1} on the history $X_{[\![0,m-2]\!]}$ and $S_{[\![0,m-2]\!]}$ is encapsulated in the objective function ϕ_{m-2} and the feasible set $\mathcal{A}_{m-2}(X_{m-2})$, the optimal value of the inner most optimisation problem of (3.40) is a function of X_{m-2} and S_{m-2}:

$$\min_{q\in\mathcal{A}_{m-2}(X_{m-2})} \phi_{m-2}(S_{m-2}, X_{m-2}, q) = V_{m-2}(S_{m-2}, X_{m-2}). \quad (3.45)$$

The optimal volume Q^*_{m-1} to be traded during τ_{m-1} is also a function of S_{m-2} and X_{m-2} and is given by

$$Q^*_{m-1} \coloneqq \operatorname*{arg\,min}_{q\in\mathcal{A}_{m-2}(X_{m-2})} \phi_{m-2}(S_{m-2}, X_{m-2}, q). \quad (3.46)$$

By injecting (3.45) in (3.40), we obtain

$$\min_{\mathbf{Q}} \varrho_{[\![0,m-1]\!]}\Big[\mathcal{C}_0^{cum}\Big(\mathbf{\Delta}_{[\![0,m]\!]}\Big)\Big] = \min_{Q_1,\dots,Q_{m-2}} \varrho_0\left[\mathcal{C}_0 + \cdots + \varrho_{m-3}\left[\mathcal{C}_{m-3} + V_{m-2}(S_{m-2}, X_{m-2})\right]\cdots\right]. \quad (3.47)$$

A similar argument can be made for $k \in [\![m-3, 0]\!]$, which leads to the desired conclusion. □

3.4 Scenario tree

In the previous section, we have seen that, at any stage $k \in [\![0, m-1]\!]$, the information needed by the trader to make a decision on the volume q_{k+1} to be traded during the next trading period τ_{k+1} can be reduced to (S_k, X_k). In the next proposition, similarly to Lin et al. (2015, Theorem 1), we show that, for all $k \in [\![0, m-1]\!]$, the optimal volume to be traded on τ_{k+1}, which is the solution of the optimisation problem in (3.30), only depends on X_k.

PROPOSITION 3.4. *The value function $V_k(s,x)$ satisfies*

$$V_k(s,x) = sx + f_k(x)\,, \tag{3.48}$$

for $k \in [\![0, m-1]\!]$, where the function f_k is defined as $f_{m-1}(x) \coloneqq \epsilon_m |x| + \frac{\eta_m x^2}{\tau_m}$ and

$$f_k(x) \coloneqq \min_{q \in \mathcal{A}_k(x)} \left\{ \varrho_k \left[\epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + \xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q) \right] \right\} \tag{3.49}$$

for $k \in [\![0, m-2]\!]$.

PROOF. First, we have with (3.48) that $V_{m-1}(s,x) = sx + \epsilon_m |x| + \frac{\eta_m x^2}{\tau_m}$, which corresponds to the value derived in Proposition 3.3. We now assume that the recursive assumption holds for $k+1$, *i.e.* $V_{k+1}(s,x) = sx + f_{k+1}(x)$, and show that it also holds for k. By the recursive definition (3.30) of $V_k(s,x)$, we have that

$$\begin{aligned}
V_k(s,x) = \min_{q \in \mathcal{A}_k(x)} \Bigg\{ \varrho_k \Bigg[\epsilon_{k+1} |q| + \left(s + \frac{\eta_{k+1} q}{\tau_{k+1}} \right) q - \delta_{k+1}(s + \xi_{k+1}) \\
+ V_{k+1}(s + \xi_{k+1}, x + \delta_{k+1} - q) \Bigg] \Bigg\} \\
\overset{(3.49)}{=} \min_{q \in \mathcal{A}_k(x)} \Bigg\{ \varrho_k \Bigg[\epsilon_{k+1} |q| + \left(s + \frac{\eta_{k+1} q}{\tau_{k+1}} \right) q + (s + \xi_{k+1})(-\delta_{k+1} + (x + \delta_{k+1} - q)) \\
+ f_{k+1}(x + \delta_{k+1} - q) \Bigg] \Bigg\} \\
= sx + \min_{q \in \mathcal{A}_k(x)} \left\{ \varrho_k \left[\epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + \xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q) \right] \right\} \\
\overset{(3.49)}{=} sx + f_k(x)\,,
\end{aligned}$$

where the second equality holds by the recursive assumption, and the third with Axiom 3.2. □

Proposition 3.4 implies that, at any stage $k \in [\![0, m-1]\!]$, the best estimate on the remaining volume to be traded X_k is the only piece of information needed to the trader to make the decision on the volume to be traded on the next period τ_{k+1}; we will use the shorthand $\zeta_k \coloneqq (X_k)$. The set Ω_k of all possible states at stage k then corresponds to the set of all reachable values for X_k. Proposition 3.4 also implies that the optimal volume q_{k+1} to be traded during τ_{k+1} is the solution of Optimisation Problem (3.49).

REMARK 3.3. *We should also note that if there is no uncertainty on the volume to be traded, then the only risk term in (3.49) is $\varrho_k[\xi_{k+1}(x-q)]$, which corresponds to the risk of a price shift for the remaining volume to be traded. This corresponds to the risk stated in Almgren and Chriss (2001).*

In Proposition 3.5, we show that the objective value of this optimisation problem is convex in x at any stage $k \in [\![0, m-1]\!]$, which will prove to be of practical use in solving (3.49) (see proof of Proposition 3.7).

PROPOSITION 3.5. *For all $k \in [\![0, m-1]\!]$, f_k is convex in x_k.*

PROOF. For $k = m-1$, based on its definition given in Proposition 3.4, it is straightforward that the convexity of f_{m-1} holds. For any $k \in [\![0, m-2]\!]$, we proceed by induction: we assume that f_{k+1} is convex in x_{k+1} and show that this implies that f_k is convex in x_k. Let $x_{k,1}$ and $x_{k,2}$ be any two elements of Ω_k, and q_1^* and q_2^* be the associated solutions of (3.49). For any $\lambda \in (0,1)$, we define the following quantities: $x_{k,\lambda} \coloneqq \lambda x_{k,1} + (1-\lambda)\, x_{k,2}$ and $q_\lambda \coloneqq \lambda q_1^* + (1-\lambda)\, q_2^*$. With this notation, we obtain

$$\begin{aligned}
&\lambda f_k(x_{k,1}) + (1-\lambda)\, f_k(x_{k,2}) \\
&= \lambda\, \varrho_k\left[\epsilon_{k+1}\,|q_1^*| + \frac{\eta_{k+1}(q_1^*)^2}{\tau_{k+1}} + \xi_{k+1}(x_{k,1} - q_1^*) + f_{k+1}(x_{k,1} + \delta_{k+1} - q_1^*)\right] \\
&+ (1-\lambda)\, \varrho_k\left[\epsilon_{k+1}\,|q_2^*| + \frac{\eta_{k+1}(q_2^*)^2}{\tau_{k+1}} + \xi_{k+1}(x_{k,2} - q_2^*) + f_{k+1}(x_{k,2} + \delta_{k+1} - q_2^*)\right], \qquad (3.50)
\end{aligned}$$

which implies by Axiom 3.1 that

$$\begin{aligned}
\lambda f_k(x_{k,1}) + (1-\lambda)\, f_k(x_{k,2}) \geq \varrho_k\Bigg[&\epsilon_{k+1}(\lambda\,|q_1^*| + (1-\lambda)\,|q_2^*|) + \frac{\eta_{k+1}}{\tau_{k+1}}\Big(\lambda(q_1^*)^2 + (1-\lambda)(q_2^*)^2\Big) \\
&+ \xi_{k+1}(x_{k,\lambda} - q_\lambda) + \lambda f_{k+1}(x_{k,1} + \delta_{k+1} - q_1^*) + (1-\lambda)\, f_{k+1}(x_{k,2} + \delta_{k+1} - q_2^*)\Bigg]. \qquad (3.51)
\end{aligned}$$

By the convexity of the functions $y \mapsto |y|$, $y \mapsto y^2$, and f_{k+1}, (3.51) simplifies to

$$\begin{aligned}
&\lambda f_k(x_{k,1}) + (1-\lambda)\, f_k(x_{k,2}) && (3.52)\\
&\geq \varrho_k\left[\epsilon_{k+1}\,|q_\lambda| + \frac{\eta_{k+1} q_\lambda^2}{\tau_{k+1}} + \xi_{k+1}(x_{k,\lambda} - q_\lambda) + f_{k+1}(x_{k,\lambda} + \delta_{k+1} - q_\lambda)\right] && (3.53)\\
&\overset{(3.49)}{\geq} f_k(x_{k,\lambda})\,. && (3.54)
\end{aligned}$$

This concludes the proof. □

If we desire to solve the dynamic program using the value iteration algorithm, we need to solve (3.49) for each stage $k \in [\![0, m-1]\!]$ and each possible state $\zeta_k \in \Omega_k$, which means for each possible X_k. With the current assumptions, the size of the scenario tree would be infinite, which precludes a numerical solution of the problem. The size of the tree can however be reduced if we make the following additional assumptions:

Assumption 3.5. *The smallest number of shares that can be exchanged over any period is given by* $\bar{q} > 0$. *Moreover, the amount of shares* q *that can be traded over any given trading period must be an integer multiple of* $\bar{q}$.

Assumption 3.6. *The two stochastic processes representing* (i) *the asset price independent of the trades impact,* i.e.

$$\forall k \in [\![1, m]\!] : S_k = S_{k-1} + \xi_k, \tag{3.55}$$

and (ii) *the estimate of the volume target are martingales. Consequently,*

$$\forall k \in [\![0, m-1]\!] : \mathbb{E}[\, S_{k+1} \mid \mathcal{F}_k] = S_k \Leftrightarrow \mathbb{E}[\, \xi_{k+1} \mid \mathcal{F}_k] = 0, \tag{3.56}$$

$$\forall k \in [\![0, m-1]\!] : \mathbb{E}[\, D_{k+1} \mid \mathcal{F}_k] = D_k \Leftrightarrow \mathbb{E}[\, \delta_{k+1} \mid \mathcal{F}_k] = 0. \tag{3.57}$$

Assumption 3.7. *For any trading period* $k \in [\![1, m]\!]$, *the associated price shift* ξ_k *is a discrete random variable. The set of possible outcomes is finite and denoted by* Ω_{ξ_k}, *with* $\mathbb{P}_{\xi_k}$ *the associated probability measure.*

Assumption 3.8. *For any trading period* $k \in [\![1, m-1]\!]$, *the associated forecast update* δ_k *is a discrete random variable over* $\bar{q}\,\mathbb{Z} \coloneqq \{\delta \mid \delta = \bar{q}z \text{ and } z \in \mathbb{Z}\}$. *The set of possible outcomes is finite and denoted by* Ω_{δ_k}, *with* $\mathbb{P}_{\delta_k}$ *the associated probability measure. The smallest and largest realisations are denoted by* $\delta_k^{\min}$ *and* $\delta_k^{\max}$.

Assumption 3.9. *The random variables representing the price shift* ξ_k *and the forecast update* δ_k *are independent from one another for all* $k \in [\![1, m]\!]$.

Assumptions 3.5 and 3.6 are commonly observed in practice, whereas Assumptions 3.7 and 3.8 hold from a modelling perspective. More precise distributions can always be obtained by increasing the number of outcomes. As a consequence of Assumption 3.6, we must have that $\delta_k^{\min} \leq 0$ and $\delta_k^{\max} \geq 0$ for any $k \in [\![0, m-1]\!]$.

As a consequence of Assumption 3.5, for any stage $k \in [\![0, m-2]\!]$, the feasible set $\mathcal{A}_k(X_k)$ is equal to $\bar{q}\,\mathbb{Z}$, whereas the feasible set at the last stage $m-1$ is the singleton $\mathcal{A}_{m-1}(X_{m-1}) = \{X_{m-1}\}$ since the clearing constraint must be satisfied (see Equation (3.32)). Consequently, the reachable states for stage $k \in [\![1, m-1]\!]$ is $\Omega_k = \bar{q}\,\mathbb{Z}$, which implies that the size of the tree remains infinite. The trade execution problem can consequently not be solved numerically given its current formulation. We show subsequently that this issue can be circumvented by further simplifying the scenario tree after showing that a trader has no interest in over-trading, which then implies that only a finite number of states are reachable.

Definition 3.5 (Over-trade). *For any* $k \in [\![0, m-1]\!]$, *a trader trading the quantity* q_{k+1} *during the trading period* τ_{k+1} *is said to* over-trade *if one of these conditions holds:*

$$q_{k+1} > 0 \quad \textit{and} \quad x_k - q_{k+1} + \delta_{k+1}^{\max} < -\delta_{[\![k+2,m]\!]}^{\max}, \tag{3.58}$$

$$q_{k+1} < 0 \quad \textit{and} \quad x_k - q_{k+1} + \delta_{k+1}^{\min} > -\delta_{[\![k+2,m]\!]}^{\min}, \tag{3.59}$$

where $\delta^{\min}_{[\![k_1,k_2]\!]}$ *and* $\delta^{\max}_{[\![k_1,k_2]\!]}$ *are defined as follows:*

$$\delta^{\min}_{[\![k_1,k_2]\!]} \coloneqq \sum_{j=k_1}^{k_2} \delta_j^{\min} \leq 0 \quad \text{and} \quad \delta^{\max}_{[\![k_1,k_2]\!]} \coloneqq \sum_{j=k_1}^{k_2} \delta_j^{\max} \geq 0. \tag{3.60}$$

In words, $\delta^{\min}_{[\![k_1,k_2]\!]}$ (*resp.* $\delta^{\max}_{[\![k_1,k_2]\!]}$) defined in (3.60) corresponds to the minimum (*resp.* maximum) possible cumulative volume update occurring over the trading periods spanning from τ_{k_1} to τ_{k_2}. The value $x_k - q_{k+1} + \delta_{k+1}$ corresponds to X_{k+1}, the random variable that represents the estimate on the remaining volume to be traded at stage $k+1$. If (3.58) holds, then the trader acquires some positions during τ_{k+1} as $q_{k+1} > 0$, and

$$X_{k+1} = x_k - q_{k+1} + \delta_{k+1} \leq x_k - q_{k+1} + \delta^{\max}_{k+1} < -\delta^{\max}_{[\![k+1,m]\!]} \leq 0, \tag{3.61}$$

which finally implies that

$$X_{k+1} \leq -\delta^{\max}_{[\![k+1,m]\!]} - \bar{q}. \tag{3.62}$$

This means that, no matter the realisation $\left(\omega_{\xi_{k+1}}, \omega_{\delta_{k+1}}\right) \in \Omega_{\xi_{k+1}} \times \Omega_{\delta_{k+1}}$, the trader will have to sell at least $\bar{q}$ positions with probability one over the remaining trading periods in order to satisfy the clearing constraint (3.31). The trader has thus over-traded as at least $\bar{q}$ positions have been bought during τ_{k+1} to be sold back at a later stage. A similar reasoning holds if (3.59) holds. For any stage $k \in [\![0, m-1]\!]$, let $\widetilde{\mathcal{A}}_k(x_k)$ be the set of trading quantities that does not lead to over-trading:

$$\widetilde{\mathcal{A}}_k(x_k) \coloneqq \bar{q}\,\mathbb{Z} \setminus \left(\left\{ q > 0 \,\middle|\, x_k - q + \delta^{\max}_{k+1} < -\delta^{\max}_{[\![k+2,m]\!]} \right\} \right.$$
$$\left. \cup \left\{ q < 0 \,\middle|\, x_k - q + \delta^{\min}_{k+1} > -\delta^{\min}_{[\![k+2,m]\!]} \right\} \right) \tag{3.63}$$

Proposition 3.6. *If Assumptions 3.1 to 3.9 hold, then the following assertions hold for any stage* $k \in [\![0, m-1]\!]$ *and any* $x_k \in \Omega_k$*:*

(i) $q^*_{k+1}(x_k) \in \widetilde{\mathcal{A}}_k(x_k)$*, where* $q^*_{k+1}(x_k)$ *is the solution of Optimisation Problem (3.49) for* $x = x_k$*.*

(ii) If $x_k < -\delta^{\max}_{[\![k+1,m]\!]}$*, then*

  (a) $q^*_{k+1}(x_k) < 0$*,*

  (b) $f_k(x_k + \bar{q}) \leq f_k(x_k) - \epsilon_{k+1}\bar{q} - \frac{\eta_{k+1}}{\tau_{k+1}}\bar{q}^2$*.*

(iii) If $x_k > -\delta^{\min}_{[\![k+1,m]\!]}$*, then*

  (a) $q^*_{k+1}(x_k) > 0$*,*

  (b) $f_k(x_k - \bar{q}) \leq f_k(x_k) - \epsilon_{k+1}\bar{q} - \frac{\eta_{k+1}}{\tau_{k+1}}\bar{q}^2$*.*

PROOF. In this proof, we use q^*_{k+1} as a shorthand for $q^*_{k+1}(x_k)$. For $k = m-1$, it is straightforward to show that *(i)*, *(ii)* and *(iii)* hold based on the definition of f_{m-1} and on the fact that $q^*_m(x_{m-1}) = x_{m-1}$. For any $k \in [\![0, m-2]\!]$, we proceed by induction: we assume that *(i)*, *(ii)* and *(iii)* hold for $k+1$ and show that it implies that it also holds for k:

(i) If $q^*_{k+1} \notin \widetilde{\mathcal{A}}_k(x_k)$, then either

$$q^*_{k+1} > 0 \quad \text{and} \quad x_k - q^*_{k+1} + \delta^{\max}_{k+1} < -\delta^{\max}_{[\![k+2,m]\!]} \leq 0, \tag{3.64}$$

or

$$q^*_{k+1} < 0 \quad \text{and} \quad x_k - q^*_{k+1} + \delta^{\min}_{k+1} > -\delta^{\min}_{[\![k+2,m]\!]} \geq 0. \tag{3.65}$$

We will show that (3.64) cannot occur; a similar reasoning can be applied to show that (3.65) cannot hold either. By contradiction, we assume that q^*_{k+1} is such that (3.64) holds true. Then, by the optimality of q^*_{k+1}, we must have that:

$$\begin{aligned} f_k(x_k) \leq \varrho_k \Big[\epsilon_{k+1} \left| q^*_{k+1} - \bar{q} \right| + \frac{\eta_{k+1}}{\tau_{k+1}} \left(q^*_{k+1} - \bar{q} \right)^2 + \xi_{k+1}\left(x_k - \left(q^*_{k+1} - \bar{q} \right) \right) \\ + f_{k+1}\left(x_k + \delta_{k+1} - \left(q^*_{k+1} - \bar{q} \right) \right) \Big]. \end{aligned} \tag{3.66}$$

As $x_k + \delta_{k+1} - q^*_{k+1} < -\delta^{\max}_{[\![k+2,m]\!]}$, the induction assumption implies with *(ii)* that

$$f_{k+1}\left(x_k + \delta_{k+1} - \left(q^*_{k+1} - \bar{q} \right) \right) \leq f_{k+1}\left(x_k + \delta_{k+1} - q^*_{k+1} \right) - \epsilon_{k+2}\bar{q} - \frac{\eta_{k+2}}{\tau_{k+2}}\bar{q}^2. \tag{3.67}$$

Consequently, (3.66) becomes

$$\begin{aligned} f_k(x_k) \leq & \left(\epsilon_{k+1} - \epsilon_{k+2} \right) \bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}} \bar{q}\left(\bar{q} - 2q^*_{k+1} \right) - \frac{\eta_{k+2}}{\tau_{k+2}} \bar{q}^2 + \epsilon_{k+1} \left| q^*_{k+1} \right| + \frac{\eta_{k+1}}{\tau_{k+1}} \left(q^*_{k+1} \right)^2 \\ & + \varrho_k \left[\xi_{k+1}\bar{q} + \xi_{k+1}\left(x_k - q^*_{k+1} \right) + f_{k+1}\left(x_k + \delta_{k+1} - q^*_{k+1} \right) \right]. \end{aligned} \tag{3.68}$$

Since $q^*_{k+1} > 0$ and $q^*_{k+1} \in \bar{q}\,\mathbb{Z}$, $\left(\bar{q} - 2q^*_{k+1} \right) < 0$. Moreover, with Assumption 3.2, we obtain that $\epsilon_{k+1} - \epsilon_{k+2} \leq 0$. Hence, it implies that

$$\begin{aligned} f_k(x_k) < & \epsilon_{k+1} \left| q^*_{k+1} \right| + \frac{\eta_{k+1}}{\tau_{k+1}} \left(q^*_{k+1} \right)^2 \\ & + \varrho_k \left[\xi_{k+1}\bar{q} + \xi_{k+1}\left(x_k - q^*_{k+1} \right) + f_{k+1}\left(x_k + \delta_{k+1} - q^*_{k+1} \right) \right]. \end{aligned} \tag{3.69}$$

If we define the constant $c \coloneqq -\left(x_k - q^*_{k+1} \right)$, then $c \geq \bar{q} > 0$ since $x_k + \delta_{k+1} - q^*_{k+1} < -\delta^{\max}_{[\![k+2,m]\!]} \leq 0$ and $\delta^{\max}_{k+1} \geq 0$. Therefore, if we define the random variables

$$A \coloneqq -\xi_{k+1} \quad \text{and} \quad B \coloneqq f_{k+1}\left(x_k + \delta_{k+1} - q^*_{k+1} \right), \tag{3.70}$$

then the last term of (3.69) can be written as $\varrho_k[(c - \bar{q})\, A + B]$. It is follows from the definition of A and Assumptions 3.4 and 3.6 to 3.9 that $\mathbb{E}[\, A \,|\, \mathcal{F}_k] = 0$, and that A and B are independent random variables. Consequently, we have with Corollary A.1 (see page 177) that

$$\varrho_k[(c - \bar{q})\, A + B] \leq \varrho_k[cA + B], \tag{3.71}$$

since $0 \leq c - \bar{q} < c$. We finally obtain that

$$f_k(x_k) < \epsilon_{k+1} \left| q^*_{k+1} \right| + \frac{\eta_{k+1}}{\tau_{k+1}} \left(q^*_{k+1} \right)^2 + \varrho_k [\xi_{k+1}(x_k - q^*_{k+1}) + f_{k+1}(x_k + \delta_{k+1} - q^*_{k+1})] = f_k(x_k)\,, \quad (3.72)$$

which is absurd; q^*_{k+1} can consequently not satisfy (3.64).

(ii.a) If $x_k < -\delta^{\max}_{[\![k+1,m]\!]} \leq 0$ then $x_k + \delta_{k+1} < -\delta^{\max}_{[\![k+2,m]\!]}$. Therefore, with *(i)* we must have that $q^*_{k+1} \leq 0$ since otherwise $q^*_{k+1}(x_k) \notin \widetilde{\mathcal{A}}_k(x_k)$. We will now show that trading $q_{k+1} = -\bar{q}$ cannot be worse than trading $q_{k+1} = 0$. This is shown as follows

$$\begin{aligned}
&\varrho_k \left[\epsilon_{k+1} \left| -\bar{q} \right| + \frac{\eta_{k+1}}{\tau_{k+1}} (-\bar{q})^2 + \xi_{k+1}(x_k - (-\bar{q})) + f_{k+1}(x_k + \delta_{k+1} - (-\bar{q})) \right] \\
&\leq \varrho_k \left[\epsilon_{k+1} \bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}} \bar{q}^2 + \xi_{k+1}(x_k - (-\bar{q})) + f_{k+1}(x_k + \delta_{k+1}) - \epsilon_{k+2} \bar{q} - \frac{\eta_{k+2}}{\tau_{k+2}} \bar{q}^2 \right] \\
&< \varrho_k [\xi_{k+1}(x_k + \bar{q}) + f_{k+1}(x_k + \delta_{k+1})] \\
&\leq \varrho_k [\xi_{k+1} x_k + f_{k+1}(x_k + \delta_{k+1})]\,,
\end{aligned}$$

where the first inequality holds by the induction assumption and *(ii)*, the second inequality derives from Assumption 3.2, and finally the last inequality holds based on Assumptions 3.4 and 3.6 to 3.9 and Corollary A.1 and the fact that $x_k < 0$ and $x_k \in \bar{q}\,\mathbb{Z} \Rightarrow x_k \leq x_k + \bar{q} \leq 0$.

(ii.b) Let q^*_{k+1} be the solution of Optimisation Problem (3.49) for $x = x_k$. We have just shown in *(ii.a)* that $q^*_{k+1} < 0$ if $x_k < -\delta^{\max}_{[\![k,m]\!]}$. As a consequence,

$$\begin{aligned}
f_k(x_k + \bar{q}) &\leq \varrho_k \left[\epsilon_{k+1} \left| q^*_{k+1} + \bar{q} \right| + \frac{\eta_{k+1}}{\tau_{k+1}} \left(q^*_{k+1} + \bar{q} \right)^2 + \xi_{k+1}(x_k + \bar{q} - (q^*_{k+1} + \bar{q})) \right. \\
&\qquad \left. + f_{k+1}(x_k + \bar{q} + \delta_{k+1} - (q^*_{k+1} + \bar{q})) \right] \\
&= \varrho_k \left[\epsilon_{k+1} \left| q^*_{k+1} + \bar{q} \right| + \frac{\eta_{k+1}}{\tau_{k+1}} \left(q^*_{k+1} + \bar{q} \right)^2 + \xi_{k+1}(x_k - q^*_{k+1}) \right. \\
&\qquad \left. + f_{k+1}(x_k + \delta_{k+1} - q^*_{k+1}) \right] \\
&= -\epsilon_{k+1} \bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}} \bar{q}(\bar{q} + 2q^*_{k+1}) + f_k(x_k)\,.
\end{aligned}$$

Since $q^*_{k+1} < 0$ and $q^*_{k+1} \in \bar{q}\,\mathbb{Z}$, we obtain that $\bar{q} + 2q^*_{k+1} \leq -\bar{q}$, and consequently the desired result, *i.e.* $f_k(x_k + \bar{q}) \leq f_k(x_k) - \epsilon_{k+1}\bar{q} - \frac{\eta_{k+1}}{\tau_{k+1}}\bar{q}^2$.

(iii) This can be shown using a similar manner as for *(ii)*.

□

Proposition 3.6 implies that the size of the feasible set of each stage $k \in [\![0, m-1]\!]$ can be reduced to a finite set without altering the optimal solution of Equation (3.49).

COROLLARY 3.1. *Given Assumptions 3.1 to 3.9, we have that, for all $k \in [\![0, m-2]\!]$, the solution to*

$$f_k(x) \coloneqq \min_{q \in \widetilde{\mathcal{A}}_k(x_k)} \left\{ \varrho_k \left[\epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + \xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q) \right] \right\} \tag{3.73}$$

is identical to the solution of (3.49).

PROOF. The proof is straightforward based on Proposition 3.6. □

COROLLARY 3.2. *Given Assumptions 3.1 to 3.9, the initial estimate D_0 on the volume target, and any stage $k \in [\![1, m-1]\!]$, the following inequalities holds for all $x_k \in \Omega_k$:*

$$\underline{x} + \sum_{i=1}^{k} \delta_i^{\min} \leq x_k \leq \bar{x} + \sum_{i=1}^{k} \delta_i^{\max}, \tag{3.74}$$

where $\underline{x} \coloneqq \min\left(D_0, -\delta^{\max}_{[\![1,m]\!]}\right) \leq 0$ and $\bar{x} \coloneqq \max\left(D_0, -\delta^{\min}_{[\![1,m]\!]}\right) \geq 0$. The set of reachable states is consequently lower and upper bounded at any stage.

PROOF. According to Corollary 3.1, the reachable states at a stage $k \in [\![1, m-1]\!]$ is the set expressed as follows

$$\Omega_k = \left\{ x_k \,\middle|\, x_k = x_{k-1} - q_k + \delta_k,\ q_k \in \widetilde{\mathcal{A}}_{k-1}(x_{k-1}) \ \text{ and } \ x_{k-1} \in \Omega_{k-1} \right\}. \tag{3.75}$$

First, for any given state x_{k-1} and trading quantity q_k, we can bound the state of the next stage as follows:

$$x_{k-1} - q_k + \delta_k^{\min} \leq x_k \leq x_{k-1} - q_k + \delta_k^{\max}. \tag{3.76}$$

Then, as $q_k \in \widetilde{\mathcal{A}}_{k-1}(x_{k-1})$, we have three possible cases given the definition of $\widetilde{\mathcal{A}}_{k-1}(x_{k-1})$ (see (3.63)):

$$\begin{cases} q_k < 0 \text{ and } x_{k-1} - q_k + \delta_k^{\min} \leq -\delta^{\min}_{[\![k+1,m]\!]} & \Rightarrow x_{k-1} \leq x_{k-1} - q_k \leq -\delta^{\min}_{[\![k,m]\!]}, \\ q_k = 0 & \Rightarrow x_{k-1} - q_k = x_{k-1}, \\ q_k > 0 \text{ and } x_{k-1} - q_k + \delta_k^{\max} \geq -\delta^{\max}_{[\![k+1,m]\!]} & \Rightarrow -\delta^{\max}_{[\![k,m]\!]} \leq x_{k-1} - q_k \leq x_{k-1}, \end{cases} \tag{3.77}$$

hence

$$\min\left(x_{k-1}, -\delta^{\max}_{[\![k,m]\!]}\right) \leq x_{k-1} - q_k \leq \max\left(x_{k-1}, -\delta^{\min}_{[\![k,m]\!]}\right). \tag{3.78}$$

By combining Equations (3.76) and (3.78), we obtain the following inequalities:

$$\min\left(x_{k-1}, -\delta^{\max}_{[\![k,m]\!]}\right) + \delta_k^{\min} \leq x_k \leq \max\left(x_{k-1}, -\delta^{\min}_{[\![k,m]\!]}\right) + \delta_k^{\max}. \tag{3.79}$$

D_0 is the unique possible state at the initial stage, *i.e.* $\Omega_0 = \{D_0\}$, which implies that Equation (3.74) holds for $k = 1$. We show now that, if Equation (3.74) holds for a stage $k-1 \in [\![1, m-2]\!]$, then it also holds for stage k. With the recursion assumption and (3.79), we obtain

$$x_k \geq \min\left(\bar{x} + \sum_{i=1}^{k-1} \delta_i^{\min}, -\delta^{\max}_{[\![k,m]\!]} \right) + \delta_k^{\min}$$

$$= \min\left(\min\left(D_0,\, -\delta^{\max}_{[\![1,m]\!]}\right) + \sum_{i=1}^{k-1} \delta^{\min}_i,\, -\delta^{\max}_{[\![k,m]\!]}\right) + \delta^{\min}_k$$
$$= \min\left(D_0,\, -\delta^{\max}_{[\![1,m]\!]}\right) + \sum_{i=1}^{k} \delta^{\min}_i = \bar{x} + \sum_{i=1}^{k} \delta^{\min}_i, \tag{3.80}$$

and

$$x_k \leq \max\left(\bar{x} + \sum_{i=1}^{k-1} \delta^{\max}_i,\, -\delta^{\min}_{[\![k,m]\!]}\right) + \delta^{\max}_k$$
$$= \max\left(\max\left(D_0,\, -\delta^{\min}_{[\![1,m]\!]}\right) + \sum_{i=1}^{k-1} \delta^{\max}_i,\, -\delta^{\min}_{[\![k,m]\!]}\right) + \delta^{\max}_k$$
$$= \max\left(D_0,\, -\delta^{\min}_{[\![1,m]\!]}\right) + \sum_{i=1}^{k} \delta^{\max}_i = \bar{x} + \sum_{i=1}^{k} \delta^{\max}_i, \tag{3.81}$$

where we used the fact that $\forall k \in [\![1, m]\!] : \delta^{\max}_k \geq 0$ and $\delta^{\min}_k \leq 0$. This concludes the proof. □

As a consequence of Corollary 3.1, we obtain that the scenario tree is finite. This tree can be represented as in Figure 3.2. The total number of nodes for which we will have to solve Optimisation Problem (3.73) scales linearly with the number of stages. These nodes correspond to any reachable node of any stage $k \in [\![0, m-2]\!]$; they are coloured in green nodes in Figure 3.2. For stage $m-1$ (the blue nodes), we have with Proposition 3.3 that the feasible set is a singleton and thus no optimisation problem is required to be solved. We denote by d_k the number of reachable states for stage $k \in [\![0, m-1]\!]$, and by $\{x_{k,1}, \ldots, x_{k,d_k}\}$ the increasing sequence in terms of volume left to be traded. With Assumption 3.5, we have that $x_{k,j+1} = x_{k,j} + \bar{q}$ for all $j \in [\![1, d_k - 1]\!]$.

Remark 3.4. *In practice, there sometimes exists an upper limit on the order size, which would imply a greater simplification of the scenario tree.*

Finally, given a state $\zeta_k = x_k \in \Omega_k$ and a decision $a_k = q_{k+1} \in \widetilde{\mathcal{A}}_k(x_k)$ at stage k, the probability of transition from ζ_k to any of its child node $\zeta_{k+1} = x_{k+1} \in \Omega_{k+1}$ is denoted by $\mathbb{P}[\zeta_{k+1} \mid \zeta_k, a_k]$ or $\mathbb{P}[x_{k+1} \mid x_k, q_{k+1}]$. This probability depends both on ζ_k and the action taken a_k and is given by

$$\mathbb{P}[x_{k+1} \mid x_k, q_{k+1}] \coloneqq \mathbb{P}_{\delta_{k+1}}[\delta_{k+1} = x_{k+1} - x_k + q_{k+1}]\,, \tag{3.82}$$

where $\mathbb{P}_{\delta_{k+1}}$ is the probability measure associated to δ_{k+1} as mentioned in Assumption 3.8. Consequently, given a *trading strategy* denoted by $\boldsymbol{\Delta}$, which is composed of the decision variables for each stage, the probability that scenario $\boldsymbol{\zeta}_{[\![0,m-1]\!]}$ happens is then given by $\prod_{k=0}^{m-2} \mathbb{P}[\zeta_{k+1} \mid \zeta_k, a_k]$.

3.5 Practical implementation

We have just seen in Section 3.4 that we must solve Optimisation Problem (3.73) for every reachable node of stages 0 to $m-2$ in order to solve the trade execution problem via the value iteration algorithm. Indeed, in the value iteration algorithm, the dynamic program is solved

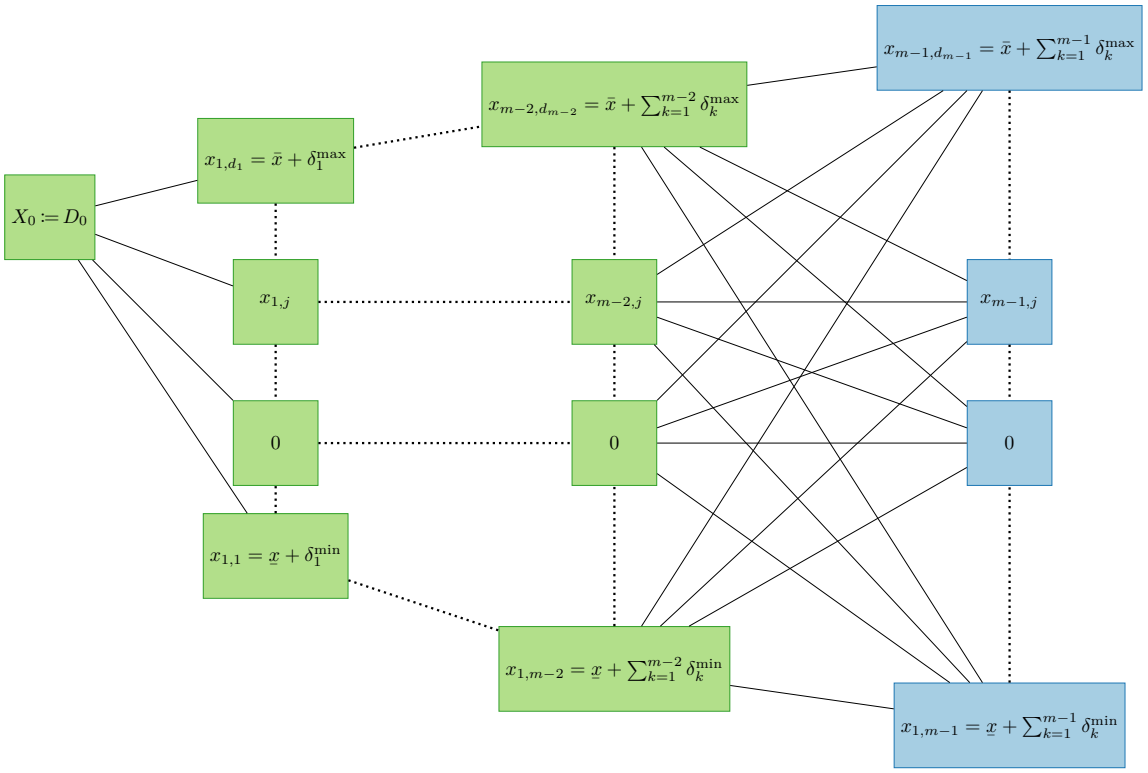


FIGURE 3.2: Scenario tree for the optimal trade execution with uncertain volume target.

by backward recursion, where the value function is calculated from the last stage backward to the first stage (Powell, 2011). Hence, Optimisation Problem (3.73) must be solved with $x = x_k$ for each state $\zeta_k = x_k \in \Omega_k$ at each stage k. In this work, we have assumed that, for any $k \in [\![0, m-1]\!]$, the one-step conditional risk measure ϱ_k corresponds to $\mathbb{E}$-$\mathrm{CVaR}_{\alpha_k,\lambda_k}$. With Axiom 3.2, part of the objective function (3.73) can be moved out of the risk measure as follows:

$$\min_{q \in \widetilde{\mathcal{A}}_k(x_k)} \left\{ \varrho_k \left[\epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + \xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q) \right] \right\}$$
$$= \min_{q \in \widetilde{\mathcal{A}}_k(x_k)} \left\{ \epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + \varrho_k[\xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q)] \right\} \quad (3.83)$$

For a fixed quantity q to be traded during τ_{k+1}, $\xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q)$ is consequently the random part of the stage cost for which we should measure the related risk with ϱ_k.

One approach to solve Optimisation Problem (3.83) is to use the definition of the CVaR_{α_k} of Rockafellar and Uryasev (2000, 2002) (see Equation (3.23)). As the sample space is assumed finite, the infimum is attained and can consequently be replaced by the minimum. Consequently, we can integrate the computation of the CVaR_{α_k} in the minimisation problem (3.83), *i.e.*

$$\min_{q \in \widetilde{\mathcal{A}}_k(x_k)} \left\{ \epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + \varrho_k[\xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q)] \right\}$$
$$= \min_{q \in \widetilde{\mathcal{A}}_k(x_k),\, W \in \mathcal{Z}_k} \left\{ \epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + (1-\lambda_k)\, \mathbb{E}[\xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q) \mid \mathcal{F}_k] \right.$$

$$+ \lambda_k \left(W + \frac{1}{\alpha_k} \mathbb{E}\Big[(\xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q) - W)_+ \,\Big|\, \mathcal{F}_k \Big] \right) \Bigg\}. \quad (3.84)$$

The efficiency of this approach is however hindered by the fact that this optimisation problem is not continuous. Another approach is to exploit the particular structure of the trade execution problem to reduce the complexity of finding the solution of Optimisation Problem (3.83), which in turns makes the resolution of our problem tractable via the value iteration algorithm. This second approach relies on the fact that the $\mathbb{E}$-CVaR$_{\alpha_k,\lambda_k}$ risk measure belongs to the class of *coherent risk measures (CRM)* (Artzner et al., 1999, Definition 2.4), and therefore inherits the property of having a dual representation formulated as follows:

$$\begin{aligned} &\epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + \varrho_k[\xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q)] \\ &\qquad = \epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + \max_{\mu_{k+1} \in \mathcal{D}_{k+1}} \mathbb{E}^{\mu_{k+1}}[\xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q)] \end{aligned} \quad (3.85)$$

where $\mathcal{D}_{k+1}$ is the associated risk-set, which, because the sample sets $\Omega_{\xi_{k+1}}$ and $\Omega_{\delta_{k+1}}$ are assumed finite, is a closed bounded convex subset of $\mathcal{P}_{k+1}$, the set of probability measures on $\Omega_{\xi_{k+1} \times \delta_{k+1}} \coloneqq \Omega_{\xi_{k+1}} \times \Omega_{\delta_{k+1}}$ (Shapiro et al., 2009, Theorems 6.4 and 6.6). Additionally, as $\mathbb{E}$-CVaR$_{\alpha_k,\lambda_k}$ belongs to the more restrictive class of *polyhedral coherent risk measures (PCRM)*, this risk-set $\mathcal{D}_{k+1}$ is a polyhedron (Kirilyuk, 2004). Let Ext$(\mathcal{D}_{k+1})$ be the set of extreme points of this polyhedron $\mathcal{D}_{k+1}$. The optimal solution of the linear program (3.85) is consequently located at one of the extreme points of $\mathcal{D}_{k+1}$, which implies that the Optimisation Problem on the right-hand-side of (3.85) simplifies to

$$\epsilon_{k+1} |q| + \frac{\eta_{k+1} q^2}{\tau_{k+1}} + \max_{\mu_{k+1} \in \mathrm{Ext}(\mathcal{D}_{k+1})} \mathbb{E}^{\mu_{k+1}}[\xi_{k+1}(x-q) + f_{k+1}(x + \delta_{k+1} - q)]\,. \quad (3.86)$$

The probability measure $\mu^*_{k+1} \in \mathcal{D}_{k+1}$ maximising (3.86) is called the *worst case probability measure*. Given the risk level α_k, the risk-aversion parameter λ_k, and the physical measure $\mathbb{P}_{k+1}$ on the finite sample set $\Omega_{\xi_{k+1} \times \delta_{k+1}}$, the risk-set $\mathcal{D}_{k+1}$ corresponding to $\mathbb{E}$-CVaR$_{\alpha_k,\lambda_k}$ is defined as follows:

$$\begin{aligned} \mathcal{D}_{k+1} \coloneqq \Bigg\{ \mu \in \mathcal{P}_{k+1} \;\Bigg|\; &\exists \upsilon \in \mathcal{P}_{k+1} \text{ such that } \forall \omega \in \Omega_{\xi_{k+1} \times \delta_{k+1}} : \\ &\mu(\omega) = (1 - \lambda_k)\, \pi(\omega) + \lambda_k \upsilon(\omega) \text{ and } 0 \leq \upsilon(\omega) \leq \frac{\pi(\omega)}{\alpha_k} \Bigg\}, \end{aligned} \quad (3.87)$$

where $\omega \coloneqq \big(\omega_{\xi_{k+1}}, \omega_{\delta_{k+1}}\big)$ with $\omega_{\xi_{k+1}} \in \Omega_{\xi_{k+1}}$ and $\omega_{\delta_{k+1}} \in \Omega_{\delta_{k+1}}$, and

$$\pi(\omega) = \mathbb{P}_{k+1}[\omega] = \mathbb{P}_{\xi_{k+1}}\Big[\omega_{\xi_{k+1}}\Big]\, \mathbb{P}_{\delta_{k+1}}\Big[\omega_{\delta_{k+1}}\Big]\,, \quad (3.88)$$

where we used the fact that ξ_{k+1} and δ_{k+1} are assumed to be independent random variables (see Assumptions 3.4 and 3.9). In words, as aforementioned, the α_k-Conditional Value-at-Risk corresponds to the expected value conditional to a quantile of worst case scenarios. Therefore,

to evaluate the risk related to the decision a_k at a stage k, we must in a first step identify the set $\Omega^*_{\xi_{k+1}\times\delta_{k+1}}$ of worst case realisations such that:

$$\mathbb{P}_{k+1}\left[\Omega^*_{\xi_{k+1}\times\delta_{k+1}}\right] \geq \alpha_k, \tag{3.89}$$

$$\forall \omega^* \in \Omega^*_{\xi_{k+1}\times\delta_{k+1}} : \mathbb{P}_{k+1}\left[\Omega^*_{\xi_{k+1}\times\delta_{k+1}} \setminus \{\omega^*\}\right] < \alpha_k \tag{3.90}$$

and $\forall \omega^* \in \Omega^*_{\xi_{k+1}\times\delta_{k+1}}, \forall \omega \in \Omega_{\xi_{k+1}\times\delta_{k+1}} \setminus \Omega^*_{\xi_{k+1}\times\delta_{k+1}}$:

$$\xi_{k+1}[\omega^*](x-q) + f_{k+1}(x+\delta_{k+1}[\omega^*]-q) \geq \xi_{k+1}[\omega](x-q) + f_{k+1}(x+\delta_{k+1}[\omega]-q) \tag{3.91}$$

where $\xi_{k+1}[\tilde{\omega}]$ and $\delta_{k+1}[\tilde{\omega}]$ are respectively the realisation of ξ_{k+1} and δ_{k+1} given $\tilde{\omega} \in \Omega_{\xi_{k+1}\times\delta_{k+1}}$.

For each stage $k \in [\![0, m-1]\!]$, we have previously shown in Corollary 3.2 and Proposition 3.6 that the set of states Ω_k and the action space $\widetilde{\mathcal{A}}_k(x_k)$ are finite. Hence, the set

$$\mathcal{S}_k = \left\{ x_k - q_{k+1} \,\middle|\, x_k \in \Omega_k \text{ and } q_{k+1} \in \widetilde{\mathcal{A}}_k(x_k) \right\} \tag{3.92}$$

is finite. Based on this observation and Equations (3.89) to (3.91), we can compute the value of $s_k \mapsto \varrho_k[\xi_{k+1}s_k + f_{k+1}(s_k+\delta_{k+1})]$ for all elements $s_k \in \mathcal{S}_k$ as follows: To find the worst case realisations, we exploit the fact that ξ_{k+1} and δ_{k+1} are independent random variables by first sorting separately the realisations of $\xi_{k+1}s_k$ and $f_{k+1}(s_k+\delta_{k+1})$ and then sort the realisations of $\xi_{k+1}s_k + f_{k+1}(s_k+\delta_{k+1})$ on the Cartesian product $\Omega_{\xi_{k+1}\times\delta_{k+1}}$. Sorting $\xi_{k+1}s_k$ and $f_{k+1}(s_k+\delta_{k+1})$ in advance, in a preliminary stage, enables to have a vector of realisations that is already partially sorted, which accelerates the sorting of the realisations on $\Omega_{\xi_{k+1}\times\delta_{k+1}}$. Additionally, if, for a given s_k, we denote by

$$\Omega_{\xi_{k+1}\times\delta_{k+1}}(s_k) \coloneqq \{\omega_1(s_k), \omega_2(s_k), \dots\} \tag{3.93}$$

the sorted set of realisations such that

$$\begin{aligned}\xi_{k+1}[\omega_j(s_k)](x-q) + f_{k+1}(x+\delta_{k+1}[\omega_j(s_k)]-q) \\ \geq \xi_{k+1}[\omega_{j+1}(s_k)](x-q) + f_{k+1}(x+\delta_{k+1}[\omega_{j+1}(s_k)]-q),\end{aligned} \tag{3.94}$$

then using $\Omega^*_{\xi_{k+1}\times\delta_{k+1}}(s_k)$ as the initial guess for $\Omega^*_{\xi_{k+1}\times\delta_{k+1}}(s_k+\bar{q})$ allows to further reduce the average complexity of the sort of the realisations associated to $s_k+\bar{q}$. Hence, to evaluate the value of $s_k \mapsto \varrho_k[\xi_{k+1}s_k + f_{k+1}(s_k+\delta_{k+1})]$ for all elements $s_k \in \mathcal{S}_k$, we should first evaluate the smallest element $s_k^{\min} \coloneqq \min\{\mathcal{S}_k\}$ and then continue in ascending order of s_k in order to exploit the similarity between the sets of worst case realisations and consequently speed up the sorting algorithm. We denote by l_k, the function

$$l_k : \mathcal{S}_k \to \mathcal{Z}_k; s_k \mapsto \varrho_k[\xi_{k+1}s_k + f_{k+1}(s_k+\delta_{k+1})]. \tag{3.95}$$

Once this function l_k is computed, the objective function of Optimisation Problem (3.73) given a decision q_{k+1} at a state $x_k \in \Omega_k$ is given by $\epsilon_{k+1}|q_{k+1}| + \frac{\eta_{k+1}q_{k+1}^2}{\tau_{k+1}} + l_k(x_k - q_{k+1})$. Finding the optimal quantity to be traded $q^*_{k+1}(x_k)$ can now be achived by evaluating every

feasible trading quantity $q_{k+1} \in \widetilde{\mathcal{A}}_k(x_k)$ and returning the one leading to the smallest value. The advantage of this approach compared to the one suggested by Rockafellar and Uryasev (2002) is that once the function l_k is computed, evaluating the objective function does not require any evaluation on $\Omega_{\xi_{k+1}\times\delta_{k+1}}$, whereas it was the case for computing the expectation in (3.84). This makes a significant difference as the function h is computed only once per stage, and is then used for all states $x_k \in \Omega_k$.

For readability purposes, we omit in the next proposition the previous result that stipulates that the set of reachable states is bounded for each stage. We thus assume that the set of reachable states for each stage is $\bar{q}\,\mathbb{Z}$, which we denote by the increasing sequence $\{j\bar{q}\}_{j\in\mathbb{Z}}$.

PROPOSITION 3.7. *Given Assumptions 3.1 to 3.9, the following assertions hold for all* $k \in [\![0, m-1]\!]$*:*

(i) For any $x_k \in \bar{q}\,\mathbb{Z} : q^*_{k+1}(x_k + \bar{q}) \in \left\{q^*_{k+1}(x_k), q^*_{k+1}(x_k) + \bar{q}\right\}$*, where* $q^*_{k+1}(x_k)$ *is the solution of Optimisation Problem (3.49) for* $x = x_k$*.*

(ii) $\{f_k(j\bar{q})\}_{j\in\mathbb{Z}}$ *is a convex sequence.*

PROOF. The proof is provided in Appendix A.2 at page 178. □

COROLLARY 3.3. *Given Assumptions 3.1 to 3.9, for any* $x_k \in \bar{q}\,\mathbb{Z}$*, we have for* $\mu \in \mathbb{Z}$ *that*

$$q^*_{k+1}(x_k + \mu\bar{q}) \in \left\{q^*_{k+1}(x_k) + j\bar{q} \,\middle|\, j \in [\![0, \mu]\!]\right\}, \tag{3.96}$$

where, if $\mu < 0$*,* $[\![0, \mu]\!]$ *must be understood as* $[\![\mu, 0]\!]$*.*

Corollary 3.3 allows to further reduce the feasible set $\mathcal{A}_k(x_k)$ in the search of the optimal quantity to be traded if we know the optimal trading quantity $q^*_{k+1}(\tilde{x}_k)$ for another state $\tilde{x}_k \in \Omega_k$. Moreover, the convexity of the sequence $\{f_k(j\bar{q})\}_{j\in\mathbb{Z}}$ allows to simplify greatly the search of the optimal trading quantity. Algorithm 3.1 shows how the dynamic program can be solved by exploiting the previous results.

ALGORITHM 3.1 (DYNAMIC PROGRAM).

Input: $\forall k \in [\![1, m]\!] : \epsilon_k, \eta_k, \xi_k, \delta_k$.

Initialisation: *Compute* f_{m-1}.

Iteration $(k \in \{m-2, m-3, \dots, 0\})$**:**

$\forall s_k \in \mathcal{S}_k$: *Compute* $l_k(s_k)$,
$x_k^* \coloneqq \arg\min_{s_k \in \mathcal{S}_k} l_k(s_k)$ *and* $q_{k+1}^*(x_k^*) = 0$,
Set $q^- = 0$ *and* $q^+ = 0$.

For $(j \geq 1)$**:**

$x^- = x_k^* - j\bar{q}$
If $x^- \in \Omega_k$**:**
$q_{k+1}^*(x^-) = \arg\min_{q \in \{q^+,\, q^+ - \bar{q}\}} \left\{ \epsilon_{k+1} |q| + \frac{\eta_{k+1}}{\tau_{k+1}} q^2 + l_k(x^- - q) \right\}$
Set $q^+ = q_{k+1}^*(x^-)$

$x^+ = x_k^* + j\bar{q}$
If $x^+ \in \Omega_k$**:**
$q_{k+1}^*(x^+) = \arg\min_{q \in \{q^-,\, q^- + \bar{q}\}} \left\{ \epsilon_{k+1} |q| + \frac{\eta_{k+1}}{\tau_{k+1}} q^2 + l_k(x^+ - q) \right\}$
$q^- = q_{k+1}^*(x^+)$

Output: $\forall k \in [\![0, m-1]\!] : q_{k+1}^* : \Omega_k \to \mathbb{R}; x_k \mapsto q_{k+1}^*(x_k^*)$.

3.6 Numerical results

We first present in Section 3.6.1 numerical results illustrating the gain in the objective value when considering volume uncertainty as part of the model. Then, in Section 3.6.2, we consider the problem of futures trading on the power market and compare the results obtained when using the dynamic coherent risk measure induced by the $\mathbb{E}$-CVaR$_{\alpha,\lambda}$ risk measure suggested in Section 3.2.5 against those obtained by using some of the other methods suggested in the literature.

3.6.1 Synthetic data

We apply our model to the test case initially provided by Almgren and Chriss (2001) and adapted by Vaes and Hauser (2022) in order to consider the uncertainty related to the volume target. The market parameters are fixed similarly to Almgren and Chriss (2001); Lin et al. (2015); Vaes and Hauser (2022): We assume a market with a single asset of initial price $S_0 = 50$, a median daily trading volume V of 5 million shares, and a minimum trading quantity $\bar{q}$ of 10^2 shares. For each trading period $k \in [\![1, m]\!]$, given the bid-ask spread b_k, we assume that the temporary impact of a trade is parametrised as follows: The fixed cost factor ϵ_k corresponds to half of the bid-ask spread b_k, *i.e.* $\epsilon_k = 0.5 b_k$. Additionally, a trader is assumed to incur a price impact equal to a bid-ask spread for every percent of the daily volume, *i.e.* $\eta_k = b_k/(0.01V)$ (Almgren and Chriss, 2001).

We now specify the uncertainty related to the market. The volatility σ of the asset is assumed constant and equal to 0.95. Each of the price shift ξ_k, $k \in [\![1, m]\!]$, follows a discrete distribution that corresponds to the discretization of the continuous zero mean Gaussian distribution of variance $\tau_k\sigma^2$, *i.e.* $\mathcal{N}(0, \tau_k\sigma^2)$. We denote this probability distribution as follows $\xi_k \overset{d=10^2}{\sim} \mathcal{N}(0, \tau_k\sigma^2)$, where $d = 10^2$ means that the continuous distribution is discretized using the 10^2-quantiles.

Concerning the volume to be traded, we assume that its initial forecast D_0 at t_s amounts to 0.5 million shares. For each trading period τ_k with $k \in [\![1, m-1]\!]$, we assume that the forecast update δ_k follows a discrete distribution that corresponds to the discretization of a zero mean Gaussian distribution of variance $\tau_k\nu^2$, where ν is equal to 10% of the initial volume target: $\nu := 0.1D_0$. We denote this probability distribution as follows $\delta_k \overset{d=10^2}{\sim} \mathcal{N}(0, \tau_k\nu^2)$.

With regard to the risk aversion of the trader, we assume that the parameters α_k and λ_k are constant for all stages. We present the results for different risk-level and risk-aversion parameters: $\alpha_k \in \{0.1, 0.25\}$ and $\lambda_k \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$.

Finally, we apply our model to two distinct bid-ask spread profiles:

(a) **Constant bid-ask spread.** The market liquidity remains constant throughout the trading period; similarly to Chapter 2, we assume a bid-ask spread b_k equal to 0.25% of the initial asset price S_0 for each trading period τ_k, *i.e.* $\forall k \in [\![1, m]\!] : b_k = 0.125$.

(b) **Market liquidity dries up.** The market suffers from lack of liquidity as the trading period τ_k gets closer to the delivery time t_f, which translates to a greater value for the bid-ask spread b_k. We assume that the asset has a bid-ask spread that increases linearly from 0.125 to 0.25 over the course of the execution period: the profile for the bid-ask spread over the 5 trading periods is thus given by $\boldsymbol{b} := [b_1, \dots, b_5]^{\mathrm{T}} = [0.125, 0.15625, 0.1875, 0.21875, 0.25]^{\mathrm{T}}$, which correspond to 0.25%, 0.3125%, 0.375%, 0.4375% and 0.5% of the initial asset price S_0.

Table 3.1 summarises the market details corresponding to this test case.

3.6.1.1 Objective value and trading cost comparison

Table 3.2 presents the objective value of Optimisation Problem (3.49) for all stages $k \in [\![0, m-1]\!]$ for $x_k = D_0$, where we have assumed that the parameters α_k and λ_k are constant and given in Table 3.1. Note that a similar analysis can be done for any level x_k, D_0 has been chosen as it is the estimate on the volume to be traded available at the initial stage.

By comparing the values of Tables 3.2a and 3.2b, we first observe the reduction in objective value when accounting for the uncertainty related to the volume to be traded while deciding of the trading quantities. The smaller the risk-aversion level α_k, and thus the more the trader focuses on the extreme trading cost realisations when evaluating the risk related to a trading strategy, the more pronounced this cost reduction. We also observe that the smaller α_k, or the greater λ_k, the larger the objective value. This is consistent with the intuition that the

TABLE 3.1: Market parameters

Acquisition time:	$T = 5$ [day]
Number of trading periods:	$m = 5$
Trading period length:	$\forall k \in [\![1, m]\!] : \tau_k = T/m = 1$ [day]
Initial asset price:	$S_0 = 50$ [\$/share]
Asset's volatility:	$\sigma = 0.95$ [(\$/share)/day$^{1/2}$]
Price shifts distribution:	$\forall k \in [\![1, m]\!] : \xi_k \overset{d=10^2}{\sim} \mathcal{N}(0, \tau_k \sigma^2)$
Daily volume:	$V = 5 \cdot 10^6$ [share]
Minimum trading quantity:	$\bar{q} = 10^2$ [share]
Initial volume target forecast:	$D_0 = 0.5 \cdot 10^6$ [share]
Volume forecast updates distribution:	$\forall k \in [\![1, m-1]\!] : \delta_k \overset{d=10^2}{\sim} \mathcal{N}(0, \tau_k \nu^2)$ with $\nu = 0.1 D_0$
Bid-ask spread:	Case (a): $\forall k \in [\![1, m]\!] : b_k = 0.125$ [\$/share]
	Case (b): $\boldsymbol{b} = [0.125, 0.15625, 0.1875, 0.21875, 0.25]$ [\$/share]
Fixed cost:	$\forall k \in [\![1, m]\!] : \epsilon_k = 0.5 b_k$ [\$/share]
Impact at 1% of market:	$\forall k \in [\![1, m]\!] : \eta_k = b_k/(0.01V)$ [(\$/share)/(share/day)]
CVaR$_{\alpha_k}$ parameter:	$\forall k \in [\![0, m-1]\!] : \alpha_k \in \{0.1, 0.25\}$
Risk aversion parameter:	$\forall k \in [\![0, m-1]\!] : \lambda_k \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$

smaller α_k is or the larger λ_k is, the more weight the trader attributes to high transaction costs in their risk estimation.

We now analyse the evolution of the objective function through the stages for a same volume left to be traded $x_k = D_0$. For a risk-neutral trader, we observe in Table 3.2b that the earlier the stage, the lower the objective function. This observation reflects the idea that the trader is insensitive to possible updates of the volume to be traded, and considers that the earlier the volume to be traded is estimated, the better they can spread its trade over the remaining trading periods in order to reduce the trading cost. This observation does not hold as the risk-aversion parameter λ_k increases. Indeed, the larger λ_k, the more the trader factors in the objective value the risk of a change in the volume to be traded. Therefore, this anticipation of the change in the quantity to be traded (*e.g.* the fear of having an increase in the volume to be traded) justifies why an exclusively risk-focused trader ($\lambda_k = 1.0$) evaluates, for instance, that $f_1(D_0) > f_3(D_0)$; the values for stage 1 and 3 respectively are given in the last line of Table 3.2.

In Figure 3.3, we represent the probability density function (pdf) of the trading cost in Case (a) for the risk-level of $\alpha_k = 0.25$ (similar plots are obtained for Case (b) and/or $\alpha_k = 0.1$). For any risk-aversion parameter λ_k, we observe that the mass of the probability density is shifted towards the left (small trading cost) when considering the uncertainty of the volume to be traded while solving Optimisation Problem (3.49) in the Dynamic Program (PVU) framework compared against the case where it is ignored.

TABLE 3.2: Objective value of Optimisation Problem (3.49) for different risk-aversion parameters values α_k and λ_k for $k \in [\![0, m-1]\!]$ and $x_k = D_0$.

(a) Dynamic Program (PU): the trader does not anticipate the uncertainty related to the volume to be trader and considers that the only source of uncertainty is the price dynamics.

Risk Aversion λ_k	$\alpha_k = 0.25$					$\alpha_k = 0.1$				
	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
0.00	1.692e5	1.988e5	2.488e5	3.499e5	6.562e5	1.692e5	1.988e5	2.488e5	3.499e5	6.562e5
0.25	3.933e5	3.839e5	3.840e5	4.232e5	6.562e5	4.482e5	4.348e5	4.257e5	4.484e5	6.562e5
0.50	5.141e5	4.985e5	4.836e5	4.869e5	6.562e5	5.839e5	5.631e5	5.423e5	5.288e5	6.562e5
0.75	5.947e5	5.754e5	5.561e5	5.409e5	6.562e5	6.715e5	6.446e5	6.177e5	5.909e5	6.562e5
1.00	6.537e5	6.309e5	6.080e5	5.853e5	6.562e5	7.340e5	7.010e5	6.680e5	6.350e5	6.562e5

(b) Dynamic Program (PVU): the trader considers that the uncertainty arises from both the price dynamics and the volume to be traded.

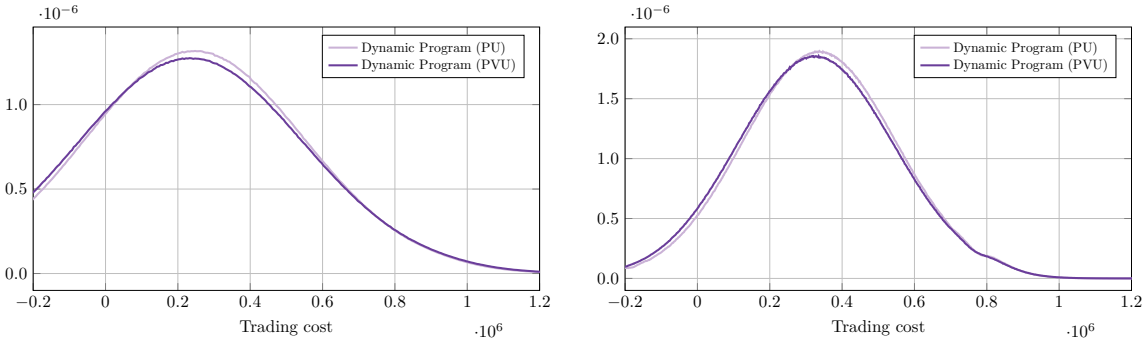
Risk Aversion λ_k	$\alpha_k = 0.25$					$\alpha_k = 0.1$				
	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
0.00	1.692e5	1.988e5	2.488e5	3.499e5	6.562e5	1.692e5	1.988e5	2.488e5	3.499e5	6.562e5
0.25	3.920e5	3.836e5	3.840e5	4.232e5	6.562e5	4.456e5	4.334e5	4.257e5	4.484e5	6.562e5
0.50	5.113e5	4.969e5	4.834e5	4.868e5	6.562e5	5.773e5	5.582e5	5.391e5	5.287e5	6.562e5
0.75	5.895e5	5.714e5	5.532e5	5.409e5	6.562e5	6.642e5	6.389e5	6.135e5	5.908e5	6.562e5
1.00	6.484e5	6.266e5	6.047e5	5.853e5	6.562e5	7.288e5	6.973e5	6.658e5	6.349e5	6.562e5

3.6.1.2 Optimal trading quantity

In this section we analyse how the quantity to be traded is impacted by the risk-averseness λ_k and the stage k for the risk-level $\alpha_k = 0.25$; similar conclusions are obtained for $\alpha_k = 0.1$.

First, as previously observed in Almgren and Chriss (2001), we observe in Figures 3.4 to 3.7 that a trader who only takes into account the uncertainty in the price process (Dynamic Program (PU)), front loads trades to avoid the risk related to an increase in price. The more risk-averse the trader, the faster the positions are traded. However, once the trader factors in the uncertainty related to the volume to be traded (Dynamic Program (PVU)), the front loading is reduced. This matches with the following observation made in Chapter 2: a risk-averse trader prefers to partially delay their trade and wait for a more accurate estimate on the volume to be traded. By doing so, they avoid over-trading. We can indeed observe on Figure 3.5, that the volume left to be traded in the case of the method Dynamic Program (PVU) is greater compared to the method Dynamic Program (PU).

The liquidity profile has also an impact on the speed at which the trades are executed. In Case (b) the market is assumed more liquid during the first trading periods compared to the last ones. It is therefore preferable to execute more transactions during these early periods, as they are less subject to temporary impact. This is observed by comparing the volume traded during the first trading period in the situations (a) and (b) which is illustrated in Figures 3.4 and 3.6 respectively: the quantity traded during τ_1 is significantly greater in case (b) compared



(a) $\lambda_k = 0.25$ (b) $\lambda_k = 0.5$

(c) $\lambda_k = 0.75$ (d) $\lambda_k = 1.0$

FIGURE 3.3: Comparison of the pdfs of the trading cost when the uncertainty related to the total volume to be traded is or is not taken into account when solving Optimisation Problem (3.49) (the pdfs are estimated based on 10^8 realisations of the underlying processes).

to case (a). The profile of the volume left to be traded in both situations is illustrated in Figures 3.5 and 3.7. Hence, if the liquidity of the market dries out, then the rate at which the trades are executed is increased compared to the situation where the market liquidity stays stable.

Finally, we analyse the optimal quantity to be traded for a given volume x_k according to the stages. The optimal quantity to be traded q^*_{k+1} given the volume left to be traded x_k is represented in Figures 3.8 and 3.9 for the situation (a) and (b) respectively. This quantity q^*_{k+1} corresponds to the solution of Optimisation Problem (3.49) for $x = x_k$. The last row of both figures represents the difference in the trading quantities between the optimal quantity obtained by solving (3.49) using the method Dynamic Program (PVU) with the optimal quantity obtained when using method Dynamic Program (PU). In other words, it represents how the strategy under PU must be adapted to obtain the optimal strategy under PVU.

On the one hand, we observe that for small estimates on the volume left to be traded, *i.e.* $x_k < 3 \cdot 10^5$, the trader mitigates the strategy under PU in the sense that $\left|q_k^{*,\mathrm{PVU}}\right| \leq \left|q_k^{*,\mathrm{PU}}\right|$, where $q_k^{*,\mathrm{PU}}$ (*resp.* $q_k^{*,\mathrm{PVU}}$) is the optimal trading quantity under PU (*resp.* PVU). This is explained by the fact that, for small values of x_k, the quantities traded are relatively small and does not suffer as much of the temporary impact as large trades. Consequently, the major

risk is to over-trade and to have to trade out this volume excess over the remaining periods. This explains why a risk-averse trader prefers to decrease the size of their next period order and to wait for a better estimate on D_{t_f}.

On the other hand, for large values of x_k, the major risk is to under-trade and to have to trade large quantities over the remaining periods, which goes hand in hand with a large additional cost due to the temporary impact. In that case, if $x > 0$ (*resp.* $x < 0$), the optimal strategy under PVU anticipates any increase (*resp.* decrease) in the estimate of D_{t_f}, and prefers to increase (*resp.* decrease) the quantity traded on the next trading period. In this way, the additional cost due to the temporary impact of a potential increase (*resp.* decrease) in the estimate of D_{t_f} would be split over the next trading period and the remaining ones rather than just the latter.

Finally, we note that, the later the stage, the sooner the risk to under-trade dominates the risk to over-trade. Indeed, the later the stage, the smaller the value of $|x|$ for which $\left|q_k^{*,\mathrm{PVU}}\right|$ becomes greater than $\left|q_k^{*,\mathrm{PU}}\right|$. This is explained by the fact that they are fewer remaining trading periods to absorb any potential increase in the estimate of D_{t_f}, as well as less change that this estimate later decreases due to a future volume update.

3.6.2 Real world application: power trading

Similarly to the model presented in Chapter 2, the dynamic program presented in this chapter to solve the liquidation problem when being subject to volume uncertainty does not rely on any particular assumptions on the distributions of the price moves and forecast updates, except the fact that they are martingales and independent. Consequently, the dynamic program approach can be applied to a wider range of applications compared to the models of Almgren and Chriss (2001) and Cheng et al. (2017). We now apply the dynamic programming approach presented in this chapter to the power trading problem introduced in Section 2.3.4. In this problem, power wholesalers (*resp.* producers) trade power future contracts in the intra-day market with the objective of achieving, for each half hourly settlement period, a contractual net position that matches the expected power demand of their end-consumers (*resp.* power generation).

We make here similar assumptions as in Chapter 2 and consider power trading in Great Britain (GB): Any power power wholesaler has access to the intra-day market, which opens 48 half-hours and closes one hour before the start of delivery. For any given settlement period starting at time t_f, the wholesaler aims to achieve a final net position close to the realised power demand D_{t_f} of their end-consumers, which we assume to be equal to $\rho = 1/3$ of the total power demand of the population V_{t_f}. We assume that the wholesaler is risk-averse and desired to minimise the risk related to their trading cost.

The market is modelled as follows (see Figure 3.10): the intra-day market is subdivided in 46 one half-hour long trading periods and spans from 48 to 2 half-hours ahead of delivery time. We denote the start of this trading window by t_s, which corresponds to 48 half-hours ahead t_f. Once the intra-day market closes, power wholesalers have a last chance to trade power on the real time market, which is modelled by adding a virtual 47th trading period. To incentivise traders

to trade the entirety of their position by the closure of the intra-day market, we penalise any outstanding volume by fixing a large temporary impact parameter η_{47}. There is consequently no obligation for a market participant to balance their contractual positions against the expected demand. However, the exposure to the price of the real-time market drives the incentive to avoid any residual volume.

Concerning the market parameters and the price shifts distributions, we use that same values as in Section 2.3.4 and they are reported in Table 3.3.

With regard to the distribution of the forecast updates, we assume that the average realised national power demand $\bar{V}_{t_f}$ equals $3 \cdot 10^4$MW, that $\bar{D}_{t_f} = \rho \bar{V}_{t_f} = 10^4$MW and that the initial position equals 0. We estimate the forecast updates of each half-hour period based on the operational data published by the *Balancing Mechanism Reporting Service* (BMRS). For each period τ_k, $k \in [\![1, 45]\!]$, we assume that the forecast update distribution is as follows: Firstly, there is a probability $\hat{p}_k$ of having a forecast update. Secondly, conditional on a forecast update occurring, the forecast update δ_k is assumed to follow the zero-mean normal distribution $\mathcal{N}\left(0, \hat{s}_k^2 \bar{D}_{t_f}^2\right)$, where $\hat{s}_k$ is the estimate of the standard deviation of the forecast update between stage $k-1$ and k, *i.e.* $\left(D_k - D_{k-1}\right)/D_{t_f}$. Finally, we assume that the last forecast update is released during τ_{46}, and that it follows the normal distribution $\delta_{46} \overset{d=10^2}{\sim} \mathcal{N}\left(0, (\hat{s}_{45}^{err})^2 \bar{D}_{t_f}^2\right)$, where $\hat{s}_{45}^{err}$ is the estimate of the forecast error at stage 45, *i.e.* $\left(D_{t_f} - D_{45}\right)/D_{t_f}$.

Figure 3.11 represents the estimate of the necessary values in order to model the forecast update distributions based on 5 years of data on the national power demand in GB: Figure 3.11a represents the estimated probability $\hat{p}_k$ of having a forecast update of the power demand at time t_k. Conditional on there being a forecast update, Figure 3.11b represents the estimated standard deviation $\hat{s}_k$ of the relative forecast update.

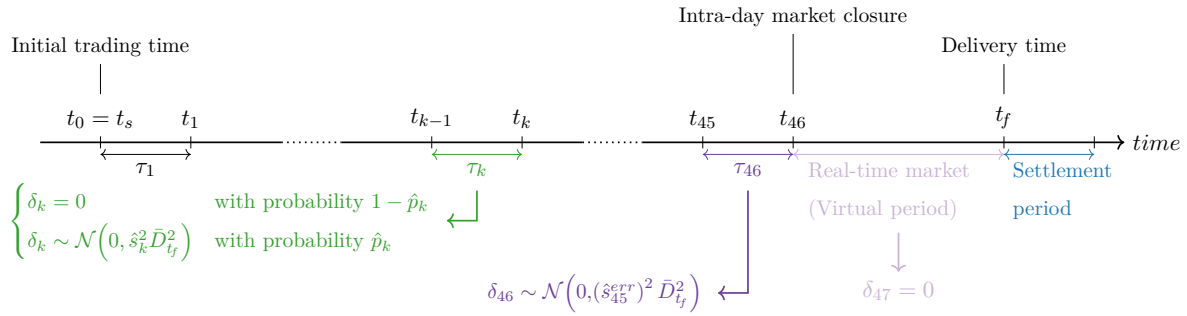
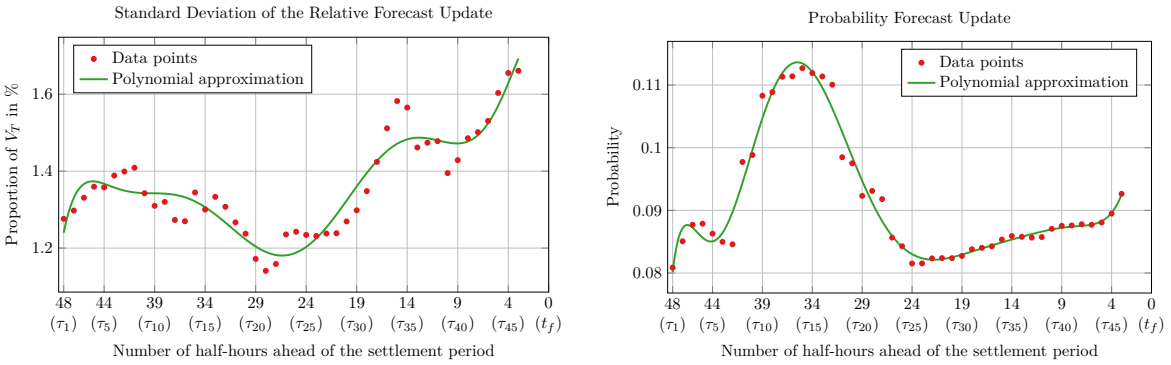


FIGURE 3.10: Power trading time line and modelling assumptions.

TABLE 3.3: Power market parameters

Number of trading periods:	$m = 47$
Trading period length:	$\forall k \in [\![1, m]\!] : \tau_k = \frac{1}{48}$ [day]
Initial future contract price:	$S_0 = 90$ [£/MW]
Asset's volatility:	$\sigma = 0.9$ [(£/MW)/day$^{1/2}$]
Price shifts distribution:	$\forall k \in [\![1, m]\!] : \xi_k \overset{d=10^2}{\sim} \mathcal{N}(0, \tau_k \sigma^2)$
Average realised national power demand:	$\bar{V}_{t_f} = 3 \cdot 10^4$ [MW]
Proportion of the population supplying:	$\rho = \frac{1}{3}$
Minimum exchangeable volume:	$\bar{q} = 1$ [MW]
Power demand initial estimate:	$D_0 = \bar{D}_{t_f} = \rho \bar{V}_{t_f} = 10^4$ [MW]
Volume forecast updates distribution:	$\forall k \in [\![1, 45]\!] : \delta_k \overset{d=10^2}{\sim} \mathcal{N}\left(0, \hat{s}_k^2 \bar{D}_{t_f}^2\right)$
	$\delta_{46} \overset{d=10^2}{\sim} \mathcal{N}\left(0, (\hat{s}_{45}^{err})^2 \bar{D}_{t_f}^2\right)$
	$\delta_{47} = 0$
Bid-ask spread:	$\forall k \in [\![1, m]\!] : b_k = 0.25\% \, S_0 = 0.225$ [£/MW]
Fixed cost:	$\forall k \in [\![1, m]\!] : \epsilon_k = 0.5 b_k = 0.1125$ [£/MW]
Impact at 2% of market:	$\forall k \in [\![1, 46]\!] : \eta_k = b_k / (0.02 \bar{V}_{t_f}) = 3.75 \cdot 10^{-5}$ [(£/MW)/(MW/day)]
	$\eta_{47} = 3 b_{47} / (0.02 \bar{V}_{t_f}) = 11.25 \cdot 10^{-5}$ [(£/MW)/(MW/day)]
Permanent impact parameter:	$\forall k \in [\![1, m]\!] : \gamma_k = 0$ [£/MW^2]
CVaR$_\alpha$ parameter (Mean-CVaR$_\alpha$):	$\alpha = 0.05$
CVaR$_{\alpha_k}$ parameter (Dynamic Program):	$\alpha_k = 0.6$



(a) (b)

FIGURE 3.11: Analysis of the forecast updates of the national power demand in GB.

To compare our approach with the one of Chapter 2, we use the parameter value $\alpha = 0.05$ in their Mean-CVaR$_\alpha$ model. In the dynamic program, it is not easy to find values for α_k that would translate an equivalent risk-level as for the Mean-CVaR$_\alpha$ model. Indeed, in the model suggested in Chapter 2, risk is measured by evaluating the *Conditional Value-at-Risk* of the resulting trading cost, whereas in the dynamic program we successively assess risk by evaluating the *Conditional Value-at-Risk* of the stage cost. Therefore, the resulting risk-level of the dynamic programming approach is intricate to estimate as it results of an composition of successive risk measures. Unlike the risk-level in the Mean-CVaR$_\alpha$ model, the risk-level of the dynamic approach depends on the number of trading periods since for each period we estimate

risk by taking the expectation of a certain percentage of the worst case realisations. Here, in order to make a comparison, we assume a constant risk-level for the dynamic program of $\alpha_k = 0.6$.

A similar analysis to that of the synthetic data can be carried out in terms of objective value (see Table 3.4) and changes in trading strategies when the uncertainty of the volume to be traded is taken into account. Table 3.5 compares the Mean-CVaR$_\alpha$ framework of Chapter 2 with the Dynamic Program framework stated in this work. We observe that the Mean-CVaR$_\alpha$ (PVU) is the method that leads to the best Mean-CVaR$_{0.05}$ risk-return tradeoff. This is not surprising since, contrarily to the Mean-CVaR$_\alpha$ framework, the dynamic program presented in this work does not measure risk in the same manner and consequently does not have the same objective function. Nevertheless, we observe that the Dynamic Program (PVU) framework gives better risk-return trade-offs compared to the case where the uncertainty of the volume target is not taken into account (Dynamic Program (PU)). A comparison of the pdfs of the trading cost for the Mean-CVaR$_{0.05}$ (PVU) and the Dynamic Program (PVU) frameworks is illustrated in Figure 3.12: we observe that as risk-aversion grows, the expectation of the trading cost grows at a slower rate for the Dynamic Program (PVU), which reflects that the composite dynamic $\mathbb{E}$-CVaR$_{\alpha_k,\lambda_k}$ focus more on the contribution of the expectation of the expected cost compared to the Mean-CVaR$_{0.05}$ risk measure. It is laborious to draw any other conclusions about the comparison of the frameworks because they measure risk differently and there is no straightforward equivalence relationship between the risk-aversion parameters that would lead to a meaningful comparison. In conclusion, which framework is preferred depends on how risk is measured, no conclusions can be drawn as to whether one method performs better than the other. However, we should mention that, unlike the Mean-CVaR$_\alpha$ framework, the optimal strategy obtained via the dynamic program is consistent over time, any future recourse action would lead to the same optimal strategy.

TABLE 3.4: Power trading: objective value of Optimisation Problem (3.49) for $x_k = D_0$.

(a) Dynamic Program (PU): the trader does not anticipate the uncertainty related to the volume to be trader and considers that the only source of uncertainty is the price dynamics.

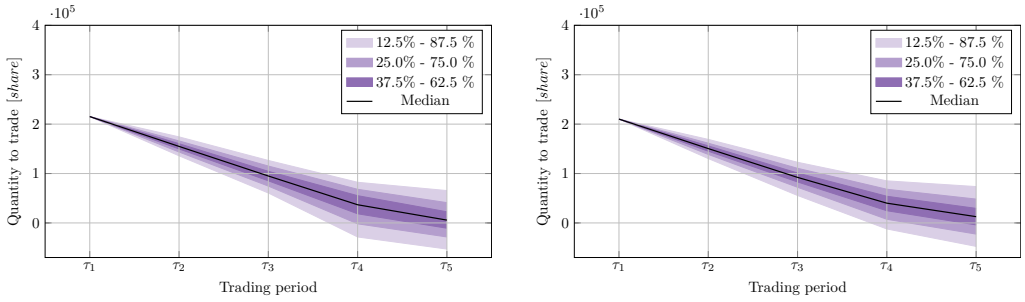
λ_k	Stage 0	Stage 5	Stage 10	Stage 15	Stage 20	Stage 25	Stage 30	Stage 35	Stage 40	Stage 45
0.00	4.282e5	4.752e5	5.351e5	6.141e5	7.231e5	8.832e5	1.141e6	1.627e6	2.881e6	1.353e7
0.25	4.838e5	5.260e5	5.811e5	6.553e5	7.597e5	9.158e5	1.170e6	1.654e6	2.912e6	1.363e7
0.50	5.384e5	5.761e5	6.266e5	6.962e5	7.962e5	9.482e5	1.199e6	1.682e6	2.942e6	1.373e7
0.75	5.920e5	6.256e5	6.718e5	7.369e5	8.325e5	9.807e5	1.228e6	1.709e6	2.973e6	1.382e7
1.00	6.448e5	6.744e5	7.165e5	7.772e5	8.687e5	1.013e6	1.258e6	1.736e6	3.005e6	1.392e7

(b) Dynamic Program (PVU): the trader considers that the uncertainty arises from both the price dynamics and the volume to be traded.

λ_k	Stage 0	Stage 5	Stage 10	Stage 15	Stage 20	Stage 25	Stage 30	Stage 35	Stage 40	Stage 45
0.00	4.282e5	4.752e5	5.351e5	6.141e5	7.231e5	8.832e5	1.141e6	1.627e6	2.881e6	1.353e7
0.25	4.837e5	5.259e5	5.81e5	6.551e5	7.595e5	9.155e5	1.170e6	1.654e6	2.911e6	1.363e7
0.50	5.381e5	5.757e5	6.262e5	6.957e5	7.955e5	9.473e5	1.198e6	1.679e6	2.940e6	1.372e7
0.75	5.914e5	6.249e5	6.709e5	7.358e5	8.311e5	9.787e5	1.225e6	1.704e6	2.967e6	1.382e7
1.00	6.438e5	6.733e5	7.151e5	7.755e5	8.664e5	1.009e6	1.252e6	1.728e6	2.993e6	1.391e7

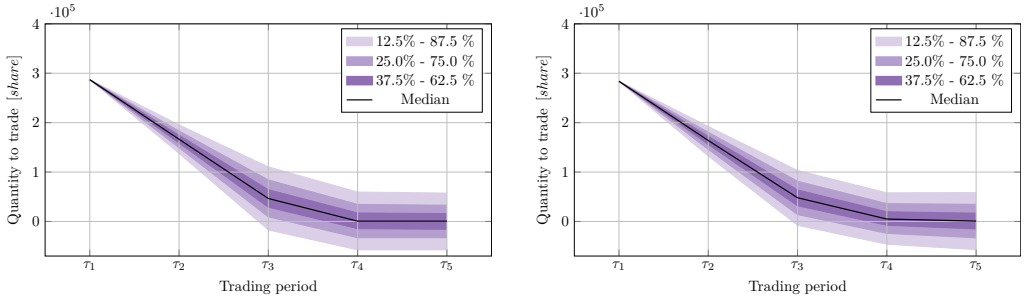
3.7 Conclusion

In this chapter we considered the specific optimal trade execution problem where, unlike what is usually assumed in the existing literature, the required trade volume is uncertain but forecasts with increasing accuracy become available throughout the trading period. We have formulated this problem as a dynamic program and have demonstrated that the size of the scenario tree associated to this dynamic program can be significantly reduced owing to the fact that, at each stage, the decision on the next action to take is only function of the best estimate on the volume left to be traded. This enables to circumvent the combinatorial explosion traditionally associated with dynamic programming approaches. Moreover, we show that a trader has no interest in over-trading, which leads to a further reduction of the size of the scenario tree. We proved that the value function f_k is convex in x_k, which is useful to simplify the search of the optimal decision at each stage k; see Proposition 3.7 and Corollary 3.3, which show that a large part of the feasible set can be pruned given the optimal decision of another state of the same stage. We also demonstrated that the optimal solution is time-consistent. Finally, we showed that the optimal strategies obtained with this model produce similar trading costs in comparison to the other approaches in the literature.



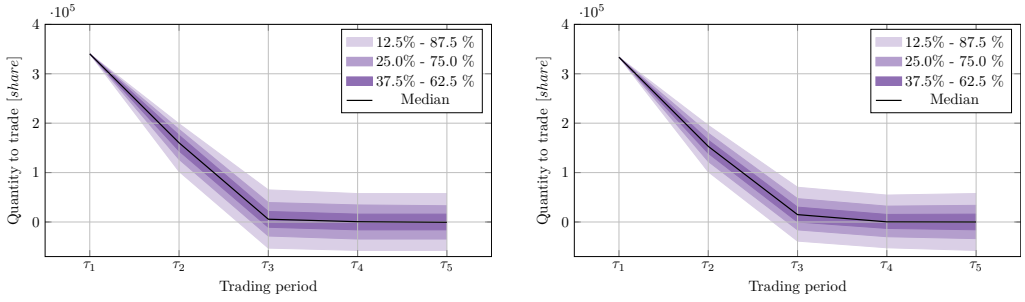
(a) Dynamic Program (PU): $\lambda_k = 0.25$

(b) Dynamic Program (PVU): $\lambda_k = 0.25$



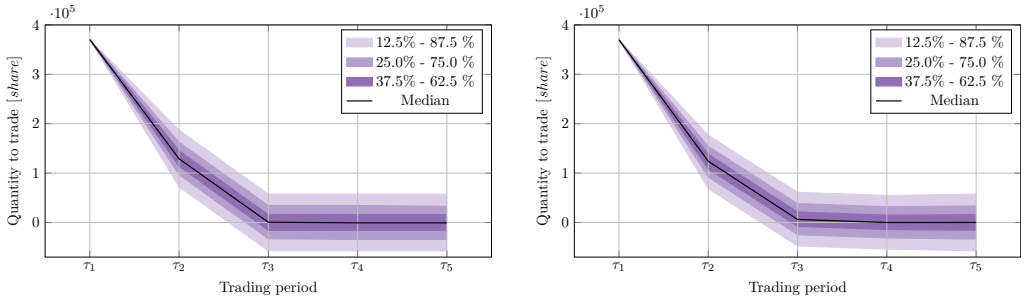
(c) Dynamic Program (PU): $\lambda_k = 0.50$

(d) Dynamic Program (PVU): $\lambda_k = 0.50$



(e) Dynamic Program (PU): $\lambda_k = 0.75$

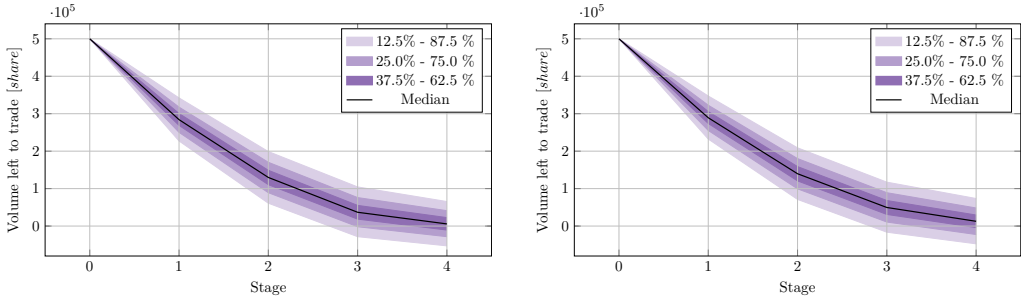
(f) Dynamic Program (PVU): $\lambda_k = 0.75$



(g) Dynamic Program (PU): $\lambda_k = 1.00$

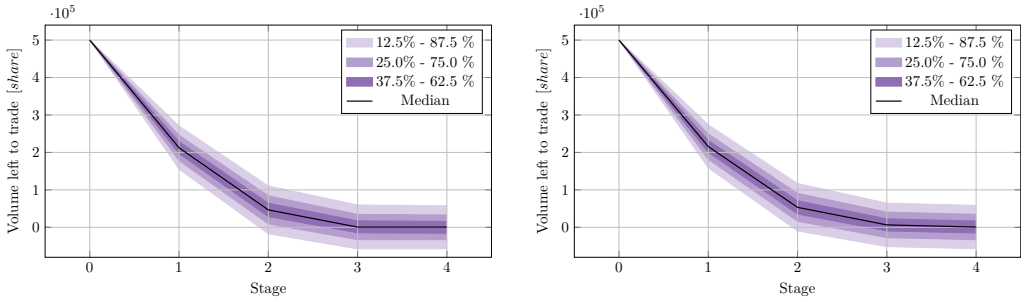
(h) Dynamic Program (PVU): $\lambda_k = 1.00$

FIGURE 3.4: Case (a) with $\alpha_k = 0.25$: Median of the quantity to be traded together with the 25%, 50% and 75% confidences intervals around it (10^6 random trading paths were generated to estimate the percentiles).



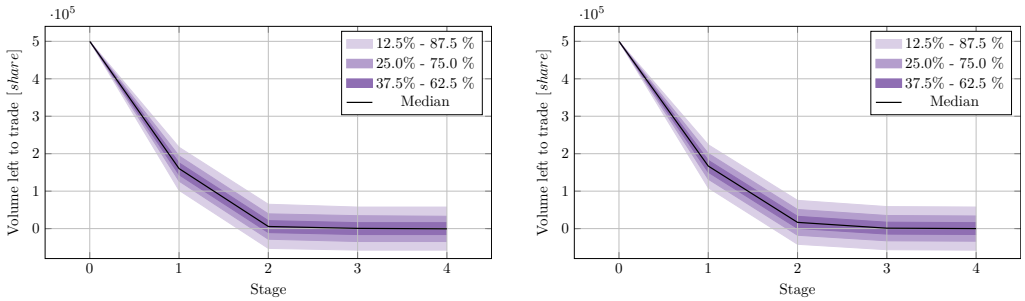
(a) Dynamic Program (PU): $\lambda_k = 0.25$

(b) Dynamic Program (PVU): $\lambda_k = 0.25$



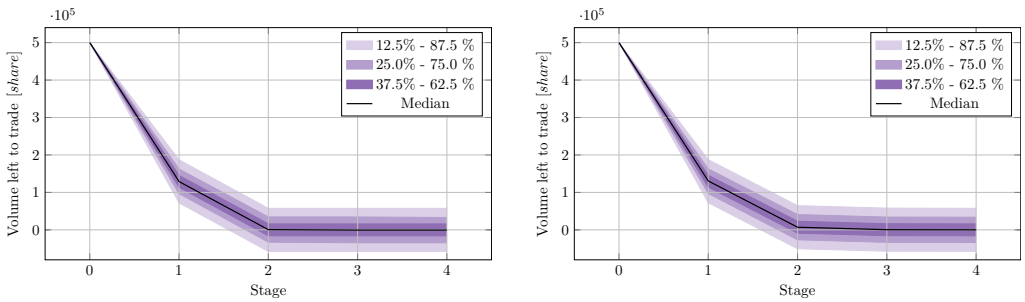
(c) Dynamic Program (PU): $\lambda_k = 0.50$

(d) Dynamic Program (PVU): $\lambda_k = 0.50$



(e) Dynamic Program (PU): $\lambda_k = 0.75$

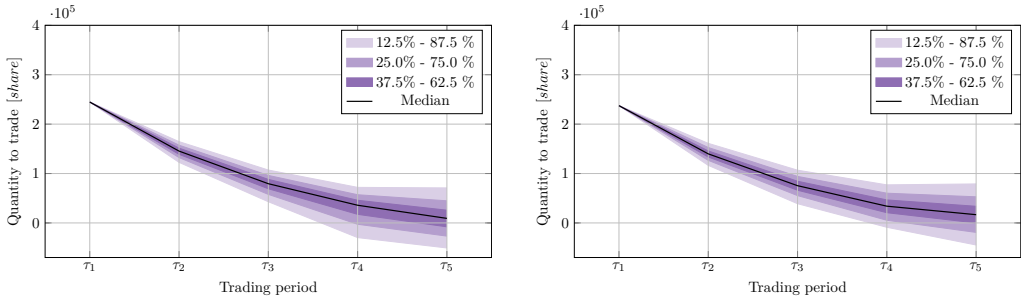
(f) Dynamic Program (PVU): $\lambda_k = 0.75$



(g) Dynamic Program (PU): $\lambda_k = 1.00$

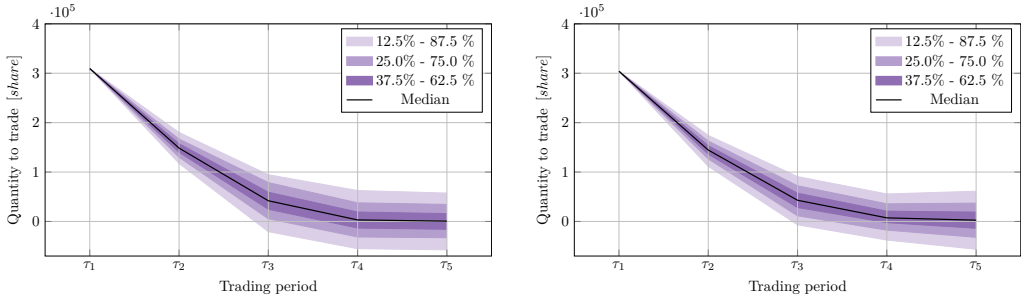
(h) Dynamic Program (PVU): $\lambda_k = 1.00$

FIGURE 3.5: Case (a) with $\alpha_k = 0.25$: Median of the volume left to be traded together with the 25%, 50% and 75% confidences intervals around it (10^6 random trading paths were generated to estimate the percentiles).



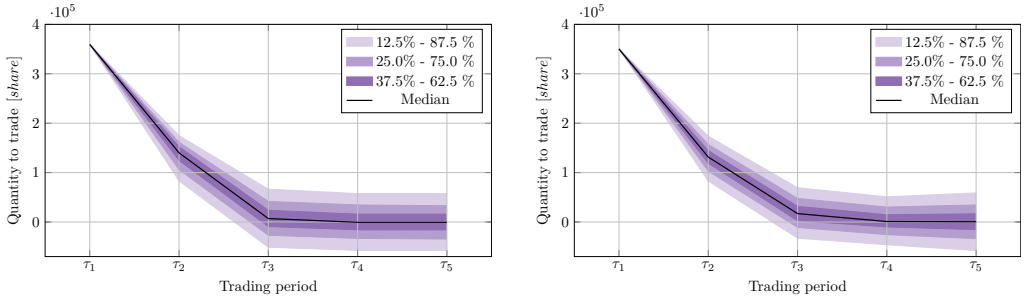
(a) Dynamic Program (PU): $\lambda_k = 0.25$

(b) Dynamic Program (PVU): $\lambda_k = 0.25$



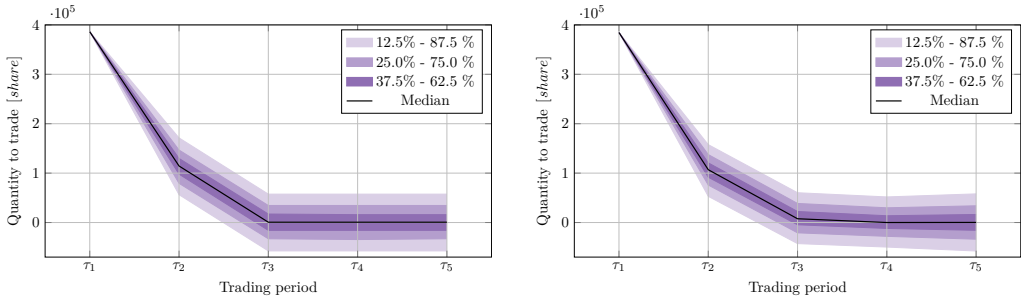
(c) Dynamic Program (PU): $\lambda_k = 0.50$

(d) Dynamic Program (PVU): $\lambda_k = 0.50$



(e) Dynamic Program (PU): $\lambda_k = 0.75$

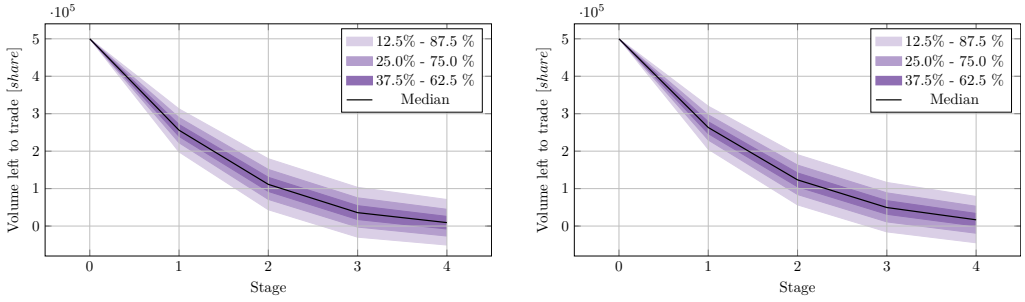
(f) Dynamic Program (PVU): $\lambda_k = 0.75$



(g) Dynamic Program (PU): $\lambda_k = 1.00$

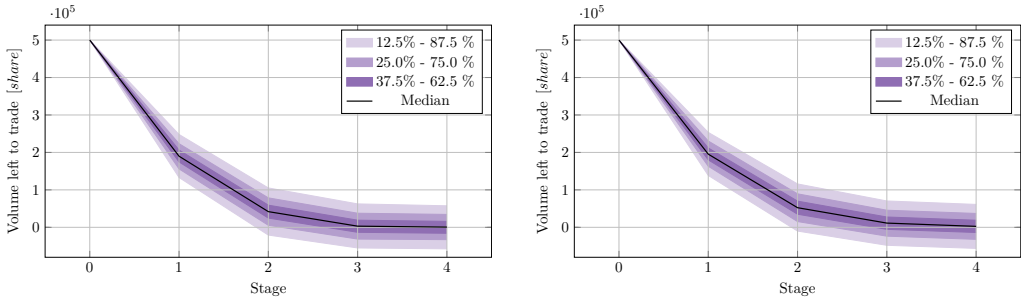
(h) Dynamic Program (PVU): $\lambda_k = 1.00$

FIGURE 3.6: Case (b) with $\alpha_k = 0.25$: Median of the quantity to be traded together with the 25%, 50% and 75% confidences intervals around it (10^6 random trading paths were generated to estimate the percentiles).



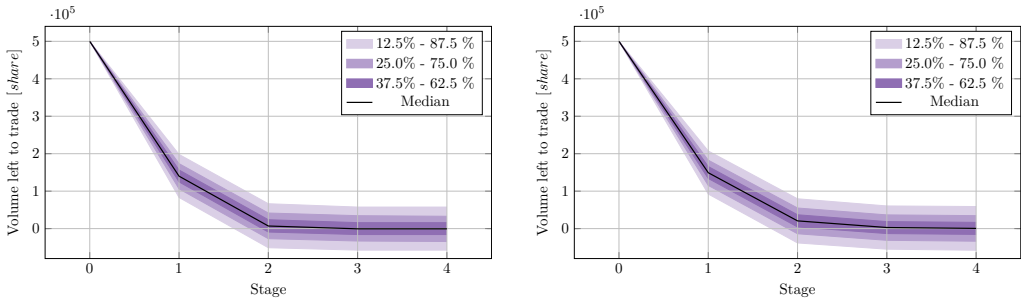
(a) Dynamic Program (PU): $\lambda_k = 0.25$

(b) Dynamic Program (PVU): $\lambda_k = 0.25$



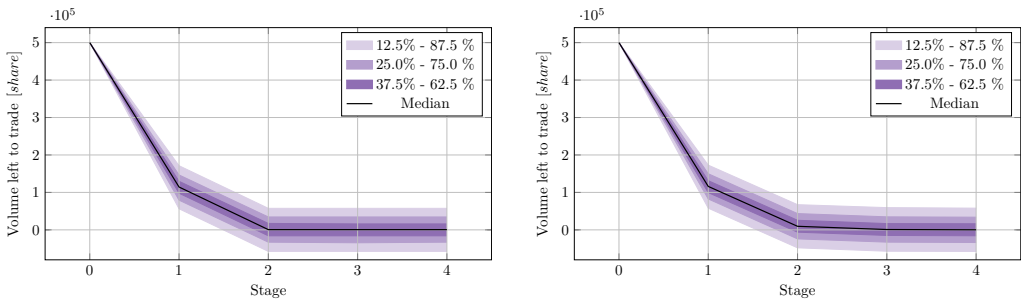
(c) Dynamic Program (PU): $\lambda_k = 0.50$

(d) Dynamic Program (PVU): $\lambda_k = 0.50$



(e) Dynamic Program (PU): $\lambda_k = 0.75$

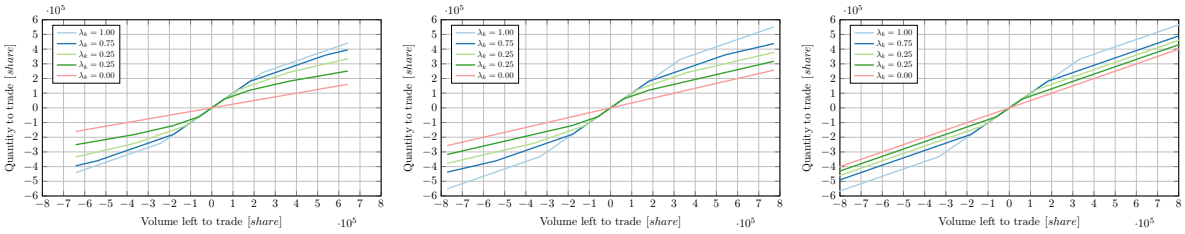
(f) Dynamic Program (PVU): $\lambda_k = 0.75$



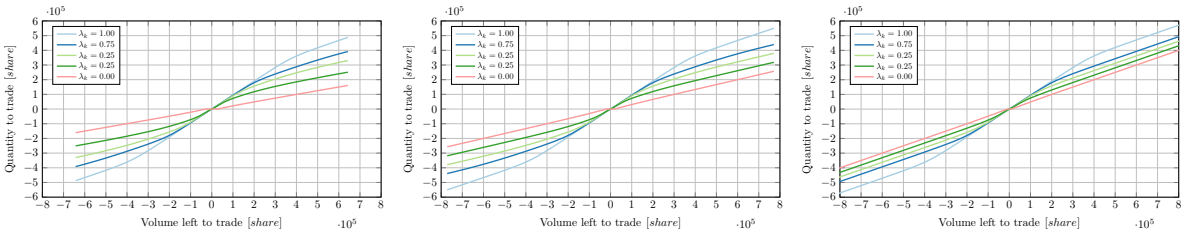
(g) Dynamic Program (PU): $\lambda_k = 1.00$

(h) Dynamic Program (PVU): $\lambda_k = 1.00$

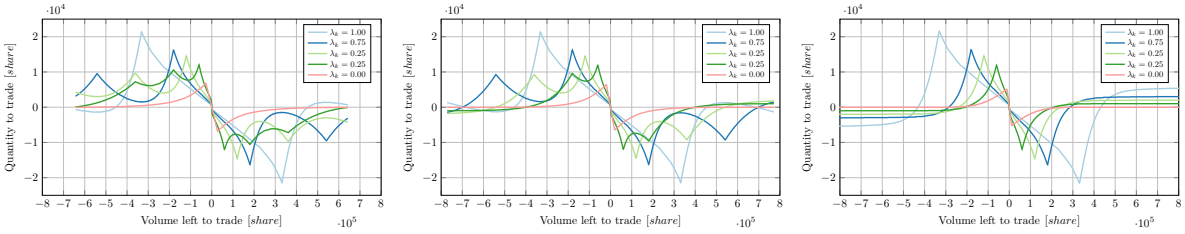
FIGURE 3.7: Case (b) with $\alpha_k = 0.25$: Median of the volume left to be traded together with the 25%, 50% and 75% confidences intervals around it (10^6 random trading paths were generated to estimate the percentiles).



(a) Dyn. Prog. (PU): Stage 1 (b) Dyn. Prog. (PU): Stage 2 (c) Dyn. Prog. (PU): Stage 3

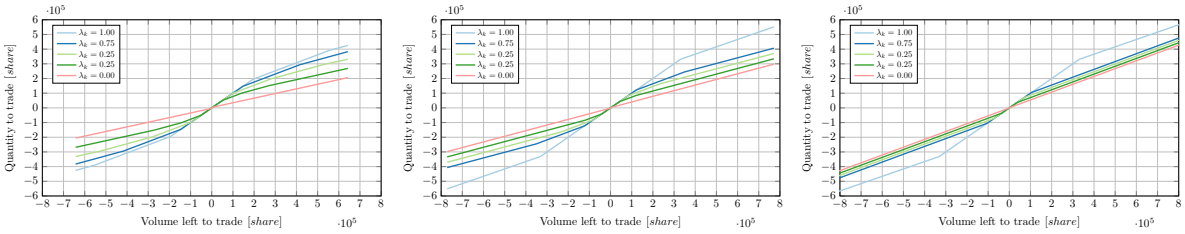


(d) Dyn. Prog. (PVU): Stage 1 (e) Dyn. Prog. (PVU): Stage 2 (f) Dyn. Prog. (PVU): Stage 3

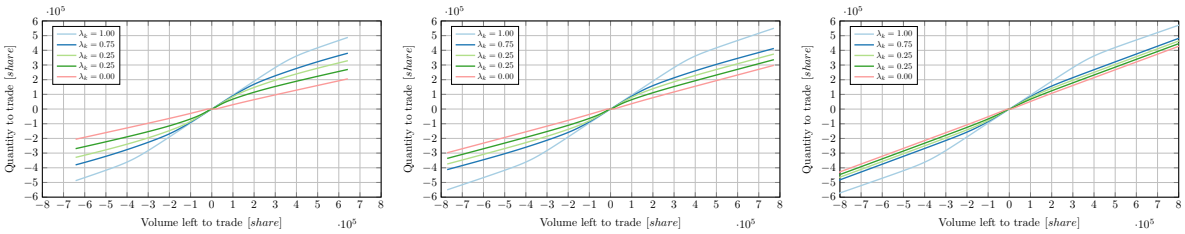


(g) Dyn. Prog. (PVU): Stage 1 (h) Dyn. Prog. (PVU): Stage 2 (i) Dyn. Prog. (PVU): Stage 3

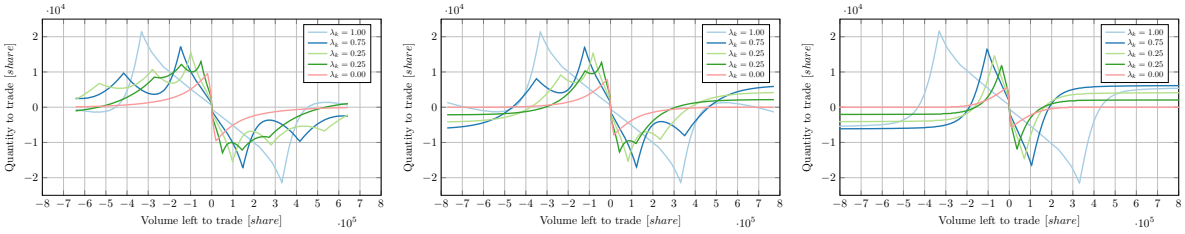
FIGURE 3.8: Case (a) with $\alpha_k = 0.25$: Optimal trading quantity q^*_{k+1} given the volume left to be traded x_k for any stage $k \in [\![1, 3]\!]$.



(a) Dyn. Prog. (PU): Stage 1 (b) Dyn. Prog. (PU): Stage 2 (c) Dyn. Prog. (PU): Stage 3



(d) Dyn. Prog. (PVU): Stage 1 (e) Dyn. Prog. (PVU): Stage 2 (f) Dyn. Prog. (PVU): Stage 3



(g) Dyn. Prog. (PVU): Stage 1 (h) Dyn. Prog. (PVU): Stage 2 (i) Dyn. Prog. (PVU): Stage 3

FIGURE 3.9: Case (b) with $\alpha_k = 0.25$: Optimal trading quantity q^*_{k+1} given the volume left to be traded x_k for any stage $k \in [\![1, 3]\!]$.

TABLE 3.5: Trading cost of the optimal trading strategies in the Mean-CVaR$_{0.05}$ and the Dynamic Program frameworks when considering or not the uncertainty related to the volume to be traded.

(a) Mean-CVaR$_{0.05}$ (PU)

λ_{CVaR} [1]	Expectation [£]	CVaR$_{0.1}$ [£]	CVaR$_{0.05}$ [£]	CVaR$_{0.01}$ [£]	Variance [£^{2}]	Mean-CVaR$_{0.05}$ [£]
0.00	4.286e5	1.078e6	1.207e6	1.470e6	1.231e11	4.286e5
0.25	4.285e5	1.077e6	1.206e6	1.469e6	1.228e11	6.230e5
0.50	4.288e5	1.077e6	1.206e6	1.468e6	1.225e11	8.175e5
0.75	4.294e5	1.076e6	1.205e6	1.467e6	1.221e11	1.011e6
1.00	4.300e5	1.077e6	1.205e6	1.468e6	1.219e11	1.205e6

(b) Mean-CVaR$_{0.05}$ (PVU)

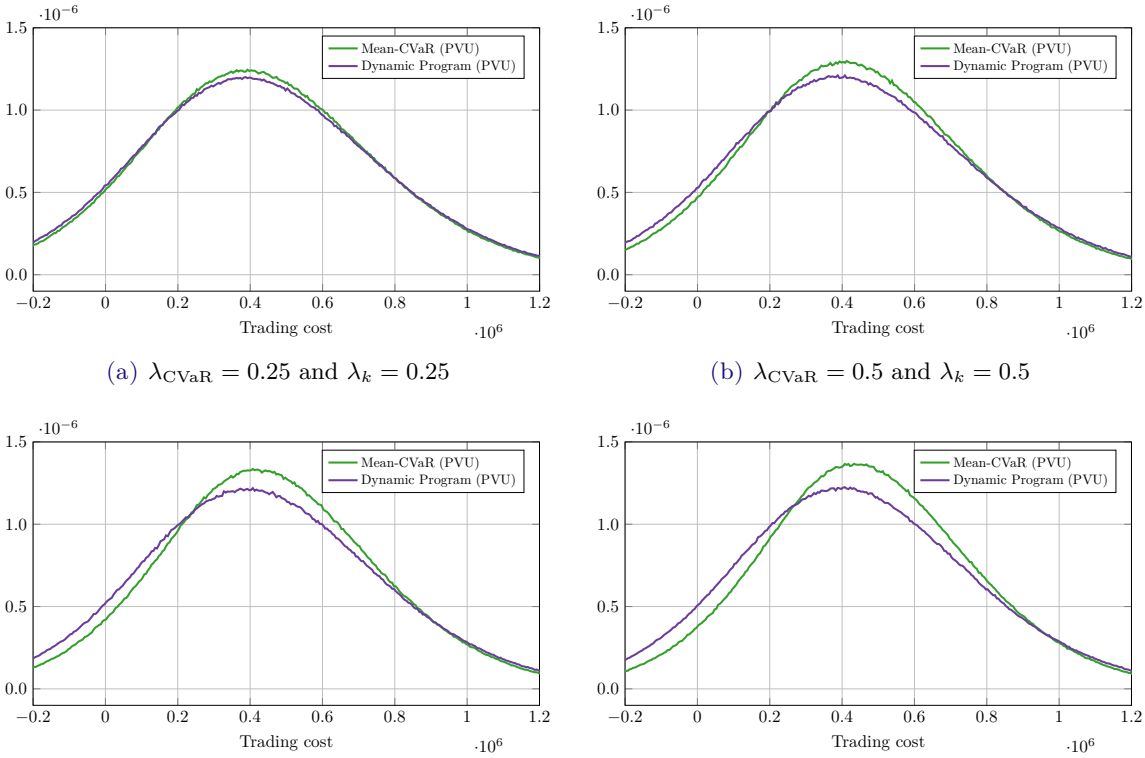
λ_{CVaR} [1]	Expectation [£]	CVaR$_{0.1}$ [£]	CVaR$_{0.05}$ [£]	CVaR$_{0.01}$ [£]	Variance [£^{2}]	Mean-CVaR$_{0.05}$ [£]
0.00	4.284e5	1.078e6	1.207e6	1.469e6	1.231e11	4.284e5
0.25	4.323e5	1.055e6	1.179e6	1.433e6	1.137e11	6.192e5
0.50	4.421e5	1.043e6	1.163e6	1.409e6	1.062e11	8.027e5
0.75	4.549e5	1.038e6	1.155e6	1.396e6	1.007e11	9.805e5
1.00	4.699e5	1.039e6	1.153e6	1.390e6	9.626e10	1.153e6

(c) Dynamic Program (PU)

$\lambda_k, \forall k$ [1]	Expectation [£]	CVaR$_{0.1}$ [£]	CVaR$_{0.05}$ [£]	CVaR$_{0.01}$ [£]	Variance [£^{2}]	Mean-CVaR$_{0.05}$ [£]
0.00	4.283e5	1.075e6	1.202e6	1.459e6	1.230e11	4.283e5
0.25	4.287e5	1.073e6	1.200e6	1.457e6	1.221e11	6.216e5
0.50	4.302e5	1.072e6	1.199e6	1.456e6	1.211e11	8.147e5
0.75	4.329e5	1.073e6	1.199e6	1.454e6	1.203e11	1.007e6
1.00	4.363e5	1.074e6	1.199e6	1.454e6	1.193e11	1.199e6

(d) Dynamic Program (PVU)

$\lambda_k, \forall k$ [1]	Expectation [£]	CVaR$_{0.1}$ [£]	CVaR$_{0.05}$ [£]	CVaR$_{0.01}$ [£]	Variance [£^{2}]	Mean-CVaR$_{0.05}$ [£]
0.00	4.282e5	1.075e6	1.203e6	1.461e6	1.232e11	4.282e5
0.25	4.290e5	1.070e6	1.197e6	1.454e6	1.210e11	6.210e5
0.50	4.311e5	1.067e6	1.193e6	1.447e6	1.193e11	8.121e5
0.75	4.347e5	1.067e6	1.192e6	1.445e6	1.179e11	1.003e6
1.00	4.396e5	1.068e6	1.192e6	1.443e6	1.165e11	1.192e6



(a) $\lambda_{\text{CVaR}} = 0.25$ and $\lambda_k = 0.25$

(b) $\lambda_{\text{CVaR}} = 0.5$ and $\lambda_k = 0.5$

(c) $\lambda_{\text{CVaR}} = 0.75$ and $\lambda_k = 0.75$

(d) $\lambda_{\text{CVaR}} = 1.0$ and $\lambda_k = 1.0$

FIGURE 3.12: Comparison of the pdfs of the trading cost when the uncertainty related to the total volume to be traded is or is not taken into account when solving Optimisation Problem (3.49) (the pdfs are estimated based on 10^7 realisations of the underlying processes).

Chapter 4

Nash equilibrium for risk averse traders with uncertain volume targets

We focus on the derivation of a trade equilibrium where all traders are in competition with each other and, therefore, the actions of some influence the cost of others.

This chapter builds on the background material on the optimal execution problem (Section 1.1) and on game theory (Section 1.3). The content of this chapter is the object of an article currently in preparation and soon available on *arXiv* (Vaes and Hauser, c).

4.1 Introduction

In this chapter, we are interested in studying the optimal execution problem when several traders with market power compete in the market and are subject to the uncertainty that arises from the price dynamics and the uncertain volume target. To define their trading strategies, traders will consequently have to consider the adverse price impact of their own trading, of the trades of the others, all the while considering the risk associated to the price dynamics and the uncertain volume target. This situation is for example encountered in the electricity market, where power wholesalers hedge against risk via forward/future contracts to provide electricity to their end-users, whose demand is intrinsically uncertain.

Modelling this interaction between traders in the optimal execution problem has only been studied in a few previous works. To the best of our knowledge, it has been initiated by Schöneborn (2008) who, based on the results of Obizhaeva and Wang (2013); Carlin et al. (2007), studied the effect of the presence of other large players on the market in the continuous formulation of the execution problem. In the formulation of Schöneborn (2008), players are assumed risk-neutral and derive strategies that are independent of the uncertainty related to the fundamental market price. In the discrete framework, Schied and Zhang (2019) consider a two-player game with transient price impact and where players are risk-neutral. They

prove the existence and uniqueness of a Nash equilibrium in a class of static and deterministic strategies in explicit form. This result has then been extended by Luo and Schied (2018) to an n-player framework, where traders aim to minimise a mean-variance functional of their costs or to maximise the expected exponential utility of their revenues. Finally, Ohnishi and Shimoshimizu (2020) presented a two-player game with risk-averse players who estimate risk with the expected value of the exponential utility of their payoff. The authors solve this problem via dynamic programming and highlight the fact that the optimal strategy is not deterministic in the case where trades of minors actors are not considered deterministic.

None of the aforementioned works consider the case where traders have uncertain volume targets, which is the object of this chapter. In this perspective, we extend to the multiplayer case the single-player model presented in Chapter 2 on the optimal execution problem with uncertain volume target. In the model presented hereafter, we formulate an n-player non-cooperative Nash game with the following attributes: *(i)* traders are risk-averse and estimate risk with $\mathbb{E}$-CVaR$_{\alpha,\lambda}$, *(ii)* traders consider the influence of their own trades as well as those of other players on the price dynamics via permanent and temporary impact functions, *(iii)* traders consider the uncertainty related to the fundamental market price, and finally *(iv)* traders take into account the uncertain nature of their volume targets.

The rest of this chapter is structured as follows: In Section 4.2, we present the n-player framework, and formulate it as a Nash game. We then demonstrate in Section 4.2.7 that this game can be reformulated as an equivalent game for which the payoff functions are differentiable with regards to the decision variables. This is achieved by lifting the set of strategies presented in Chapter 2. Based on this equivalent game formulation, we demonstrate in Section 4.2.8, that the viability of the market induces the uniqueness of the trading strategy of a player given those of the others. We then show in Section 4.3 that the Nash game has an equilibrium, and that this is unique if and only if the solution to the associated variational inequality problem is unique. We then prove uniqueness for the latter problem in the risk-neutral case and formulate a conjecture, supported by numerical evidence, under which the uniqueness also holds in the risk-averse case. Based on the numerical results of Section 4.4, we come to the conclusion that traders who are less risk-sensitive benefit the most from considering volume uncertainty due to their lack of sensitivity with regards to other players' strategies and consequently due to their increased influence when deriving the equilibrium. Section 4.5 then summarises the content of this chapter.

4.2 Framework

We consider the model introduced in Section 1.1, where we assumed that the trading window spans the time interval $[\,t_s, t_f\,]$ with $t_s < t_f$, and that it is subdivided into m trading periods. As mentioned in the introduction, the market is subject to the uncertainty related to the price dynamics and to the volume to be traded, for which the estimate becomes increasingly precise

over time. We will make assumptions similar to those in Chapter 2 regarding these two sources of uncertainty.

4.2.1 Price dynamics

Single-player framework We first derive the model of the price dynamics in the case of a single player; the model is subsequently extended in order to capture explicitly the influence of the trades of all players.

Let S_0 be the initial price of the security, and let $\{S_k, k \in [\![0, m]\!]\}$ be the stochastic process that represents the fair price dynamics. The price process is assumed to be a random walk where the random step or price shift between two successive decision times t_{k-1} and t_k is modelled by a random variable ξ_k, *i.e.* $\forall k \in [\![1, m]\!] : S_k = S_{k-1} + \xi_k$. Furthermore, we assume that the market has limited liquidity and that trades induce linear temporary and permanent market impacts on the price dynamics. The price dynamics is then formulated as follows (see Section 1.1.2 for more details):

$$S_k = S_{k-1} + \xi_k + \gamma_k q_k, \tag{4.1a}$$

$$\widetilde{S}_k = S_{k-1} + \epsilon_k \operatorname{sgn}\left(\frac{q_k}{\tau_k}\right) + \frac{\eta_k}{\tau_k} q_k, \tag{4.1b}$$

where S_k is the fair price at decision time t_k and $\widetilde{S}_k$ is the effective price for the trades executed during the trading period τ_k.

Multiplayer framework In the case where n actors with market power are in competition with each other, the decisions of one can have a significant impact on the transaction costs of others. There is thus need for a model that integrates explicitly the trades of the all players in the price dynamics rather than modelling them with a random variable as done in Equation (4.1). To do so, we denote by $\mathbf{q}_\nu \coloneqq [q_{\nu,1}, \dots, q_{\nu,m}]^{\mathrm{T}}$ the trading quantities of player ν on each trading period and by $\mathbf{q}_{-\nu}$ the trading quantities of the rest of the players with market power. As for the single player case, we also assume that a player ν models the fluctuations in the price dynamics due to external factors, *e.g.* public news or information about the economic situation, and to the trades of minor players with a random walk $(\xi_{\nu,k})_{k \in [\![1,m]\!]}$ (Ohnishi and Shimoshimizu, 2020). Moreover, a player ν integrates in the price dynamics the permanent impact of the trades of the major players. Hence, given $\mathbf{q} = (\mathbf{q}_\nu, \mathbf{q}_{-\nu})$, the fair price dynamics estimated by player ν is as follows

$$S_{\nu,k} = S_{\nu,k-1} + \xi_{\nu,k} + \gamma_k \left(\sum_{\tilde{\nu} \in [\![1,n]\!]} q_{\tilde{\nu},k} \right), \tag{4.2}$$

where $S_{\nu,k}$ is the fair price of the asset at decision time t_k, and $S_{\nu,0} = S_0$ is the initial fair price of the asset. Throughout this paper, the index ν is used in order to designate any values estimated by player ν, *e.g.* $(S_{\nu,k})_{k \in [\![0,m]\!]}$ refers to the fair price dynamics estimated by player ν. We also denote by $\mathbf{q}_\nu^+ \coloneqq \left[q_{\nu,1}^+, \dots, q_{\nu,m}^+\right]^{\mathrm{T}} \geq \mathbf{0}$ (*resp.* $\mathbf{q}_\nu^-$) the quantities of long (*resp.* short) positions traded by player ν on each trading period.

In the single player case, the additional cost to purchase $q^+_{\nu,k}$ shares on τ_k due the temporary impact is modelled as in Equation (1.5). In the multiplayer case, if for a trading period τ_k the buy order of a player $\tilde{\nu} \neq \nu$ is executed prior the one of player ν, then the latter has on top of the previous cost an additional cost that amounts to $2\eta_k q^+_{\tilde{\nu},k}$ per unit traded (see illustration of the assumption on the linear impact of trades on the price dynamics in Figure 1.3). This is referred as the *latency cost*. This is intuitively explained by the fact that player $\tilde{\nu}$ walks the limit order book before player ν. As mentioned in Luo and Schied (2018), we can assume that fifty percent of the time, the orders of player ν will be executed before the orders of player $\tilde{\nu}$. Hence, the temporary transaction cost of a buy market order of player ν in the multiplayer case can be estimated as follows

$$\epsilon_k q^+_{\nu,k} + \frac{\eta_k}{\tau_k}\left(\sum_{\tilde{\nu}\in[\![1,n]\!]} q^+_{\nu,k}\, q^+_{\tilde{\nu},k}\right) + \epsilon_k q^-_{\nu,k} + \frac{\eta_k}{\tau_k}\left(\sum_{\tilde{\nu}\in[\![1,n]\!]} q^-_{\nu,k}\, q^-_{\tilde{\nu},k}\right). \tag{4.3}$$

Note that in Equation (4.3), we dissociate the temporary cost related to the long and short positions. This is because we assume that the players are price takers and not price makers: a player thus assumes that the limit order book is not refilled during a trading period. Hence, a long and a short position cannot be compensated, *i.e.* a player taking a long and a short position still has a temporary transaction cost even if the net position is 0. The transaction price $\widetilde{S}^+_{\nu,k}$ (*resp.* $\widetilde{S}^-_{\nu,k}$) for player ν for taking long (*resp.* short) positions on the k-th trading period is thus affected by the quantities traded by all players and is given by:

$$\widetilde{S}^+_{\nu,k} = S_{\nu,k-1} + \epsilon_k + \frac{\eta_k}{\tau_k}\left(\sum_{\tilde{\nu}\in[\![1,n]\!]} q^+_{\tilde{\nu},k}\right), \tag{4.4a}$$

$$\widetilde{S}^-_{\nu,k} = S_{\nu,k-1} - \epsilon_k - \frac{\eta_k}{\tau_k}\left(\sum_{\tilde{\nu}\in[\![1,n]\!]} q^-_{\tilde{\nu},k}\right). \tag{4.4b}$$

4.2.2 Volume target dynamics

Additionally to the uncertainty related to the price dynamics, we assume that, at the start of the execution period t_s, a player ν has only a forecast on the volume D_{ν,t_f} that must be traded by time t_f. To incorporate this source of uncertainty in the trading strategy of a player, we adopt the framework proposed in Chapter 2. We denote by $D_{\nu,k} \coloneqq \mathbb{E}^{\mathbb{P}}[D_{\nu,m} \mid \mathcal{F}_k]$ the random variable corresponding to the forecast of the volume target of player ν at time t_k, where $\mathbb{P}$ is the probability measure on the sample space Ω and $(\mathcal{F}_k)_{k=1}^m$ is the filtration of sigma algebras that represent the information available at time t_k. We assume that a player ν knows with certainty their volume target at decision time t_{m-1}, which enables the player to match with their volume target by adjusting their position on the last trading period τ_m. Two successive forecasts on the volume target, $D_{\nu,k-1}$ and $D_{\nu,k}$, are assumed linked by a continuous random variable $\delta_{\nu,k}$ for $k < m$. With these assumptions, the forecast on the volume target follows the dynamics

$$\begin{cases} D_{\nu,k} = D_{\nu,k-1} + \delta_{\nu,k}, & \text{if } k \in [\![1, m-1]\!], \\ D_{\nu,m} = D_{\nu,m-1}, \end{cases} \tag{4.5}$$

where $D_{\nu,m} = D_{\nu,t_f}$, and $\delta_{\nu,k}$ is the *forecast update* that comes to the knowledge of player ν at time t_k.

4.2.3 Trading strategy

A *trading strategy* of a player ν is defined as in Chapter 2: It corresponds to a tuple $(\boldsymbol{y}_\nu, \boldsymbol{\beta}_\nu) \in \mathbb{R}^m \times \mathbb{R}^{m-1\times m}$, where the first argument represents the proportions of the most up-to-date forecast on the volume target to be traded on each trading period, and the second argument tells how the observed trading error due to past forecast updates is dealt with on the remaining trading periods.

As previously mentioned, numerical simulations that we carried suggest that the reduction in the objective function obtained by considering $\boldsymbol{\beta}_\nu$ as part of the variables is mild compared to the case where $\boldsymbol{\beta}_\nu$ is defined as parameter the redistribution matrix implied by the risk-neutral strategy computed at time t_0 in the mean-variance framework of Almgren and Chriss (2001) (Vaes and Hauser, 2022, Section 2.3). We therefore assume in this work that the redistribution matrix $\boldsymbol{\beta}_\nu$ of any player ν is fixed and known by every other player. We omit to explicitly mention $\boldsymbol{\beta}$ as part of the strategy for readability purposes.

Like in the single-player framework (see Equation (2.6)), the volume $Q_{\nu,k}$ traded on a trading period τ_k by a player ν having as strategy $\boldsymbol{y}_\nu$ is the following random variable:

$$Q_{\nu,k}(\boldsymbol{y}_\nu) = y_{\nu,k} D_{\nu,k-1} + \sum_{i=1}^{k-1} \beta_{\nu,i,k} \delta^{\varepsilon}_{\nu,i} = y_{\nu,k} D_{\nu,0} + \sum_{i=1}^{k-1} \delta_{\nu,i} \left(y_{\nu,k} + \beta_{\nu,i,k} \sum_{r=1}^{i} y_{\nu,r} \right). \tag{4.6}$$

The quantity $Q_{\nu,k}$ is composed of two elements: The first contribution is the proportion $y_{\nu,k}$ of the last forecast released on the volume target, *i.e.* $D_{\nu,k-1}$. Then, every *forecast update* $\delta_{\nu,i}$, $i \in [\![1, m-1]\!]$, introduces a *forecast error* $\delta^{\varepsilon}_{\nu,i} \coloneqq \delta_{\nu,i} \sum_{r=1}^{i} y_{\nu,r}$, which corresponds to the additional volume, positive or negative, that should have been traded by player ν between t_0 and t_i if $\delta_{\nu,i}$ had been known from time t_0. The second contribution is therefore the sum of the proportion $\beta_{i,k}$ of each past forecast error $\delta^{\varepsilon}_{\nu,i}$, $i < k$, *i.e.* $\sum_{i=1}^{k-1} \beta_{\nu,i,k} \delta^{\varepsilon}_{\nu,i}$. The matrix $\boldsymbol{\beta}_\nu \in \mathbb{R}^{m-1\times m}$ is called the redistribution matrix, where each component $\beta_{\nu,i,k}$ denotes the proportion of $\delta^{\varepsilon}_{\nu,i}$ to trade during τ_k in order to progressively correct for the error caused by the forecast update of τ_i (a more detailed explanation is given in Chapter 2).

Given a realisation $\omega \in \Omega$, there exists a non-singular matrix $\mathbf{L}_\nu(\boldsymbol{\beta}_\nu\,;\,\omega)$ such that, for any $\boldsymbol{y}_\nu \in \mathcal{Y}_\nu$, the vector of the volumes traded on each trading period by player ν following strategy $\boldsymbol{y}_\nu$, written as $\mathbf{Q}_\nu(\boldsymbol{y}_\nu\,;\,\omega) \coloneqq [Q_{\nu,1}(\boldsymbol{y}_\nu\,;\,\omega), \ldots, Q_{\nu,m}(\boldsymbol{y}_\nu\,;\,\omega)]^{\mathrm{T}}$, is given by

$$\mathbf{Q}_\nu(\boldsymbol{y}_\nu\,;\,\omega) = \mathbf{L}_\nu(\boldsymbol{\beta}_\nu\,;\,\omega)\, \boldsymbol{y}_\nu. \tag{4.7}$$

We should note that by adopting this strategy structure, it can be shown that the final net position is equal to the realised volume target:

$$\forall \omega \in \Omega : \sum_{i=1}^{m} Q_\nu(\boldsymbol{y}_\nu\,;\,\omega) = D_{\nu,m}(\omega) = D_{\nu,t_f}(\omega)\,. \tag{4.8}$$

4.2.4 Trading cost

Given the deterministic volumes traded over the course of the execution period by the other players $\left(\mathbf{q}^+_{-\nu}, \mathbf{q}^-_{-\nu}\right)$, we define the *liquidation cost* of a player ν that trades the deterministic volumes $\left(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu\right) \in \mathbb{R}^m_+ \times \mathbb{R}^m_+$ as the following random variable:

$$\begin{aligned}\mathcal{C}_\nu\left(\mathbf{q}^+, \mathbf{q}^-\right) &\coloneqq \sum_{k=1}^{m}\left(q^+_{\nu,k}\widetilde{S}^+_{\nu,k} + q^-_{\nu,k}\widetilde{S}^-_{\nu,k}\right) - d_{\nu,m}S_0 \\ &= \sum_{k=1}^{m-1}\left(\xi_{\nu,k} + \gamma_k\left(\sum_{\tilde{\nu}\in[\![1,n]\!]}\left(q^+_{\tilde{\nu},k} - q^-_{\tilde{\nu},k}\right)\right)\right)\left(d_{\nu,m} - \sum_{i=1}^{k}\left(q^+_{\nu,i} - q^-_{\nu,i}\right)\right) \\ &\quad + \sum_{k=1}^{m}\epsilon_k\left(q^+_{\nu,k} + q^-_{\nu,k}\right) + \sum_{k=1}^{m}\left(\left(\frac{\eta_k}{\tau_k}\right)\left(\sum_{\tilde{\nu}\in[\![1,n]\!]} q^+_{\nu,k}\, q^+_{\tilde{\nu},k} + \sum_{\tilde{\nu}\in[\![1,n]\!]} q^-_{\nu,k}\, q^-_{\tilde{\nu},k}\right)\right), \end{aligned} \tag{4.9}$$

where $d_{\nu,m} \coloneqq \sum_{k=1}^{m} q^+_{\nu,k} - q^-_{\nu,k}$ is the net position traded by player ν over the entire execution period $[t_s, t_f]$. This corresponds to the difference between the effective trading cost incurred at the end of the execution period and the trading cost ideally obtained in an hypothetical infinitely liquid market where the entire position would be traded at the beginning of the execution period.

In the present case, we assume that the volume target is uncertain, which implies that the trading quantities are random variables. We will consequently adopt the trading cost definition formulated in Equation (2.5). Given the trading strategy structure introduced in Section 4.2.3, we define the *trading cost* of a player ν as the difference between the cost incurred at the end of the execution period by following the trading strategy with the cost incurred if the final volume target $D_{\nu,m}$ was perfectly known at time t_0 and if the entire position was traded at time t_0 on a hypothetical market free of market frictions:

$$\begin{aligned}\mathcal{C}_\nu(\boldsymbol{y}) = \sum_{k=1}^{m-1}\left(\xi_{\nu,k} + \gamma_k\left(\sum_{\tilde{\nu}\in[\![1,n]\!]} Q_{\tilde{\nu},k}(\boldsymbol{y}_{\tilde{\nu}})\right)\right)\left(D_{\nu,m} - \sum_{i=1}^{k} Q_{\nu,i}(\boldsymbol{y}_\nu)\right) \\ + \sum_{k=1}^{m}\epsilon_k\left|Q_{\nu,k}(\boldsymbol{y}_\nu)\right| + \sum_{k=1}^{m}\frac{\eta_k}{\tau_k}\left(\sum_{\tilde{\nu}\in[\![1,n]\!]}\max(0,\, Q_{\nu,k}(\boldsymbol{y}_\nu)\, Q_{\tilde{\nu},k}(\boldsymbol{y}_{\tilde{\nu}}))\right). \end{aligned} \tag{4.10}$$

4.2.5 Risk minimisation

In this chapter, we assume that the sample space Ω is finite, *i.e.* $\Omega = \{\omega_1, \ldots, \omega_{n_\Omega}\}$ with $n_\Omega = |\Omega|$, with associated physical probability measure $\mathbb{P}$. We assume that each player ν quantifies the risk associated with their trading strategy based on the $\mathbb{E}$-$\mathrm{CVaR}_{\alpha,\lambda}$ risk measure, which is defined as the trade-off parametrised by λ between the expectation and CVaR_α (see definition in Equation (1.25)). Hence, to partially hedge the risk associated to unfavourable scenarios, a player ν then seeks the trading strategy that minimises the weighted sum, parametrised by $\lambda_\nu \in [0,1]$, of the expectation and the α-Conditional Value-at-Risk of their trading cost:

$$\underset{\boldsymbol{y}_\nu}{\text{minimise}} \quad (1-\lambda_\nu)\,\mathbb{E}[\mathcal{C}_\nu(\boldsymbol{y})] + \lambda_\nu\,\mathrm{CVaR}_{\alpha_\nu}[\mathcal{C}_\nu(\boldsymbol{y})] \tag{4.11a}$$

$$\text{subject to} \quad \boldsymbol{y}_\nu \in \mathcal{Y}_\nu, \tag{4.11b}$$

where $\mathrm{CVaR}_{\alpha_\nu}[\mathcal{C}_\nu(\boldsymbol{y})]$ corresponds to the expectation of the trading cost conditional to the fact that it exceeds the $(1-\alpha_\nu)$-quantile.

4.2.6 Trading game

We are consequently in the situation where n players (or traders) trades on the same market and where the actions of one player influence the objective function of any other player. It is consequently of interest to investigate whether this situation has an equilibrium point, *i.e.* a set of strategies for each player, such that no player has interest in deviating from their strategy.

To carry out this analysis, we formulate this problem within the framework of game theory. Using such a framework allows us to capture the private interest that drives each player. As previously mentioned, given the decision variables $\boldsymbol{y}_{-\nu}$ of the other players, each player $\nu \in [\![1, n]\!]$ solves the following optimisation problem:

$$\mathbf{P}_\nu \coloneqq \underset{\boldsymbol{y}_\nu}{\text{minimise}} \quad \theta_\nu(\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu}) \coloneqq \mathbb{E}\text{-}\mathrm{CVaR}_{\alpha_\nu,\lambda_\nu}[\mathcal{C}_\nu(\boldsymbol{y})]\,, \tag{4.12a}$$

$$\text{subject to} \quad \boldsymbol{y}_\nu \in \mathcal{Y}_\nu. \tag{4.12b}$$

The objective function θ_ν is called the *payoff* function of player ν. We use the following notation $\boldsymbol{\Theta} \coloneqq \{\theta_\nu\}_{\nu=1}^n$. We can now define the trading game as follows:

Definition 4.1. *The n-player game where each trader $\nu \in [\![1, n]\!]$ solves (4.12) is denoted by $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ and is called the* trading game.

We can now define the notions of risk-neutral and risk-averse equilibrium as follows:

Definition 4.2. *Given the physical probability measure $\mathbb{P}$ on the sample space Ω, a* risk-neutral Nash equilibrium *of the n-player game $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$, noted $\mathrm{RnEq}(\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta}))$, is a set of strategies $\{\boldsymbol{y}_\nu^\star \in \mathcal{Y}_\nu\}_{\nu \in [\![1,n]\!]}$ such that, for each player $\nu \in [\![1, n]\!]$, $\boldsymbol{y}_\nu^\star$ is a solution of $\mathbf{P}_\nu$ given $\boldsymbol{y}_{-\nu}^\star$ when all players are risk-neutral,* i.e. $\forall \nu : \lambda_\nu = 0$.

Definition 4.3. *Given the physical probability measure $\mathbb{P}$ on the sample space Ω, a* risk-averse Nash equilibrium *of the n-player game $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$, noted $\mathrm{RaEq}(\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta}))$, is a set of strategies $\{\boldsymbol{y}_\nu^\star \in \mathcal{Y}_\nu\}_{\nu \in [\![1,n]\!]}$ such that, for each player $\nu \in [\![1, n]\!]$, $\boldsymbol{y}_\nu^\star$ is a solution of $\mathbf{P}_\nu$ given $\boldsymbol{y}_{-\nu}^\star$ and where all players can potentially be risk-averse,* i.e. $\forall \nu : \lambda_\nu \in [0, 1]$.

In this work, we investigate the existence and uniqueness of a pure Nash equilibria in the sense that we do not assume that the strategy of a player ν can be a randomisation over several points of $\mathcal{Y}_\nu$.

We should note that an optimal strategy resulting from the optimisation of (4.12) is *deterministic*, *i.e.* a player decides of their strategy at the initial trading time t_0 and does not modify it over the course of the execution period $[t_s, t_f]$. Consequently, the strategy of a trader is relatively small and of size m, which explains, among other reasons, why we have privileged to formulate the execution problem using the framework introduced in Chapter 2 rather than

the dynamical program presented in Chapter 3 for which the optimal strategy is defined as the quantity to trade for any realisation $\omega \in \Omega$ and consequently the set $\mathcal{Y}_\nu$ would be of a much larger size.

In the rest of this work, we denote by $\mathcal{Y} \coloneqq \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n$ the set of all feasible strategies and we interchangeably use the notations $\boldsymbol{y}$, $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_\nu, \ldots, \boldsymbol{y}_n)$ and $(\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu})$. Hence, $\theta_\nu(\boldsymbol{y})$, $\theta_\nu(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_\nu, \ldots, \boldsymbol{y}_n)$ and $\theta_\nu(\boldsymbol{y}_\nu, \boldsymbol{y}_{-\nu})$ will thus for instance be used as equivalent notations.

4.2.7 Equivalent game formulation with differentiable cost function

In the next section, we find an equivalent formulation to (4.12) by lifting the feasible set in order to obtain an expression for the trading cost that is $\mathcal{C}^\infty$ in terms of the variables of all players. This has a two-fold interest: Firstly, as shown subsequently in Section 4.2.8, it allows us to prove the uniqueness of a trading strategy if it is free of any price manipulation as formulated in Lemma 1.1. Secondly, it eases the proof of the existence and uniqueness of a Nash equilibrium by allowing to apply without any further modifications the existence and uniqueness results presented in Sections 1.3.3 and 1.3.4 (see Section 4.3).

In Model (4.12), the quantities traded by a player $\nu \in [\![1, n]\!]$ in the case of realisation $\omega \in \Omega$ are given by Equation (4.7). These quantities are the net positions taken. On each trading period, a trader takes consequently either exclusively long or short positions. To allow a trader ν to take both long and short positions on each trading period τ_k, $k \in [\![1, m]\!]$, we lift their feasible set $\mathcal{Y}_\nu$ as follows:

$$\widehat{\mathcal{Y}}_\nu \coloneqq \left\{ \left(\mathbf{q}_\nu^+, \mathbf{q}_\nu^- \right) \in \mathbb{R}_+^{(m-1)n_\Omega} \times \mathbb{R}_+^{mn_\Omega} \,\middle|\, \exists \boldsymbol{y}_\nu \in \mathcal{Y}_\nu : (4.14) \text{ and } (4.15) \text{ hold} \right\}, \tag{4.13}$$

with

$$\forall \omega \in \Omega, \forall k \in [\![1, m-1]\!] : q^+_{\nu,\omega,k} - q^-_{\nu,\omega,k} = Q_{\nu,k}(\boldsymbol{y}_\nu \,;\, \omega)\,, \tag{4.14}$$

$$\forall \omega \in \Omega : q^-_{\nu,\omega,m} \geq \sum_{k=1}^{m-1} \left(q^+_{\nu,\omega,k} - q^-_{\nu,\omega,k} \right) - D_{\nu,m}(\omega)\,, \tag{4.15}$$

where $q^+_{\nu,\omega,k}$ (*resp.* $q^-_{\nu,\omega,k}$) denotes the number of long (*resp.* short) positions to trade during trading period τ_k in the case of realisation $\omega \in \Omega$. Constraint (4.14) imposes that, for any $\omega \in \Omega$, the net position traded by player ν on each τ_k, $k \in [\![1, m-1]\!]$, is equal the net position of a trader following the strategy $\boldsymbol{y}_\nu$ (see Equation (4.6)). Note that the amount of long positions $q^+_{\nu,\omega,m}$ to trade during the final trading period is not part of a strategy $\hat{\boldsymbol{y}}_\nu \in \widehat{\mathcal{Y}}_\nu$, however it can be obtained using the constraint that the final net position must be equal to the volume target (see Equation (4.8)), *i.e.*

$$\forall \omega \in \Omega : q^+_{\nu,\omega,m} = \sum_{k=1}^{m} q^-_{\nu,\omega,k} - \sum_{k=1}^{m-1} q^+_{\nu,\omega,k} + D_{\nu,m}(\omega)\,, \tag{4.16}$$

which is greater than or equal to 0 since (4.15) must hold. Moreover, if $q^+_{\nu,\omega,m}$ is defined as in (4.16), then, for all $\omega \in \Omega$, $q^+_{\nu,\omega,m} - q^-_{\nu,\omega,m} = Q_{\nu,m}(\boldsymbol{y}_\nu \,;\, \omega)$ holds. We use $\hat{\boldsymbol{y}}_\nu$ as shorthand for the

tuple $(\mathbf{q}_\nu^+, \mathbf{q}_\nu^-)$ and define $\widehat{\mathcal{Y}}$ as $\widehat{\mathcal{Y}} \coloneqq \widehat{\mathcal{Y}}_1 \times \cdots \times \widehat{\mathcal{Y}}_n$. Hence, by lifting the feasible set of (4.12), a player ν then solves the following optimisation problem $\widehat{\mathbf{P}}_\nu$:

$$\widehat{\mathbf{P}}_\nu \coloneqq \underset{\hat{\boldsymbol{y}}_\nu}{\text{minimise}} \quad \hat{\theta}_\nu(\hat{\boldsymbol{y}}_\nu, \hat{\boldsymbol{y}}_{-\nu}) \coloneqq \mathbb{E}\text{-CVaR}_{\alpha_\nu,\lambda_\nu}\left[\widehat{\mathcal{C}}_\nu(\hat{\boldsymbol{y}})\right], \tag{4.17a}$$

$$\text{subject to} \quad \hat{\boldsymbol{y}}_\nu \in \widehat{\mathcal{Y}}_\nu, \tag{4.17b}$$

where $\widehat{\mathcal{C}}_\nu(\hat{\boldsymbol{y}})$ is the random variable depending on the strategies $\hat{\boldsymbol{y}}_{-\nu}$ that represents the trading cost of player ν following the strategy $\hat{\boldsymbol{y}}_\nu$. The realisation of this random variable given $\omega \in \Omega$ is denoted by $\widehat{\mathcal{C}}_\nu(\hat{\boldsymbol{y}}\,;\,\omega)$ and is defined as follows:

$$\begin{aligned}\widehat{\mathcal{C}}_\nu(\hat{\boldsymbol{y}}\,;\,\omega) = \sum_{k=1}^{m-1}\left(\xi_{\nu,k}(\omega) + \gamma_k\left(\sum_{\tilde{\nu}\in[\![1,n]\!]}\left(q^+_{\tilde{\nu},\omega,k} - q^-_{\tilde{\nu},\omega,k}\right)\right)\right)\left(D_{\nu,m}(\omega) - \sum_{i=1}^{k}\left(q^+_{\nu,\omega,i} - q^-_{\nu,\omega,i}\right)\right) \\ + \sum_{k=1}^{m}\epsilon_k\left(q^+_{\nu,\omega,k} + q^-_{\nu,\omega,k}\right) + \sum_{k=1}^{m}\frac{\eta_k}{\tau_k}\left(\sum_{\tilde{\nu}\in[\![1,n]\!]}\left(q^+_{\nu,\omega,k}\,q^+_{\tilde{\nu},\omega,k} + q^-_{\nu,\omega,k}\,q^-_{\tilde{\nu},\omega,k}\right)\right).\end{aligned} \tag{4.18}$$

Definition 4.4. *The n-player game where each trader $\nu \in [\![1,n]\!]$ solves (4.17) is denoted by $\mathcal{G}\left(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\right)$ and is called the* lifted trading game.

Definition 4.5. *Given the physical probability measure $\mathbb{P}$ on the sample space Ω, a* risk-neutral Nash equilibrium *of the n-player game $\mathcal{G}\left(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\right)$, noted $\text{RnEq}\left(\mathcal{G}\left(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\right)\right)$, is a set of strategies $\left\{\hat{\boldsymbol{y}}^\star_\nu \in \widehat{\mathcal{Y}}_\nu\right\}_{\nu\in[\![1,n]\!]}$ such that, for each player $\nu \in [\![1,n]\!]$, $\hat{\boldsymbol{y}}^\star_\nu$ is a solution of $\widehat{\mathbf{P}}_\nu$ given $\hat{\boldsymbol{y}}^\star_{-\nu}$ when all players are risk-neutral,* i.e. $\forall \nu : \lambda_\nu = 0$.

Definition 4.6. *Given the physical probability measure $\mathbb{P}$ on the sample space Ω, a* risk-averse Nash equilibrium *of the n-player game $\mathcal{G}\left(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\right)$, noted $\text{RaEq}\left(\mathcal{G}\left(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\right)\right)$, is a set of strategies $\left\{\hat{\boldsymbol{y}}^\star_\nu \in \widehat{\mathcal{Y}}_\nu\right\}_{\nu\in[\![1,n]\!]}$ such that, for each player $\nu \in [\![1,n]\!]$, $\hat{\boldsymbol{y}}^\star_\nu$ is a solution of $\widehat{\mathbf{P}}_\nu$ given $\hat{\boldsymbol{y}}^\star_{-\nu}$ and where all players can potentially be risk-averse,* i.e. $\forall \nu : \lambda_\nu \in [0,1]$.

The following results prove the existence of a one-to-one mapping between every Nash equilibrium of the trading game $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ and the lifted trading game $\mathcal{G}\left(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\right)$.

Lemma 4.1 (Complementarity condition). *For any player $\nu \in [\![1,n]\!]$ with risk-aversion parameter $\lambda_\nu \in [0,1)$, if $\widehat{\mathcal{Y}}^\star_\nu \subseteq \widehat{\mathcal{Y}}_\nu$ denotes the set of optimal execution strategies of (4.17) given the strategies $\hat{\boldsymbol{y}}_{-\nu}$ of the other players. Then, we have for all $\hat{\boldsymbol{y}}^\star_\nu = (\mathbf{q}^{+,\star}_\nu, \mathbf{q}^{-,\star}_\nu) \in \widehat{\mathcal{Y}}^\star_\nu$, and $\omega \in \Omega$, that the following complementarity condition holds for all $k \in [\![1,m]\!]$:*

$$q^{+,\star}_{\nu,\omega,k} \geq 0 \quad \perp \quad q^{-,\star}_{\nu,\omega,k} \geq 0, \tag{4.19}$$

which means that at least one of these inequalities must be an equality.

Proof. By contradiction, we assume that there exist an optimal solution $\hat{\boldsymbol{y}}^\dagger_\nu \coloneqq \left(\mathbf{q}^{+,\dagger}_\nu, \mathbf{q}^{-,\dagger}_\nu\right) \in \widehat{\mathcal{Y}}^\star_\nu$, a trading period $k^\dagger \in [\![1,m]\!]$, and a realisation $\omega^\dagger \in \Omega$, such that (4.19) is not respected, *i.e.*

$$q^{+,\dagger}_{\nu,\omega^\dagger,k^\dagger} > 0 \quad \text{and} \quad q^{-,\dagger}_{\nu,\omega^\dagger,k^\dagger} > 0. \tag{4.20}$$

Then, based on this strategy $\hat{\boldsymbol{y}}_\nu^\dagger$, we define for a parameter value $\mu \in \mathbb{R}_{++}$ the strategy $\hat{\boldsymbol{y}}_\nu^\ddagger := \left(\mathbf{q}_\nu^{+,\ddagger}, \mathbf{q}_\nu^{-,\ddagger}\right)$, where $\mathbf{q}_\nu^{+,\ddagger}$ and $\mathbf{q}_\nu^{-,\ddagger}$ are defined as follows:

$$\forall (k,\omega) \neq \left(k^\dagger, \omega^\dagger\right): \quad q_{\nu,\omega,k}^{-,\ddagger} := q_{\nu,\omega,k}^{-,\dagger} \quad \text{and} \quad q_{\nu,\omega,k}^{+,\ddagger} := q_{\nu,\omega,k}^{+,\dagger}, \tag{4.21a}$$

$$q_{\nu,\omega^\dagger,k^\dagger}^{-,\ddagger} := q_{\nu,\omega^\dagger,k^\dagger}^{-,\dagger} - \mu, \tag{4.21b}$$

$$q_{\nu,\omega^\dagger,k^\dagger}^{+,\ddagger} := q_{\nu,\omega^\dagger,k^\dagger}^{+,\dagger} - \mu. \tag{4.21c}$$

For any value of μ such that

$$\mu < \min\left(q_{\nu,\omega^\dagger,k^\dagger}^{+,\dagger},\, q_{\nu,\omega^\dagger,k^\dagger}^{-,\dagger}\right), \tag{4.22}$$

the strategy defined as in Equation (4.21) belongs to $\widehat{\mathcal{Y}}_\nu$. For all $\omega \neq \omega^\dagger$ the strategies $\hat{\boldsymbol{y}}_\nu^\dagger$ and $\hat{\boldsymbol{y}}_\nu^\ddagger$ are equivalent and has consequently the same trading cost:

$$\forall \omega \neq \omega^\dagger : \widehat{\mathcal{C}}_\nu\left(\hat{\boldsymbol{y}}_\nu^\dagger, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega\right) - \widehat{\mathcal{C}}_\nu\left(\hat{\boldsymbol{y}}_\nu^\ddagger, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega\right) = 0. \tag{4.23}$$

However, for $\omega^\dagger$, the trading cost of $\hat{\boldsymbol{y}}_\nu^\ddagger$ is strictly lower than the one of $\hat{\boldsymbol{y}}_\nu^\dagger$ as show as follows:

$$\widehat{\mathcal{C}}_\nu\left(\hat{\boldsymbol{y}}_\nu^\dagger, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega^\dagger\right) - \widehat{\mathcal{C}}_\nu\left(\hat{\boldsymbol{y}}_\nu^\ddagger, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega^\dagger\right) = \mu\left(2\epsilon_{k^\dagger} + \frac{\eta_{k^\dagger}}{\tau_{k^\dagger}}\left(\underbrace{2q_{\nu,\omega^\dagger,k^\dagger}^{+,\dagger} + 2q_{\nu,\omega^\dagger,k^\dagger}^{-,\dagger} - 2\mu}_{>0 \text{ with } (4.22)} + \sum_{\tilde{\nu}\neq\nu}\left(q_{\tilde{\nu},\omega^\dagger,k^\dagger}^{+} + q_{\tilde{\nu},\omega^\dagger,k^\dagger}^{-}\right)\right)\right) > 0, \tag{4.24}$$

as by assumption $\epsilon_{k^\dagger}, \tau_{k^\dagger}, \eta_{k^\dagger} > 0$. As a consequence, since $\lambda_\nu \in [0,1)$, the objective function charges all realisations $\omega \in \Omega$, thus $\hat{\theta}_\nu\left(\hat{\boldsymbol{y}}_\nu^\dagger, \hat{\boldsymbol{y}}_{-\nu}\right) > \hat{\theta}_\nu\left(\hat{\boldsymbol{y}}_\nu^\ddagger, \hat{\boldsymbol{y}}_{-\nu}\right)$. This contradicts the assumption that $\hat{\boldsymbol{y}}_\nu^\dagger$ is optimal and terminates the proof. □

Corollary 4.1. *For a player $\nu \in [\![1,n]\!]$ with risk-aversion parameter $\lambda_\nu \in [0,1)$, if $\hat{\boldsymbol{y}}_\nu^\star = (\mathbf{q}_\nu^{+,\star}, \mathbf{q}_\nu^{-,\star})$ is an optimal strategy of (4.17), then there exists a $\boldsymbol{y}_\nu^\star \in \mathcal{Y}_\nu$ that replicates $\hat{\boldsymbol{y}}_\nu^\star$ in the sense that the volumes traded on each trading period are equal for both strategies,* i.e. *for all $\omega \in \Omega$ and $k \in [\![1,m]\!]$ the following equalities hold:*

$$q_{\nu,\omega,k}^{+,\star} = \max(Q_{\nu,k}(\boldsymbol{y}_\nu^\star\,;\,\omega)\,,\,0) \quad \text{and} \quad q_{\nu,\omega,k}^{-,\star} = \max(-Q_{\nu,k}(\boldsymbol{y}_\nu^\star\,;\,\omega)\,,\,0)\,. \tag{4.25}$$

Proof. The proof follows from Lemma 4.1 and Equation (4.13). □

Theorem 4.1 (One-to-one mapping RaEq($\mathcal{G}(\mathcal{Y},\boldsymbol{\Theta})$) and RaEq$\left(\mathcal{G}\left(\widehat{\mathcal{Y}},\widehat{\boldsymbol{\Theta}}\right)\right)$).
If $\{\boldsymbol{y}_\nu^\star \in \mathcal{Y}_\nu\}_{\nu\in[\![1,n]\!]}$ is a risk-averse Nash equilibrium of $\mathcal{G}(\mathcal{Y},\boldsymbol{\Theta})$, then the lifted strategies denoted by $\{\hat{\boldsymbol{y}}_\nu = (\mathbf{q}_\nu^+, \mathbf{q}_\nu^-)\}_{\nu\in[\![1,n]\!]}$ defined as follows

$$\forall \omega \in \Omega, \forall k \in [\![1, m-1]\!]: \quad q_{\nu,\omega,k}^{+} = \max(Q_{\nu,k}(\boldsymbol{y}_\nu^\star\,;\,\omega)\,,\,0)\,, \tag{4.26}$$

$$\forall \omega \in \Omega, \forall k \in [\![1, m]\!]: \quad q_{\nu,\omega,k}^{-} = \max(-Q_{\nu,k}(\boldsymbol{y}_\nu^\star\,;\,\omega)\,,\,0)\,, \tag{4.27}$$

is a risk-averse Nash equilibrium of $\mathcal{G}\left(\widehat{\mathcal{Y}},\widehat{\boldsymbol{\Theta}}\right)$. Conversely, if

$$\left\{\hat{\boldsymbol{y}}_\nu^\diamond = \left(\mathbf{q}_\nu^{+,\diamond}, \mathbf{q}_\nu^{-,\diamond}\right) \in \widehat{\mathcal{Y}}_\nu\right\}_{\nu\in[\![1,n]\!]} \tag{4.28}$$

is a risk-averse Nash equilibrium of $\mathcal{G}\left(\widehat{\mathcal{Y}},\widehat{\boldsymbol{\Theta}}\right)$, then the projected strategies $\{\boldsymbol{y}_\nu^\diamond, \nu \in [\![1,n]\!]\}$ (the strategies in $\mathcal{Y}_\nu$ such that Equation (4.13) holds) is a risk-averse Nash equilibrium of $\mathcal{G}(\mathcal{Y},\boldsymbol{\Theta})$.

PROOF. The proof follows directly from Corollary 4.1. □

With Theorem 4.1, proving the existence and uniqueness of the Nash equilibrium for $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ is equivalent to proving the existence and uniqueness of the Nash equilibrium for $\mathcal{G}\left(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\right)$. We will see that proving the latter is easier than the former since the cost function $\widehat{\mathcal{C}}_\nu$ is differentiable in $\mathbf{q}^+$ and $\mathbf{q}^-$.

4.2.8 A necessary condition for market viability

In Lemma 1.1 (Vaes and Hauser (2022, Lemma 2.3)), we derived for the single player framework a necessary condition for the absence of price manipulation. In this section, we show that this this condition remains necessary for the absence of price manipulation, and additionally, we show that this condition is also sufficient to prevent any unbounded price manipulation. A (unbounded) price manipulation as defined in Huberman and Stanzl (2004) relies on *round-trip strategies*, which is defined as follows:

DEFINITION 4.7 (ROUND-TRIP STRATEGY). *A (deterministic)* round-trip strategy $(\mathbf{q}^+, \mathbf{q}^-) \in \mathbb{R}_+^m \times \mathbb{R}_+^m \setminus \{\mathbf{0}\}$ *is defined as the quantities of long and short positions to trade on each trading period such that the cumulated net position equates to zero,* i.e. $\sum_{k=1}^m q_k^+ - q_k^- = 0$, *and if for all period* $k \in [\![1, m]\!]$*:* $q_k^+ \neq q_k^-$ *if* $q_k^+ > 0$ *and* $q_k^- > 0$. *We denote by* $\mathcal{RT}$ *the set of round-trip strategies.*

This definition generalises Definition 1.1 as a trader can on any trading period τ_k, $k \in [\![1, m]\!]$, trade both long and short positions, whereas in Definition 1.1 a trader takes exclusively either long or short positions on each trading period. Based on Definition 4.7 and on the definition of price manipulation of Huberman and Stanzl (2004, Definition 1), we define, in the context of the multiplayer game previously presented, a *price manipulation* as follows:

DEFINITION 4.8 (PRICE MANIPULATION). *Given the strategies of the other player* $\hat{\boldsymbol{y}}_{-\nu}$*, a* (risk neutral) price manipulation *for a player* ν *is a round-trip strategy* $(\mathbf{q}_\nu^+, \mathbf{q}_\nu^-)$ *with a strictly negative expected cost,* i.e.

$$\mathbb{E}\left[\widehat{\mathcal{C}}_\nu\left(\mathbf{q}_\nu^+, \mathbf{q}_\nu^-, \hat{\boldsymbol{y}}_{-\nu}\right)\right] < 0, \tag{4.29}$$

where $\widehat{\mathcal{C}}_\nu(\mathbf{q}_\nu^+, \mathbf{q}_\nu^-, \hat{\boldsymbol{y}}_{-\nu})$ *is the random variable depending on the strategies* $\hat{\boldsymbol{y}}_{-\nu}$ *that represents the trading cost of the round-trip strategy* $(\mathbf{q}_\nu^+, \mathbf{q}_\nu^-)$ *of player* ν. *The realisation of this random variable given* $\omega \in \Omega$ *is noted* $\widehat{\mathcal{C}}_\nu(\mathbf{q}_\nu^+, \mathbf{q}_\nu^-, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega)$ *and defined as follows:*

$$\begin{aligned}
&\widehat{\mathcal{C}}_\nu\left(\mathbf{q}_\nu^+, \mathbf{q}_\nu^-, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega\right) \\
&\quad := \sum_{k=1}^{m-1}\left(\xi_{\nu,k}(\omega) + \gamma_k\left[\left(q_{\nu,k}^+ - q_{\nu,k}^-\right) + \left(\sum_{\tilde{\nu}\neq\nu}\left(q_{\tilde{\nu},\omega,k}^+ - q_{\tilde{\nu},\omega,k}^-\right)\right)\right]\right)\left(-\sum_{r=1}^{k}\left(q_{\nu,r}^+ - q_{\nu,r}^-\right)\right) \\
&\qquad + \sum_{k=1}^{m}\epsilon_k\left(q_{\nu,k}^+ + q_{\nu,k}^-\right) + \sum_{k=1}^{m}\left(\left(\frac{\eta_k}{\tau_k}\right)\left(\left(q_{\nu,k}^+\right)^2 + \left(q_{\nu,k}^-\right)^2\right)\right) \\
&\qquad + \sum_{k=1}^{m}\left(\left(\frac{\eta_k}{\tau_k}\right)\left(\sum_{\tilde{\nu}\neq\nu} q_{\nu,k}^+\, q_{\tilde{\nu},\omega,k}^+ + \sum_{\tilde{\nu}\neq\nu} q_{\nu,k}^-\, q_{\tilde{\nu},\omega,k}^-\right)\right).
\end{aligned} \tag{4.30}$$

Similarly to Huberman and Stanzl (2004, Definition 2), an *unbounded price manipulation* is defined as follows:

Definition 4.9 (Unbounded price manipulation). *Given the strategies of the other player* $\hat{\boldsymbol{y}}_{-\nu}$*, an* unbounded (risk neutral) price manipulation *for a player* ν *is a sequence* $\left\{\left(\mathbf{q}^+_{\nu,l}, \mathbf{q}^-_{\nu,l}\right)\right\}_{l=1}^{\infty}$ *of round-trip strategies with* $\lim_{l\to\infty} \mathbb{E}\left[\widehat{\mathcal{C}}_\nu\left(\mathbf{q}^+_{\nu,l}, \mathbf{q}^-_{\nu,l}, \hat{\boldsymbol{y}}_{-\nu}\right)\right] = -\infty$.

In Definition 4.10 we define some matrices, to which we often refer in the rest of this chapter. We should note that the matrix $\mathbf{M}$ defined in Definition 4.10 is the same as the matrix $\mathbf{M}$ defined in Lemma 1.1.

Definition 4.10.

$$\boldsymbol{\Lambda} = \begin{bmatrix} \gamma_1 & \gamma_2 & & \gamma_i & & \gamma_{m-1} \\ 0 & \gamma_2 & & \vdots & & \vdots \\ \vdots & \ddots & \ddots & \vdots & & \vdots \\ \vdots & & \ddots & \gamma_i & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \gamma_{m-1} \end{bmatrix}, \quad \boldsymbol{\Gamma} = \boldsymbol{\Lambda} + \boldsymbol{\Lambda}^{\mathrm{T}} = \begin{bmatrix} 2\gamma_1 & \gamma_2 & & \gamma_k & & \gamma_{m-1} \\ \gamma_2 & 2\gamma_2 & & \gamma_k & & \gamma_{m-1} \\ & & \ddots & \vdots & & \vdots \\ \gamma_k & \gamma_i & \cdots & 2\gamma_k & & \vdots \\ & & & & \ddots & \vdots \\ \gamma_{m-1} & \gamma_{m-1} & \cdots & \cdots & \cdots & 2\gamma_{m-1} \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} \frac{\eta_1}{\tau_1} & & \\ & \ddots & \\ & & \frac{\eta_{m-1}}{\tau_{m-1}} \end{bmatrix}, \qquad \mathbf{E}_m = \frac{\eta_m}{\tau_m}\, \mathbf{1}_{m-1\times m-1}, \qquad \mathbf{M} = 2\,\mathbf{E} + 2\,\mathbf{E}_m - \boldsymbol{\Gamma},$$

$$\widetilde{\mathbf{M}} := \begin{bmatrix} 2\,\mathbf{E} + 2\,\mathbf{E}_m - \boldsymbol{\Gamma} & -(2\,\mathbf{E}_m - \boldsymbol{\Gamma}) \\ -(2\,\mathbf{E}_m - \boldsymbol{\Gamma}) & 2\,\mathbf{E} + 2\,\mathbf{E}_m - \boldsymbol{\Gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{M} & -(2\,\mathbf{E}_m - \boldsymbol{\Gamma}) \\ -(2\,\mathbf{E}_m - \boldsymbol{\Gamma}) & \mathbf{M} \end{bmatrix},$$

$$\bar{\mathbf{M}} := \left[\begin{array}{cc|c} & \widetilde{\mathbf{M}} & \begin{matrix} -\left(\frac{2\eta_m}{\tau_m}\right)\mathbf{1}_{m-1} \\ \left(\frac{2\eta_m}{\tau_m}\right)\mathbf{1}_{m-1} \end{matrix} \\ \hline -\left(\frac{2\eta_m}{\tau_m}\right)\mathbf{1}^{\mathrm{T}}_{m-1} & \left(\frac{2\eta_m}{\tau_m}\right)\mathbf{1}^{\mathrm{T}}_{m-1} & \frac{4\eta_m}{\tau_m} \end{array}\right].$$

Lemma 4.2. *The matrix* $\mathbf{M}$ *is positive (*resp. *semi-) definite if and only if for all round trip strategy* $(\boldsymbol{q}^+, \boldsymbol{q}^-)$ *the following inequality holds:* $\boldsymbol{r}^{\mathrm{T}}\bar{\mathbf{M}}\boldsymbol{r} > 0$ *(*resp. $\boldsymbol{r}^{\mathrm{T}}\bar{\mathbf{M}}\boldsymbol{r} \geq 0$*) where* $\boldsymbol{r}$ *is defined as such* $\boldsymbol{r} := \left[\boldsymbol{q}^+_{[\![1,m-1]\!]}, \boldsymbol{q}^-\right]^{\mathrm{T}}$, i.e.

$$\mathbf{M} \succ 0 \Leftrightarrow \boldsymbol{r}^{\mathrm{T}}\bar{\mathbf{M}}\boldsymbol{r} > 0, \quad \forall \boldsymbol{r} := \left[\boldsymbol{q}^+_{[\![1,m-1]\!]}, \boldsymbol{q}^-\right]^{\mathrm{T}} \text{ with } \left(\boldsymbol{q}^+, \boldsymbol{q}^-\right) \in \mathcal{RT}. \tag{4.31}$$

Proof.

$\Leftarrow$: We assume that $\mathbf{M}$ has a non-positive eigenvalue $\lambda \leq 0$, then there exists a vector $\boldsymbol{d} \in \mathbb{R}^{m-1}\setminus\{\mathbf{0}\}$ such that $\boldsymbol{d}^{\mathrm{T}}\,\mathbf{M}\,\boldsymbol{d} \leq 0$. Based on $\boldsymbol{d}$, we define the round trip strategy $(\mathbf{q}^+, \mathbf{q}^-)$ where, for all $k \in [\![1, m-1]\!]$, we have

$$q^+_k = \max(d_k, 0) \quad \text{and} \quad q^-_k = \max(-d_k, 0)\,, \tag{4.32}$$

and

$$q^+_m = \max\left(-\sum_{k=1}^{m-1} d_k, 0\right) \quad \text{and} \quad q^-_m = \max\left(\sum_{k=1}^{m-1} d_k, 0\right). \tag{4.33}$$

If $\boldsymbol{r}$ is defined as $\left[\mathbf{q}^+_{[\![1,m-1]\!]}, \mathbf{q}^-\right]^{\mathrm{T}}$, then we have that $\boldsymbol{r}^{\mathrm{T}}\bar{\mathbf{M}}\boldsymbol{r} = \boldsymbol{d}^{\mathrm{T}}\,\mathbf{M}\,\boldsymbol{d} \leq 0$, which proves the implication $\Leftarrow$ of Equation (4.31).

$\Rightarrow$: We assume that there exists a round trip strategy $(\mathbf{q}^+, \mathbf{q}^-)$ such that $\boldsymbol{r}^{\mathrm{T}}\bar{\mathbf{M}}\boldsymbol{r} \leq 0$ with $\boldsymbol{r} = \left[\mathbf{q}^+_{[\![1,m-1]\!]}, \mathbf{q}^-\right]^{\mathrm{T}}$. Then, using a similar argument that in Lemma 4.1, we can define the round trip strategy $(\tilde{\mathbf{q}}^+, \tilde{\mathbf{q}}^-)$ with

$$\forall k \in [\![1,m]\!]: \quad \tilde{q}^+_k = \max\left(q^+_k - q^-_k,\, 0\right) \quad \text{and} \quad \tilde{q}^-_k = \max\left(q^-_k - q^+_k,\, 0\right) \tag{4.34}$$

such that $\tilde{\boldsymbol{r}}^{\mathrm{T}}\bar{\mathbf{M}}\tilde{\boldsymbol{r}} \leq \boldsymbol{r}^{\mathrm{T}}\bar{\mathbf{M}}\boldsymbol{r} \leq 0$, where $\tilde{\boldsymbol{r}} = \left[\tilde{\mathbf{q}}^+_{[\![1,m-1]\!]}, \tilde{\mathbf{q}}^-\right]^{\mathrm{T}}$. Moreover, if we define $\boldsymbol{d} \in \mathbb{R}^{m-1}$ with $d_k = q^+_k - q^-_k$ for all $k \in [\![1, m-1]\!]$, then $\boldsymbol{d}^{\mathrm{T}}\,\mathbf{M}\,\boldsymbol{d} = \tilde{\boldsymbol{r}}^{\mathrm{T}}\bar{\mathbf{M}}\tilde{\boldsymbol{r}} \leq \boldsymbol{r}^{\mathrm{T}}\bar{\mathbf{M}}\boldsymbol{r} \leq 0$, which contradicts the fact that $\mathbf{M}$ is positive definite.

The proof for the semi-positive definiteness follows the same logic. □

PROPOSITION 4.1. *If a market is free of price manipulation, then the matrix* $\mathbf{M}$ *is positive semi-definite.*

PROOF. The proof follows from Lemma 4.2 and Vaes and Hauser (2022, Lemma 2.3). □

Proposition 4.1 implies that the condition $\mathbf{M} \succ 0$ is necessary for having a viable market in the multiplayer framework is the same as the one derived in Vaes and Hauser (2022, Lemma 2.3) for the single player framework. In the next proposition, we demonstrate that the positive definiteness of the matrix $\mathbf{M}$ is also a sufficient condition for not having an unbounded price manipulation.

PROPOSITION 4.2. *If* $\mathbf{M}$ *is positive definite, then the market is free of unbounded price manipulations.*

PROOF. We assume by contradiction that there exists an unbounded price manipulation for a player $\nu \in [\![1,n]\!]$. Given an outcome $\omega \in \Omega$ and the strategies $\hat{\boldsymbol{y}}_{-\nu}$, the trading cost related to a round-trip strategy $(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu)$ of player ν is given by Equation (4.30). It can be split in two parts:

(i) $\mathrm{Lin}(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega)$, which grows linearly with $\mathbf{q}_\nu \coloneqq [\mathbf{q}^+_\nu, \mathbf{q}^-_\nu]^{\mathrm{T}}$ and depends on ω,

(ii) a quadratic term in $\mathbf{q}_\nu$ independent of ω,

and can be reformulated as follows:

$$\begin{aligned}\widehat{\mathcal{C}}_\nu\left(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega\right) &= \mathrm{Lin}\left(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega\right) \\ &\quad + \sum_{k=1}^{m-1} \gamma_k\left(q^+_{\nu,k} - q^-_{\nu,k}\right)\left(-\sum_{i=1}^{k}\left(q^+_{\nu,i} - q^-_{\nu,i}\right)\right) + \sum_{k=1}^{m}\left[\frac{\eta_k}{\tau_k}\left(\left(q^+_{\nu,k}\right)^2 + \left(q^-_{\nu,k}\right)^2\right)\right].\end{aligned}$$

As $(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu)$ is a round-trip strategy, the following constraint must be satisfied

$$\sum_{k=1}^{m} q^+_{\nu,k} - q^-_{\nu,k} = 0, \tag{4.35}$$

which means that the amount of long positions traded during the last trading period can be expressed in terms of the positions constituting the round-trip strategy: $q^+_{\nu,m} = q^-_{\nu,m} - \sum_{k=1}^{m-1}\left(q^+_{\nu,k} - q^-_{\nu,k}\right)$. Consequently, given a round-trip strategy $(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu)$, the associated trading cost can be reformulated as

$$\begin{aligned}
&\widehat{\mathcal{C}}_\nu\left(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega\right) \\
&\quad= \mathrm{Lin}\left(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega\right) + \sum_{k=1}^{m-1}\gamma_k\left(q^+_{\nu,k} - q^-_{\nu,k}\right)\left(-\sum_{i=1}^{k}\left(q^+_{\nu,i} - q^-_{\nu,i}\right)\right) \\
&\qquad+ \sum_{k=1}^{m-1}\left[\frac{\eta_k}{\tau_k}\left(\left(q^+_{\nu,k}\right)^2 + \left(q^-_{\nu,k}\right)^2\right)\right] + \left(\frac{\eta_m}{\tau_m}\right)\left(\left(q^-_{\nu,m} - \sum_{k=1}^{m-1}\left(q^+_{\nu,k} - q^-_{\nu,k}\right)\right)^2 + \left(q^-_{\nu,m}\right)^2\right) \\
&\quad= \mathrm{Lin}\left(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega\right) + \frac{1}{2}\boldsymbol{r}_\nu^{\mathrm{T}}\bar{\mathbf{M}}\boldsymbol{r}_\nu, \qquad (4.36)
\end{aligned}$$

with $\boldsymbol{r}_\nu \coloneqq \left[\mathbf{q}^+_{\nu,[\![1,m-1]\!]}, \mathbf{q}^-_\nu\right]^{\mathrm{T}}$ and $\bar{\mathbf{M}}$ as defined in Definition 4.10. By hypothesis there exists a sequence of round-trip strategies $\left\{\left(\mathbf{q}^+_{\nu,l}, \mathbf{q}^-_{\nu,l}\right)\right\}_{l=1}^{\infty}$ such that $\lim_{l\to\infty}\mathbb{E}\left[\widehat{\mathcal{C}}_\nu\left(\mathbf{q}^+_{\nu,l}, \mathbf{q}^-_{\nu,l}, \hat{\boldsymbol{y}}_{-\nu}\right)\right] = -\infty$. This implies that

$$\lim_{l\to\infty}\min_{\omega\in\Omega}\widehat{\mathcal{C}}_\nu\left(\mathbf{q}^+_{\nu,l}, \mathbf{q}^-_{\nu,l}, \hat{\boldsymbol{y}}_{-\nu}\,;\,\omega\right) = -\infty, \qquad (4.37)$$

which in turns implies that $\lim_{l\to\infty}\left\|\boldsymbol{r}_{\nu,l}\right\|_2 = +\infty$, with $\boldsymbol{r}_{\nu,l} \coloneqq \left[\mathbf{q}^+_{\nu,l,[\![1,m-1]\!]}, \mathbf{q}^-_{\nu,l}\right]^{\mathrm{T}}$, as the trading cost is a quadratic function. The greater the norm of the round-trip strategy, the more the quadratic part dominates the linear part. However, we have with Lemma 4.2 that the quadratic part is always greater than 0 since by assumption $\mathbf{M} \succ 0$, which leads to a contradiction and terminates the proof. □

Proposition 4.3. *For a player $\nu \in [\![1,n]\!]$, if $\lambda_\nu \in [0,1]$ and if $\mathbf{M} \succeq 0$ (*resp. *$\mathbf{M} \succ 0$), then the objective function $\hat{\theta}_\nu$ of Optimisation Problem (4.17) is convex (*resp. *strictly convex) on its feasible set $\widehat{\mathcal{Y}}_\nu$ for every $\hat{\boldsymbol{y}}_{-\nu}$. Moreover, $\mathbf{M} \succeq 0$ implies the existence and uniqueness of a convex set $\widehat{\mathcal{Y}}^\star \subseteq \widehat{\mathcal{Y}}$ of optimal execution strategies for (4.17). If $\mathbf{M} \succ 0$ then this set is a singleton.*

Proof. This is shown by combining the proofs of Theorem 2.1 and Proposition 4.2. □

Remark 4.1. *Implicitly, we have also proven that lifting the set of feasible strategies does not create new price manipulation opportunities as, based on Lemma 4.1, for any lifted strategy $\hat{\boldsymbol{y}}_\nu$ we can find a strategy $\boldsymbol{y}_\nu$ that has, for any realisation $\omega \in \Omega$, a trading cost lower or equal to the one obtained with the strategy $\hat{\boldsymbol{y}}_\nu$.*

4.3 Nash equilibrium

In the multiplayer framework described in this work, the players are assumed to be market takers in the sense that they exchange their positions on the market by matching their transactions with orders in the limit order book; independent market markers are on their side responsible for refilling the order book. This justifies the impact functions proposed by Almgren and Chriss

(2001). Here, we assume that no bilateral trades can be executed between the players; no over-the-counter trades (OTC) are allowed. As a consequence, the market clearing constraint, *i.e.* the constraint imposing that for each contract the number of short and long positions are equal, is not enforced as this is the role of market makers. As the limit order book has not infinite capacity on a particular trading period τ_k, $k \in [\![1, m]\!]$, we assume that the total volume that can be traded is limited, *i.e.*

ASSUMPTION 4.1. *For all player $\nu \in [\![1, n]\!]$, $\omega \in \Omega$ and all trading period τ_k, $k \in [\![1, m]\!]$, the variables $q^+_{\nu,\omega,k}$ and $q^-_{\nu,\omega,k}$ are respectively bounded by $\bar{q}^+_{\omega,k}$ and $\bar{q}^-_{\omega,k}$.*

For simplicity, we assume w.l.o.g. that $\bar{q}^+_{\omega,k} = \bar{q}^+_{\omega,k} = \bar{q}$ for all $k \in [\![1, m]\!]$ and $\omega \in \Omega$. When Assumption 4.1 is made, we say that we are in a *traded economy* (de Maere d'Aertrycke and Smeers, 2010, 2013). It is worth mentioning that, unlike de Maere d'Aertrycke and Smeers (2013), this limit on trading capacity is not shared among players and is not applied to represent the limited liquidity of the market, which is here modelled by the impact of the trades on the transaction prices. This limit can, for example, be chosen such that it would never be reached in practice. The benefit of adding such a constraint is to bound the feasible set of each player and to be able to apply the equilibrium existence theorem stated in Proposition 1.5 as we will see in the next section.

In practice, if we desire to check whether there exists a Nash equilibrium in the case where there is no limit on the traded volumes, we can apply the following strategy. First, fix a value for $\bar{q}$ and compute all equilibrium points. If for all equilibrium points there exists a $\nu \in [\![1, n]\!]$, $k \in [\![1, m]\!]$ and $\omega \in \Omega$ such that $q^+_{\nu,\omega,k} = \bar{q}$ or $q^-_{\nu,\omega,k} = \bar{q}$, then we can recompute the equilibrium points after having increased the value of $\bar{q}$. If no matter the value of $\bar{q}$, there is a traded volume that hits the bound $\bar{q}$, then there exists no equilibrium.

In the following, we assume that we are in a traded economy for a given value $\bar{q} > 0$, and redefine the feasible set of a player $\nu \in [\![1, n]\!]$ as follows

$$\widehat{\mathcal{Y}}_\nu \coloneqq \left\{ \left(\mathbf{q}^+_\nu, \mathbf{q}^-_\nu \right) \in \mathbb{R}_+^{(m-1)n_\Omega} \times \mathbb{R}_+^{mn_\Omega} \,\middle|\, \exists \boldsymbol{y}_\nu \in \mathcal{Y}_\nu : (4.39), (4.40) \text{ and } (4.41) \text{ hold} \right\}, \tag{4.38}$$

with

$$\forall \omega \in \Omega, \forall k \in [\![1, m-1]\!] : q^+_{\nu,\omega,k} - q^-_{\nu,\omega,k} = Q_{\nu,k}(\boldsymbol{y}_\nu \,;\, \omega)\,, \tag{4.39}$$

$$\forall \omega \in \Omega : q^-_{\nu,\omega,m} \geq \sum_{k=1}^{m-1} \left(q^+_{\nu,\omega,k} - q^-_{\nu,\omega,k} \right) - D_{\nu,m}(\omega)\,, \tag{4.40}$$

$$\forall \omega \in \Omega, \forall k \in [\![1, m-1]\!] : \quad q^+_{\nu,\omega,k} \leq \bar{q}, \quad q^-_{\nu,\omega,k} \leq \bar{q} \quad \text{and} \quad q^-_{\nu,\omega,m} \leq \bar{q}. \tag{4.41}$$

For a given $\omega \in \Omega$ and player $\nu \in [\![1, n]\!]$, we define the truncated feasible set $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}_\nu$ in a traded economy as follows

$$\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}_\nu \coloneqq \left\{ \left(\mathbf{q}^+_{\nu,\omega}, \mathbf{q}^-_{\nu,\omega} \right) \in \mathbb{R}_+^{m-1} \times \mathbb{R}_+^{m} \,\middle|\, \exists \boldsymbol{y}_\nu \in \mathcal{Y}_\nu : (4.39), (4.40) \text{ and } (4.41) \text{ hold for } \omega \right\}. \tag{4.42}$$

We denote the Cartesian product of the truncated feasible set of each player by

$$\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)} \coloneqq \widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}_1 \times \cdots \times \widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}_\nu \times \ldots \widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}_n. \tag{4.43}$$

To ease the notations, we define the quantities $p_\nu = (m-1)\,n_\Omega + mn_\Omega$ and $p = np_\nu$, which implies that, for all $\nu \in [\![1, n]\!]$, $\hat{\boldsymbol{y}}_\nu \subset \mathbb{R}_+^{p_\nu}$ and $\hat{\boldsymbol{y}} \subset \mathbb{R}_+^{p}$. We should also note that $\widehat{\mathcal{Y}} = \widehat{\mathcal{Y}}^{\mathrm{tr}(\omega_1)} \times \cdots \times \widehat{\mathcal{Y}}^{\mathrm{tr}(\omega_{n_\Omega})}$.

4.3.1 Existence of a Nash equilibrium

The existence of a Nash equilibrium in a traded economy (see with Assumption 4.1) can be shown with (Rosen, 1965, Theorem 1) or Facchinei and Pang (2003a, Theorem 12.3).

Proposition 4.4 (Existence of a Nash equilibrium). *A risk-averse Nash equilibrium of the n-player convex game $\mathcal{G}\big(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\big)$ exists in a traded economy.*

Proof. The result follows from Propositions 1.5 and 4.3 and the fact the feasible set is bounded with Assumption 4.1. □

4.3.2 Uniqueness of the Nash equilibrium

To study the uniqueness of a Nash equilibrium for the game $\mathcal{G}\big(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\big)$, it is interesting to study the solutions of the associated generalised variational inequality (GVI) problem $\mathrm{GVI}\big(\widehat{\mathcal{Y}}, \boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}\big)$ (Facchinei and Pang, 2003a). The GVI problem associated to $\mathcal{G}\big(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\big)$ consists of finding a pair $(\hat{\boldsymbol{y}}^\star, \hat{\boldsymbol{u}}^\star)$ with $\hat{\boldsymbol{u}}^\star \in \boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}(\hat{\boldsymbol{y}}^\star)$ such that

$$\langle \hat{\boldsymbol{u}}^\star\,,\, \hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}^\star \rangle \geq 0, \text{ for all } \hat{\boldsymbol{y}} \in \widehat{\mathcal{Y}}, \tag{4.44}$$

where $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}} : \mathbb{R}^p \supseteq \widehat{\mathcal{Y}} \rightrightarrows \mathbb{R}^p$ is the *pseudo-subdifferential set-valued map* defined as follows:

$$\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}(\hat{\boldsymbol{y}}) \coloneqq \left\{ \begin{pmatrix} \boldsymbol{g}_1 \\ \vdots \\ \boldsymbol{g}_n \end{pmatrix} \,\middle|\, \forall \nu \in [\![1, n]\!] : \boldsymbol{g}_\nu \in \boldsymbol{\partial}_\nu \hat{\theta}_\nu(\hat{\boldsymbol{y}}) \right\}, \tag{4.45}$$

where $\boldsymbol{\partial}_\nu \hat{\theta}_\nu(\hat{\boldsymbol{y}})$ denotes the subdifferential of the function $\hat{\theta}_\nu$ with respect to $\hat{\boldsymbol{y}}_\nu$ at the point $\hat{\boldsymbol{y}}$, *i.e.* the set of subgradients of $\hat{\theta}_\nu$ in $\hat{\boldsymbol{y}}_\nu$ at the point $\hat{\boldsymbol{y}}$. Any solution of the GVI problem is called a *variational equilibrium*, and we denote by $\mathrm{SOL}\Big(\mathrm{GVI}\big(\widehat{\mathcal{Y}}, \boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}\big)\Big)$ the set of all solutions. Studying these solutions is interesting since we have with Proposition 1.4 that any variational equilibrium of $\mathrm{GVI}\big(\widehat{\mathcal{Y}}, \boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}\big)$ is a Nash equilibrium of the game $\mathcal{G}\big(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\big)$. Moreover, since the payoff function $\hat{\theta}_\nu$ of each player $\nu \in [\![1, m]\!]$ is convex in $\hat{\boldsymbol{y}}_\nu$ (see Proposition 4.3), and since the joint feasible set $\widehat{\mathcal{Y}}$ is defined as the Cartesian product of the convex feasible set $\widehat{\mathcal{Y}}_\nu$ of each player $\nu \in [\![1, m]\!]$, we know that the converse also holds: any Nash equilibrium of $\mathcal{G}\big(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\big)$ is a variational equilibrium of $\mathrm{GVI}\big(\widehat{\mathcal{Y}}, \boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}\big)$ (Facchinei and Pang, 2009, Proposition 12.3).

As a consequence, if we can prove the uniqueness of the solution of $\mathrm{GVI}\big(\widehat{\mathcal{Y}}, \boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}\big)$, then we jointly prove the uniqueness of the Nash equilibrium. To do so, it is sufficient with Proposition 1.6 to prove that the set-valued map $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ is strictly monotone on $\widehat{\mathcal{Y}}$ (see Definition 1.5). To prove

the strict monotonicity of $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$, we first prove the strict monotonicity of the map $\boldsymbol{\nabla}\widehat{\mathcal{C}}(\bullet\,;\,\omega) : \mathbb{R}^p \supset \widehat{\mathcal{Y}} \to \mathbb{R}^p$ for any $\omega \in \Omega$, where $\boldsymbol{\nabla}\widehat{\mathcal{C}}(\bullet\,;\,\omega)$ is defined as follows:

$$\boldsymbol{\nabla}\widehat{\mathcal{C}}(\hat{\boldsymbol{y}}\,;\,\omega) \coloneqq \begin{bmatrix} \nabla_1 \widehat{\mathcal{C}}_1(\hat{\boldsymbol{y}}\,;\,\omega) \\ \nabla_2 \widehat{\mathcal{C}}_2(\hat{\boldsymbol{y}}\,;\,\omega) \\ \vdots \\ \nabla_n \widehat{\mathcal{C}}_n(\hat{\boldsymbol{y}}\,;\,\omega) \end{bmatrix}, \tag{4.46}$$

where $\nabla_\nu \widehat{\mathcal{C}}_\nu(\hat{\boldsymbol{y}}\,;\,\omega)$ denotes the gradient of the cost function $\widehat{\mathcal{C}}_\nu(\bullet\,;\,\omega)$ in $\hat{\boldsymbol{y}}_p$ at the point $\hat{\boldsymbol{y}}$. Note that this is well defined since, by lifting to feasible set $\mathcal{Y}_\nu$ to $\widehat{\mathcal{Y}}_\nu$, we obtain a differentiable trading cost function $\widehat{\mathcal{C}}_\nu$ for any $\omega \in \Omega$ (see Equation (4.18)).

Remark 4.2. *In Equation (4.46), $\nabla_\nu \widehat{\mathcal{C}}_\nu(\hat{\boldsymbol{y}}\,;\,\omega)$ corresponds to the gradient of the cost function $\widehat{\mathcal{C}}_\nu(\bullet\,;\,\omega)$ of player ν with regards to the variables of player ν,* i.e. *$\hat{\boldsymbol{y}}_\nu$, albeit $\widehat{\mathcal{C}}_\nu(\bullet\,;\,\omega)$ also depends on the variables $\hat{\boldsymbol{y}}_{-\nu}$ of the other players.*

Given a realisation $\omega \in \Omega$, only the variables $\mathbf{q}^+_{\nu,\omega}$ and $\mathbf{q}^-_{\nu,\omega}$ of each player $\nu \in [\![1, n]\!]$ have an impact on $\widehat{\mathcal{C}}(\bullet\,;\,\omega)$ and consequently on $\boldsymbol{\nabla}\widehat{\mathcal{C}}(\bullet\,;\,\omega)$. Hence, if for any $\hat{\boldsymbol{y}} \in \widehat{\mathcal{Y}}$ and $\omega \in \Omega$, we denote the vector of decisions of all the players related to ω as

$$\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)} \coloneqq \left[\mathbf{q}^+_{1,\omega}, \mathbf{q}^-_{1,\omega}, \ldots, \mathbf{q}^+_{\nu,\omega}, \mathbf{q}^-_{\nu,\omega}, \ldots, \mathbf{q}^+_{n,\omega}, \mathbf{q}^-_{n,\omega}\right], \tag{4.47}$$

and use the notation $\hat{\boldsymbol{y}}_\nu^{\mathrm{tr}(\omega)} \coloneqq \left[\mathbf{q}^+_{\nu,\omega}, \mathbf{q}^-_{\nu,\omega}\right]$, the cost function given ω is then equivalent to the function $\widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)} : \mathbb{R}^{2m-1} \to \mathbb{R}^{2m-1}$ defined as follows:

$$\begin{aligned} \widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) \coloneqq \sum_{k=1}^{m-1} \left(\xi_{\nu,k}(\omega) + \gamma_k \left(\sum_{\tilde{\nu} \in [\![1,n]\!]} \left(q^+_{\tilde{\nu},\omega,k} - q^-_{\tilde{\nu},\omega,k}\right)\right)\right)\left(D_{\nu,m}(\omega) - \sum_{i=1}^{k}\left(q^+_{\nu,\omega,i} - q^-_{\nu,\omega,i}\right)\right) \\ + \sum_{k=1}^{m} \epsilon_k \left(q^+_{\nu,\omega,k} + q^-_{\nu,\omega,k}\right) + \sum_{k=1}^{m} \frac{\eta_k}{\tau_k}\left(\sum_{\tilde{\nu} \in [\![1,n]\!]} \left(q^+_{\nu,\omega,k}\, q^+_{\tilde{\nu},\omega,k} + q^-_{\nu,\omega,k}\, q^-_{\tilde{\nu},\omega,k}\right)\right). \end{aligned} \tag{4.48}$$

Indeed, for any $\hat{\boldsymbol{y}} \in \widehat{\mathcal{Y}}$ and $\omega \in \Omega$, we observe that the right-hand side of Equations (4.18) and (4.48) are the same, and we have thus that

$$\widehat{\mathcal{C}}_\nu(\hat{\boldsymbol{y}}\,;\,\omega) = \widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right). \tag{4.49}$$

The gradient of the function $\widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)}$ with regards to $\hat{\boldsymbol{y}}_\nu^{\mathrm{tr}(\omega)}$ is noted $\nabla_\nu^{\mathrm{tr}(\omega)} \widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)}$ and defined as follows:

$$\nabla_\nu^{\mathrm{tr}(\omega)} \widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) \coloneqq \left[\frac{\partial \widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)}}{\partial q^+_{\nu,\omega,1}}, \ldots, \frac{\partial \widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)}}{\partial q^+_{\nu,\omega,m-1}}, \frac{\partial \widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)}}{\partial q^-_{\nu,\omega,1}}, \ldots, \frac{\partial \widehat{\mathcal{C}}_\nu^{\mathrm{tr}(\omega)}}{\partial q^-_{\nu,\omega,m}}\right]^{\mathrm{T}}, \tag{4.50}$$

where, for readability reasons, we have omitted to mention that the partial derivatives are evaluated at the point $\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}$. Finally, we define the truncated pseudo-gradient map $\boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)} :$

$\mathbb{R}^{(2m-1)n} \to \mathbb{R}^{(2m-1)n}$ as follows:

$$\boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) \coloneqq \begin{bmatrix} \nabla_1^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}_1^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) \\ \nabla_2^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}_2^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) \\ \vdots \\ \nabla_n^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}_n^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) \end{bmatrix}. \tag{4.51}$$

PROPOSITION 4.5. *If $\bar{\mathbf{M}} \succ 0$, then, for any realisation $\omega \in \Omega$, the truncated pseudo-gradient map $\boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}$ defined in (4.51) is strongly monotone on $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$ with constant $\frac{1}{2}\lambda_{\min}\left(\bar{\mathbf{M}}\right)$, where $\lambda_{\min}\left(\bar{\mathbf{M}}\right)$ denotes the smallest eigenvalue of $\bar{\mathbf{M}}$.*

PROOF. The proof is provided in Appendix B.2 at page 183. □

Risk-neutral case. The first situation for which we are interested to study the uniqueness of the Nash equilibrium is the case where any agent $\nu \in [\![1, n]\!]$ is risk-neutral, *i.e.* $\lambda_\nu = 0$. In this situation, the pseudo-subdifferential defined in (4.45) is singled valued and can be written as follows

$$\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}(\hat{\boldsymbol{y}}) = \sum_{\omega\in\Omega} \boldsymbol{\nabla}\widehat{\mathcal{C}}(\hat{\boldsymbol{y}}\,;\,\omega)\,\mathbb{P}[\omega]\,. \tag{4.52}$$

If we denote by $\omega_{\min}$ the least probable realisation of Ω, *i.e.*

$$\forall \omega \in \Omega : \mathbb{P}[\omega_{\min}] \leq \mathbb{P}[\omega]\,, \tag{4.53}$$

we have that the following proposition holds:

LEMMA 4.3. *If $\lambda_\nu = 0$ for any player $\nu \in [\![1, n]\!]$ and if $\bar{\mathbf{M}} \succ 0$, then, the pseudo-subdifferential map $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ defined in (4.45) is strongly monotone on $\widehat{\mathcal{Y}}$ with constant $\frac{1}{2}\lambda_{\min}\left(\bar{\mathbf{M}}\right)\mathbb{P}[\omega_{\min}]$.*

PROOF. Let $\hat{\boldsymbol{y}}_{(1)}$ and $\hat{\boldsymbol{y}}_{(2)}$ be two distinct points of $\widehat{\mathcal{Y}}$. With Equation (4.52) and Proposition 4.5, we have that

$$\left\langle \boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}\left(\hat{\boldsymbol{y}}_{(2)}\right) - \boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}\left(\hat{\boldsymbol{y}}_{(1)}\right)\,,\, \hat{\boldsymbol{y}}_{(2)} - \hat{\boldsymbol{y}}_{(1)} \right\rangle \tag{4.54}$$

$$\overset{(4.52)}{=} \sum_{\omega\in\Omega} \left\langle \boldsymbol{\nabla}\widehat{\mathcal{C}}\left(\hat{\boldsymbol{y}}_{(2)}\,;\,\omega\right) - \boldsymbol{\nabla}\widehat{\mathcal{C}}\left(\hat{\boldsymbol{y}}_{(1)}\,;\,\omega\right)\,,\, \hat{\boldsymbol{y}}_{(2)} - \hat{\boldsymbol{y}}_{(1)} \right\rangle \mathbb{P}[\omega] \tag{4.55}$$

$$= \sum_{\omega\in\Omega} \left\langle \boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}_{(2)}\right) - \boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}_{(1)}\right)\,,\, \hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)} \right\rangle \mathbb{P}[\omega] \tag{4.56}$$

$$\geq \frac{\lambda_{\min}\left(\bar{\mathbf{M}}\right)}{2} \sum_{\omega\in\Omega} \left(\left\| \hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)} \right\|^2 \mathbb{P}[\omega] \right) \tag{4.57}$$

$$\overset{(4.53)}{\geq} \frac{\lambda_{\min}\left(\bar{\mathbf{M}}\right)}{2} \mathbb{P}[\omega_{\min}] \sum_{\omega\in\Omega} \left(\left\| \hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)} \right\|^2 \right) \tag{4.58}$$

$$\geq \frac{\lambda_{\min}\left(\bar{\mathbf{M}}\right)}{2} \mathbb{P}[\omega_{\min}] \left\| \hat{\boldsymbol{y}}_{(2)} - \hat{\boldsymbol{y}}_{(1)} \right\|^2 > 0. \tag{4.59}$$

This completes the proof. □

THEOREM 4.2. *If* $\bar{\mathbf{M}} \succ 0$ *and if a risk-neutral Nash equilibrium to the game* $\mathcal{G}\left(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\right)$ *exists, then it is unique.*

PROOF. The results follows from Corollary 1.1 and Lemma 4.3 and the fact that the feasible sets of the players are independent of each other. □

REMARK 4.3. *Theorem 4.2 also holds for an infinite sample space* Ω.

THEOREM 4.3. *If* $\bar{\mathbf{M}} \succ 0$ *and if a risk-neutral Nash equilibrium to the game* $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ *exists, then it is unique.*

PROOF. The proof follows from Theorems 4.1 and 4.2. □

Risk-averse case. We now investigate the uniqueness of a risk-averse Nash equilibrium if one exists. As mentioned previously, we assume that each player ν evaluates risk with $\mathbb{E}\text{-CVaR}_{\alpha_\nu,\lambda_\nu}$, α_ν denoting the level of risk and λ_ν the risk-aversion. We have seen that such a risk measure accepts a dual representation (see Section 1.2), which corresponds to the expectation of the cost under a particular probability measure, called the worst case probability measure, that belongs to the associated risk-set $\mathcal{D}_{\mathbb{E}\text{-CVaR}_{\alpha,\lambda}}$ defined in Equation (1.26) as follows:

$$\mathcal{D}_{\mathbb{E}\text{-CVaR}_{\alpha,\lambda}} \coloneqq \left\{ \mu \in \mathcal{P}_\Omega \;\middle|\; \exists \upsilon \in \mathcal{P}_\Omega \text{ such that } \forall \omega \in \Omega : \begin{cases} \mu(\omega) = (1-\lambda)\,\mathbb{P}[\omega] + \lambda \upsilon(\omega) \\ 0 \leq \upsilon(\omega) \leq \frac{\mathbb{P}[\omega]}{\alpha} \end{cases} \right\}, \tag{4.60}$$

where $\mathcal{P}_\Omega$ is the set of probability measures on Ω.

Therefore, each player can potentially load each realisation $\omega \in \Omega$ in different ways into the expectation corresponding to the dual representation of their risk-measures since they can have different worst case probability measures $\{\mu_\nu^*\}_{\nu \in [\![1,n]\!]}$. According to (4.60), a player ν attributes a probability $\mu_\nu^*(\omega)$ to a realisation $\omega \in \Omega$ that must be comprised in the following range:

$$(1-\lambda_\nu)\,\mathbb{P}[\omega] \leq \mu_\nu^*(\omega) \leq \left((1-\lambda_\nu) + \frac{\lambda_\nu}{\alpha_\nu}\right)\mathbb{P}[\omega]\,. \tag{4.61}$$

If, for a player $\nu \in [\![1,n]\!]$ and a realisation $\omega \in \Omega$, we define the *risk-factor* $r_\nu(\omega)$ as the relative probability with regards to $\mathbb{P}[\omega]$ attributed to ω in the worst case probability measure μ_ν^*, then, we must have that

$$(1-\lambda_\nu) \leq r_\nu(\omega) \leq (1-\lambda_\nu) + \frac{\lambda_\nu}{\alpha_\nu}. \tag{4.62}$$

Hence, for a given realisation $\omega \in \Omega$, and for the risk-factors $\boldsymbol{r}(\omega) \coloneqq \{r_\nu(\omega)\}_{\nu \in [\![1,n]\!]}$, it is interesting to study whether the truncated pseudo-subdifferential map weighted by $\boldsymbol{r}(\omega)$, noted $\boldsymbol{\partial}^{\mathrm{tr}(\omega)}_{\boldsymbol{r}(\omega)}\widehat{\boldsymbol{\Theta}}(\bullet\,;\,\omega)$ and defined as follows in (4.63), is strongly monotone on $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$:

$$\boldsymbol{\partial}^{\mathrm{tr}(\omega)}_{\boldsymbol{r}(\omega)}\widehat{\boldsymbol{\Theta}}(\hat{\boldsymbol{y}}\,;\,\omega) \coloneqq \begin{bmatrix} r_1(\omega)\,\nabla_1^{\mathrm{tr}(\omega)}\,\widehat{C}_1^{\mathrm{tr}(\omega)}(\hat{\boldsymbol{y}}\,;\,\omega) \\ r_2(\omega)\,\nabla_2^{\mathrm{tr}(\omega)}\,\widehat{C}_2^{\mathrm{tr}(\omega)}(\hat{\boldsymbol{y}}\,;\,\omega) \\ \vdots \\ r_n(\omega)\,\nabla_n^{\mathrm{tr}(\omega)}\,\widehat{C}_n^{\mathrm{tr}(\omega)}(\hat{\boldsymbol{y}}\,;\,\omega) \end{bmatrix}. \tag{4.63}$$

To do so, we define the notions of risk-factors set and strong risk monotonicity as follows:

DEFINITION 4.11 (RISK-FACTORS). *Given the risk-aversion parameter λ_ν and the risk-level α_ν of each player $\nu \in [\![1, n]\!]$, we define the* risk-factors set $\mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ *as the polytope defined as follows:*

$$\mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \left\{ \boldsymbol{r} \in \mathbb{R}^n \,\middle|\, (1 - \lambda_\nu) \leq r_\nu \leq (1 - \lambda_\nu) + \frac{\lambda_\nu}{\alpha_\nu} \right\}. \tag{4.64}$$

DEFINITION 4.12 (RISK MONOTONICITY).

- *Given a realisation $\omega \in \Omega$, we say that the set-valued map $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ is* strongly risky monotone *on $\widehat{\mathcal{Y}}$ conditionally to the realisation ω, if the set-valued map $\boldsymbol{\partial}_{\boldsymbol{r}}^{\mathrm{tr}(\omega)}\widehat{\boldsymbol{\Theta}}(\bullet\,; \omega)$ defined in (4.63) is strongly monotone on $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$ for any $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$.*
- *We say that $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ is* strongly risky monotone *on $\widehat{\mathcal{Y}}$ if it is strongly risky monotone on $\widehat{\mathcal{Y}}$ conditionally to any realisation $\omega \in \Omega$.*

REMARK 4.4. *Since $\mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ is a polyhedron, if we desire to prove that the set-valued map $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ is strongly risky monotone on $\widehat{\mathcal{Y}}$ conditionally to the realisation ω, then it is sufficient to prove the set-valued map $\boldsymbol{\partial}_{\boldsymbol{r}}^{\mathrm{tr}(\omega)}\widehat{\boldsymbol{\Theta}}(\bullet\,; \omega)$ is strongly monotone on $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$ for any extreme point of $\mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$,* i.e. $\boldsymbol{r} \in \mathrm{Ext}(\mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda}))$.

Studying whether the set-valued map $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ is strongly risky monotone on $\widehat{\mathcal{Y}}$ is interesting as we hope that it implies that the set-valued map $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ is strongly monotone on $\widehat{\mathcal{Y}}$. To prove that the set-valued map $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ is strongly risky monotone on $\widehat{\mathcal{Y}}$, we define the following matrices that are parametrised by a risk-factors $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$:

$$\mathbf{J}_{\boldsymbol{r}}^{\mathrm{tr}} \coloneqq \begin{bmatrix} r_1 \mathbf{J}_{=}^{\mathrm{tr}} & r_1 \mathbf{J}_{\neq}^{\mathrm{tr}} & \cdots & \cdots & r_1 \mathbf{J}_{\neq}^{\mathrm{tr}} \\ r_2 \mathbf{J}_{\neq}^{\mathrm{tr}} & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & r_{n-1} \mathbf{J}_{\neq}^{\mathrm{tr}} \\ r_n \mathbf{J}_{\neq}^{\mathrm{tr}} & \cdots & \cdots & r_n \mathbf{J}_{\neq}^{\mathrm{tr}} & r_n \mathbf{J}_{=}^{\mathrm{tr}} \end{bmatrix}, \tag{4.65}$$

and

$$\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}} = \frac{1}{2}\left(\mathbf{J}_{\boldsymbol{r}}^{\mathrm{tr}} + \left(\mathbf{J}_{\boldsymbol{r}}^{\mathrm{tr}} \right)^{\mathrm{T}} \right), \tag{4.66}$$

where $\mathbf{J}_{=}^{\mathrm{tr}}$ and $\mathbf{J}_{\neq}^{\mathrm{tr}}$ are the matrices defined in Equations (B.11) and (B.12) respectively.

PROPOSITION 4.6. *If $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}} \succ 0$ for any risk-factors $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$, then the set-valued map $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ is strongly risky monotone on $\widehat{\mathcal{Y}}$.*

PROOF. According to Definition 4.12, $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ is strongly risk monotone on $\widehat{\mathcal{Y}}$ if the set-valued map $\boldsymbol{\partial}_{\boldsymbol{r}}^{\mathrm{tr}(\omega)}\widehat{\boldsymbol{\Theta}}(\bullet\,; \omega)$ is strongly monotone on $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$ for any $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ and $\omega \in \Omega$. By using a similar proof technique as Proposition 4.5, it can be shown that, for any $\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)} \in \widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$, the Jacobian of the truncated pseudo-gradient $\boldsymbol{\partial}_{\boldsymbol{r}}^{\mathrm{tr}(\omega)}\widehat{\boldsymbol{\Theta}}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\,; \omega \right)$ is equal to $\mathbf{J}_{\boldsymbol{r}}^{\mathrm{tr}}$. The rest of the proof follows the same steps as the one of Proposition 4.5 and uses the assumption that

$$\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}} = \frac{1}{2}\left(\mathbf{J}_{\boldsymbol{r}}^{\mathrm{tr}} + \left(\mathbf{J}_{\boldsymbol{r}}^{\mathrm{tr}} \right)^{\mathrm{T}} \right) \succ 0, \tag{4.67}$$

for any $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$. □

REMARK 4.5. *The condition that the matrix* $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}}$ *is positive definite for any* $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ *is a generalisation of the condition* $\bar{\mathbf{M}} \succ 0$ *of Theorem 4.3 in the risk-neutral case. Indeed, if* $\lambda_\nu = 0$ *for all* $\nu \in [\![1, n]\!]$*, then the risk-factors set* $\mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ *reduces to the singleton* $\{\mathbf{1}\}$*. Hence, if* $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}}$ *is positive definite, then so is* $\boldsymbol{\Delta}^{\mathrm{tr}}$*, which in turn implies that* $\bar{\mathbf{M}} \succ 0$ *with Corollary B.1.*

REMARK 4.6. *Verifying the condition that the matrix* $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}}$ *is positive definite for any* $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ *requires to check the positive definiteness of* 2^n *matrices, where* n *is the number of players. This holds since we have mentioned in Remark 4.4 that it suffices to show that* $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}}$ *is positive definite for the* 2^n *extreme points of* $\mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$*. Moreover, the number of elements in* $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}}$ *grows quadratically with the number of players* n *and with the number of trading periods* m*,* i.e. $\mathcal{O}\Big((mn)^2\Big)$*.*

REMARK 4.7. *The larger the number of players, the more restrictive is the condition that requires that the matrix* $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}}$ *is positive definite for any* $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$*. However, in practice only a small amount of players has market power, which implies that this condition holds true for reasonable risk-levels and risk-aversions.*

Based on these notions of risk-factors set and strong risk monotonicity, we formulate the following conjuncture:

CONJECTURE 4.1. *If*

(i) Ω *is a continuous sample space,*

(ii) $\forall \nu \in [\![1, n]\!]$*, the function* $\widehat{\mathcal{C}}_\nu(\bullet\,; \omega)$ *is differentiable in* $\hat{\boldsymbol{y}}$*,*

(iii) $\forall \hat{\boldsymbol{y}} \in \widehat{\mathcal{Y}}$*, the random variable* $\widehat{\mathcal{C}}_\nu(\hat{\boldsymbol{y}})$ *is absolutely continuous,*

(iv) $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ *is strongly risky monotone on* $\widehat{\mathcal{Y}}$*,*

then $\boldsymbol{\partial}\widehat{\boldsymbol{\Theta}}$ *is strongly monotone on* $\widehat{\mathcal{Y}}$*.*

THEOREM 4.4. *If Conjecture 4.1 holds, if* $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}} \succ 0$ *for any risk-factors* $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$*, and if a risk-averse Nash equilibrium to the game* $\mathcal{G}\Big(\widehat{\mathcal{Y}}, \widehat{\boldsymbol{\Theta}}\Big)$ *exists, then it is unique.*

THEOREM 4.5. *If Conjecture 4.1 holds, if* $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}} \succ 0$ *for any risk-factors* $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$*, and if a risk-averse Nash equilibrium to the game* $\mathcal{G}(\mathcal{Y}, \boldsymbol{\Theta})$ *exists, then it is unique.*

PROOF. The proof would follow from Theorems 4.1 and 4.4. □

4.4 Numerical results

In this section, we generalise the example provided in Chapter 2 to the multi-player case: we assume the situation with 3 players and an execution window subdivided into 5 trading periods, *i.e.* $n = 3$ and $m = 5$. We assume a market with a single stock that has a median daily volume V of 5 million shares. All other market parameters are shown in Table 4.1 and take the same values as those used in the numerical section of Chapter 2. It can be shown that given these market parameters values, the matrix $\bar{\mathbf{M}}$ defined in Definition 4.10 is positive definite.

TABLE 4.1: Market parameters in the multiplayer case

Acquisition time:	$T = t_f - t_s = 5$ [day]
Number of players:	$n = 3$
Risk-level parameter α:	$\forall \nu \in [\![1, n]\!] : \alpha_\nu = 0.3$
Number of trading periods:	$m = 5$
Trading period length:	$\forall k \in [\![1, m]\!] : \tau_k = T/m = 1$ [day]
Initial asset price:	$S_0 = 50$ [\$/share]
Asset's volatility:	$\sigma = 0.95$ [(\$/share)/day$^{1/2}$]
Price shifts distribution:	$\forall k \in [\![1, m]\!] : \xi_k \sim \mathcal{N}(0, \tau_k \sigma^2)$
Daily volume:	$V = 5 \cdot 10^6$ [share]
Volume forecast updates distribution:	$\forall k \in [\![1, m-1]\!] : \delta_k \sim \mathcal{N}(0, \tau_k \nu^2)$ with $\nu = 0.1 D_0$
Bid-ask spread:	$\forall k \in [\![1, m]\!] : b_k = 0.125$ [\$/share]
Fixed cost:	$\forall k \in [\![1, m]\!] : \epsilon_k = 0.5 b_k$ [\$/share]
Impact at 1% of market:	$\forall k \in [\![1, m]\!] : \eta_k = b_k/(0.01V)$ [(\$/share)/(share/day)]
Permanent impact parameter:	$\forall k \in [\![1, m]\!] : \gamma_k = (5 b_k)/(0.1V)$ [\$/share^2]

4.4.1 Risk-neutral traders

We first consider the case where all the traders are risk-neutral, *i.e.* $\forall \nu \in [\![1, n]\!] : \lambda_\nu = 0$. For this case, unlike reported in Table 4.1, we assume that the traders have different volume targets, *i.e.* $\mathbf{D}_0 = \{1.5 \cdot 10^6, 10^6, 0.5 \cdot 10^6\}$. Since the matrix $\bar{\mathbf{M}}$ is positive definite, there exists a unique risk-neutral Nash equilibrium, which is reported in Table 4.2.

Interestingly, we observe that all traders are front loading even if they are not sensitive to risk. When trading large amounts, the permanent impact of the trades shifts prices of subsequent trading periods. Hence, to avoid this price increase induced by the trades of the other players, which would be detrimental in terms of cost, traders prefer to trade more on the first trading periods even though the resulting temporary cost of these periods are higher. We also observe that the larger the volume to be traded, the more evenly the players spread their volumes over the trading periods, in order to avoid a too large temporary cost.

Finally, we clearly observe that a better social welfare could be attained if collusion and redistribution of the cost was accepted: indeed, in this case the situation could be seen as a single-player situation, where trades are more evenly distributed over the execution period and the additional cost due to the temporary impact of the trades is thereby reduced (see Chapter 2).

4.4.2 Risk-averse traders

We assume in this section that each of the traders wants to trade 0.5 million shares. We now turn our focus to the risk-averse case: we consider that the players have different values for the risk-aversion parameter, *i.e.* $\boldsymbol{\lambda} = \{0.5, 0.25, 0.0\}$. Given these and the market parameters values, we have that $\boldsymbol{\Delta}_{\boldsymbol{r}}^{\mathrm{tr}} \succ 0$ for any risk-factors $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$. This has been numerically verified by

TABLE 4.2: Risk-neutral Nash equilibrium ($\forall \nu \in [\![1, n]\!] : \lambda_\nu = 0$) given the parameters in Table 4.1

Player ν	$D_{\nu,0}$	VU	$y^\star_{\nu,1}$	$y^\star_{\nu,2}$	$y^\star_{\nu,3}$	$y^\star_{\nu,4}$	$y^\star_{\nu,5}$	$\theta_\nu(\boldsymbol{y}^\star)$ [\$]
1	$1.5 \cdot 10^6$	✓	2.479e-1	2.036e-1	1.895e-1	1.816e-1	1.772e-1	5.183e6
2	$1.0 \cdot 10^6$	✓	3.510e-1	2.649e-1	2.052e-1	1.174e-1	6.129e-2	3.234e6
3	$0.5 \cdot 10^6$	✓	5.968e-1	3.191e-1	7.713e-2	2.989e-2	-2.30e-2	1.486e6
Sum	$3.0 \cdot 10^6$	✓	-	-	-	-	-	**9.903e6**

checking the positive definiteness of $\boldsymbol{\Delta}^{\text{tr}}_{\boldsymbol{r}}$ for the 2^n extreme points of $\mathcal{R}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ (see Remark 4.6). Hence, if Conjecture 4.1 holds, we would have the uniqueness of the risk-averse Nash equilibrium. For the three different situations described below, we obtain numerical evidence for the uniqueness of this equilibrium by starting the search from different starting points and verifying that we converge to the same equilibrium.

In Tables 4.3 and 4.4, we report the risk-averse Nash equilibria and the associated payoffs in 3 distinct situations characterised by different attitudes towards the uncertainty related to the volume target:

(i) In Tables 4.3a and 4.4a, we consider the case where none of the players takes into account the uncertainty related to the volume target.

(ii) In Tables 4.3b and 4.4b, we consider the case where all players take into account the uncertainty related to the volume target.

(iii) Finally, in Tables 4.3c and 4.4c, we consider the case where only the first player considers the uncertainty related to the volume target.

In all situations, we observe that the risk-averse players (players 1 and 2) front load their trades, which is typical behaviour when being subject to risk (Almgren and Chriss, 2001). Also, we can draw conclusions similar to those in Chapter 2: players decrease their front loading when considering the uncertainty related to their volume targets.

The risk-neutral player (player 3) has on the other hand an interesting behaviour: in the situation where all players ignore volume uncertainty (VU), we observe in Table 4.3a that player 3 trades more on the last trading periods. This is explained by the fact the other players are front loading significantly and consequently induce a large temporary cost on the first trading periods, which player 3 avoids by shifting their trades to the last trading periods. However, this behaviour disappears when the traders considers VU as observed in Table 4.3b. In this situation, we have just mentioned that the risk-averse players reduce their traded volumes of the first periods when considering the risk associated to the uncertain volume target. As a consequence, the temporary cost of the first trading periods is reduced, and so player 3 is more inclined to trade during these periods in order to partially avoid the permanent effect of the trades of the other players.

Finally, when comparing the evolution of the payoffs in the different situations (see Table 4.4), we observe that the more risk-averse the trader, the less beneficial in terms of payoff is the equilibrium when all traders consider VU compared to the situation when they do not. Indeed, for player 1, the most risk-averse of all players, we observe an increase in the payoff, whereas the payoff of player 2 slightly decreases and the payoff of player 3 has a 2% decrease in payoff. Interestingly, even though player 3 is the only one considering the uncertainty related to the volume target (see Table 4.4c), we still observe an increase in the player's payoff function when compared to situation where all players ignore this source of uncertainty.

This observation leads us to draw the following conclusion: a more risk-averse trader tends to be more sensitive to the strategies of other players and more willing to adapt their strategy according to those of other players. As a result, they have less influence on the equilibrium that is reached. In other words, the less risk-averse a trader is, the less sensitive they are to the strategies of other players, and the more they can guide and force the convergence to an equilibrium which is favourable to them.

4.5 Conclusion

Our contributions in this chapter are as follows: we formulated the execution problem as an n-player risk-averse non-cooperative Nash game which includes the trades impact of all players on the price dynamics and the uncertainty that arises from both the price dynamics and the uncertain volume target. We then showed that this game can be reformulated as a generalised variational inequality, and proved the existence of a risk-averse Nash equilibrium. By proving the strict monotonicity of the payoff pseudo-subdifferential map for every realisation, we proved the uniqueness of such an equilibrium in the risk-neutral case. We also formulated a conjecture for the uniqueness of a risk-averse Nash equilibrium, which relies on the dual representation of the coherent risk measures. Finally, we presented numerical experiments which demonstrate that traders who are less risk-averse are less sensitive to the strategies of other players and benefit the most from taking into account volume uncertainty.

TABLE 4.3: Risk-averse Nash equilibrium for the risk-aversion parameters: $\boldsymbol{\lambda} = \{0.5, 0.25, 0.0\}$ and given the parameters in Table 4.1. The column VU indicates whether the player considers the uncertainty related to the volume target.

(a) None of the players consider the uncertainty related to the volume target; see associated payoffs in Table 4.4a.

Player ν	λ_ν	VU	$y^\star_{\nu,1}$	$y^\star_{\nu,2}$	$y^\star_{\nu,3}$	$y^\star_{\nu,4}$	$y^\star_{\nu,5}$
1	0.50	✗	6.513e-1	2.531e-1	9.193e-2	3.542e-3	7.25e-6
2	0.25	✗	4.246e-1	2.811e-1	1.757e-1	1.027e-1	1.566e-2
3	0.00	✗	1.856e-1	1.885e-1	1.879e-1	2.062e-1	2.316e-1

(b) All the players consider the uncertainty related to the volume target; see associated payoffs in Table 4.4b.

Player ν	λ_ν	VU	$y^\star_{\nu,1}$	$y^\star_{\nu,2}$	$y^\star_{\nu,3}$	$y^\star_{\nu,4}$	$y^\star_{\nu,5}$
1	0.50	✓	5.675e-1	2.76e-1	1.133e-1	4.287e-2	1.931e-4
2	0.25	✓	4.186e-1	2.694e-1	1.673e-1	8.941e-2	5.516e-2
3	0.00	✓	2.357e-1	2.010e-1	1.961e-1	1.931e-1	1.739e-1

(c) Only the first player considers the uncertainty related to the volume target; see associated payoffs in Table 4.4c.

Player ν	λ_ν	VU	$y^\star_{\nu,1}$	$y^\star_{\nu,2}$	$y^\star_{\nu,3}$	$y^\star_{\nu,4}$	$y^\star_{\nu,5}$
1	0.50	✓	5.581e-1	2.765e-1	1.197e-1	4.599e-2	-4.46e-4
2	0.25	✗	4.570e-1	2.764e-1	1.664e-1	8.684e-2	1.319e-2
3	0.00	✗	2.179e-1	1.881e-1	1.814e-1	1.882e-1	2.241e-1

TABLE 4.4: Payoffs associated to the risk-averse Nash equilibria reported in Table 4.3.

(a) None of the players consider the uncertainty related to the volume target; see associated strategies in Table 4.3a.

Player ν	λ_ν	VU	Expectation [$]	$\text{CVaR}_{0.3}$ [$]	$\theta_\nu(\boldsymbol{y}^\star)$ [$]
1	0.50	✗	0.879e6	1.236e6	1.057e6
2	0.25	✗	0.906e6	1.395e6	1.028e6
3	0.00	✗	0.971e6	1.709e6	0.971e6

(b) All the players consider the uncertainty related to the volume target; see associated strategies in Table 4.3b.

Player ν	λ_ν	VU	Expectation [$]	$\text{CVaR}_{0.3}$ [$]	$\theta_\nu(\boldsymbol{y}^\star)$ [$]
1	0.50	✓	0.874e5	1.276e6	1.075e6
2	0.25	✓	0.898e5	1.403e6	1.025e6
3	0.00	✓	0.948e5	1.630e6	0.948e5

(c) Only the first player considers the uncertainty related to the volume target; see associated strategies in Table 4.3c.

Player ν	λ_ν	VU	Expectation [$]	$\text{CVaR}_{0.3}$ [$]	$\theta_\nu(\boldsymbol{y}^\star)$ [$]
1	0.50	✓	0.879e6	1.285e6	1.082e6
2	0.25	✗	0.894e6	1.365e6	1.012e6
3	0.00	✗	0.961e6	1.673e6	0.961e6

Chapter 5

Solving strongly monotone generalised variational inequalities: theoretical and numerical approaches

We now focus on methods for finding the solution of a strongly monotone generalised variational inequality problem (GVIP). As mentioned in Chapter 3, GVIP are often encountered in generalised Nash equilibrium problems (GNEPs). Indeed, under appropriate assumptions, we can associate to a GNEP a GVIP, whose solutions form a subset of the set of generalised Nash equilibria.

This chapter builds on the background material on game theory (Section 1.3). The content of this chapter is the object of an article currently in preparation and soon available on *arXiv* (Vaes and Hauser, a).

5.1 Introduction

Many mathematical problems can be reduced to solving a *generalised variational inequality* (GVI) for which the operator is monotone. This is for instance the case for convex optimisation problems, fixed-point equations, equilibrium and complementarity problems (Facchinei and Pang, 2003a,b). Given a subset $\mathbf{K}$ of the d-dimensional Euclidean space $\mathbb{R}^d$ and a set-valued map $\boldsymbol{\Xi} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$, the *generalised variational inequality*, denoted by GVI$(\mathbf{K}, \boldsymbol{\Xi})$, is the problem of finding a pair $(\boldsymbol{y}^\star, \boldsymbol{u}^\star)$ with $\boldsymbol{y}^\star \in \mathbf{K}$ and $\boldsymbol{u}^\star \in \boldsymbol{\Xi}(\boldsymbol{y}^\star)$ such that

$$\langle \boldsymbol{u}^\star \, , \, \boldsymbol{y} - \boldsymbol{y}^\star \rangle \geq 0, \text{ for all } \boldsymbol{y} \in \mathbf{K} \, . \tag{5.1}$$

We denote by SOL(GVI$(\mathbf{K}, \boldsymbol{\Xi})$) the set of solutions of GVI$(\mathbf{K}, \boldsymbol{\Xi})$. If $\boldsymbol{\Xi}$ is a single-valued operator, the problem (5.1) reduces to a *variational inequality* (VI). In this work, we are interested in solving *generalised variational inequalities* (GVI) for which $\mathbf{K}$ is compact and convex, and $\boldsymbol{\Xi}$ is an upper hemicontinuous, compact- and convex-valued and strongly monotone set-valued map. These assumptions imply the existence and uniqueness of the solution of (5.1) (Chan and Pang, 1982, Corollary 3.1); we are consequently here interested in methods able to derive this solution.

From a theoretical perspective, it can be shown that finding the solution of such a GVI can be reduced to finding a point $\boldsymbol{y}^\star$ in $\mathbf{K}$ and an element $\boldsymbol{u}^\star$ of $\boldsymbol{\Xi}(\boldsymbol{y}^\star)$ such that $-\boldsymbol{u}^\star$ lies in the normal cone to $\mathbf{K}$ at the point $\boldsymbol{y}^\star$ (Facchinei and Pang, 2003a). This problem can in turn be reformulated as finding an equilibrium point to a differential inclusion of the type $\dot{\boldsymbol{y}} \in -\boldsymbol{\Phi}(\boldsymbol{y})$. In the case where $\boldsymbol{\Xi}$ is single-valued and hemicontinuous on $\mathbf{K}$, Rockafellar (1970b, Theorem 3) showed that the operator $\boldsymbol{\Phi}$ is maximally monotone. In Proposition 5.8, we show under appropriate assumptions that $\boldsymbol{\Phi}$ stays maximally monotone when it is an upper hemicontinuous set-valued map. The maximal monotonicity of the set-valued map $\boldsymbol{\Phi}$ implies that the *proximal point algorithm* (Martinet, 1970; Rockafellar, 1976) is well defined, which provides a fundamental algorithm to find an equilibrium point to to the differential inclusion (and consequently a solution to GVI($\mathbf{K}, \boldsymbol{\Xi}$)). This algorithm relies on applying successively for different values λ_i the *resolvent* operator, which, given a positive value $\lambda > 0$, is defined as $\mathbf{J}_\lambda^{\boldsymbol{\Phi}} \coloneqq (\mathbf{I} + \lambda\boldsymbol{\Phi})^{-1}$ where $\mathbf{I}$ is the identity operator. If the resolvent can be computed exactly, then strong convergence is assured if λ_i is bounded away from zero and if $\boldsymbol{\Phi}$ is strongly monotone with constant $\mu > 0$ (Rockafellar, 1976); the convergence proof relies on the fact that $\mathbf{J}_\lambda^{\boldsymbol{\Phi}}$ is single-valued and non-expansive (Ryu and Boyd, 2016), and that the maximal monotonicity of $\boldsymbol{\Phi}$ implies that $\mathbf{dom}\left(\mathbf{J}_\lambda^{\boldsymbol{\Phi}}\right) = \mathbb{R}^d$ (Minty, 1962, Corollary of Theorem 4). While being appealing by its simplicity, $\mathbf{I} + \lambda\boldsymbol{\Phi}$ can however be hard to invert, which prevents the usage of the method as it is (Eckstein and Bertsekas, 1992). Numerous of other techniques have then been proposed to solve GVI($\mathbf{K}, \boldsymbol{\Xi}$) such as, but not restricted to, *inexact methods* (*e.g.* He (1999); Burachik and Dutta (2010)), *splitting methods* (*e.g.* Eckstein and Bertsekas (1992); Lions and Mercier (1979); Passty (1979)), *gradient projection methods* (*e.g.* Nesterov and Scrimali (2011)), or *extragradient methods* (*e.g.* Nesterov (2007); Nesterov and Scrimali (2011); Censor et al. (2011); Yao et al. (2013)).

For a variational inequality problem VI($\mathbf{K}, \boldsymbol{\Xi}$) for which the operator $\boldsymbol{\Xi}$ is Lipschitz continuous and strongly monotone and therefore single-valued ($\boldsymbol{\Xi}$ is consequently a function on $\mathbb{R}^d$), Nesterov and Scrimali (2011) show that a solution can be found using a standard gradient-type method. They then provide an alternative method that relies on the technique of estimates functions (Nesterov, 2018) and on the extragradient method first suggested by Korpelevič (1976, 1983) and further developed by Nesterov (2007) in the dual space. This method has a better convergence rate compared to the standard gradient-type method while having a comparable computational cost. To the best of our knowledge, using such an algorithm to solve a generalised variational inequality, which is characterised by the fact that the operator $\boldsymbol{\Xi}$ is a set-valued map, has not been studied and constitutes the second part of this paper. As $\boldsymbol{\Xi}$ is set-valued, the conventional assumption of the Lipschitz continuity of the operator $\boldsymbol{\Xi}$ to study the convergence of the method cannot be made. Indeed, assuming the Lipschitz continuity would implicitly impose $\boldsymbol{\Xi}$ to be single-valued. We will however show that, under a similar condition as the Lipschitz continuity, the convergence to a neighbourhood of the solution of the GVI still holds.

Contents This chapter is organised as follows. In Section 5.2, we introduce all the definitions and objects used in this work. In Section 5.3, we state clearly the generalised variational inequality problem GVI($\mathbf{K}, \boldsymbol{\Xi}$) as well as the specific assumptions made in this paper. Subsequently in Section 5.3.2, we reformulate the GVI($\mathbf{K}, \boldsymbol{\Xi}$) problem as a differential inclusion of the form $\dot{\boldsymbol{y}} \in -\boldsymbol{\Phi}$ and prove that the corresponding set-valued map $\boldsymbol{\Phi}$ is *maximal monotone*, which is then used to prove the uniqueness of the equilibrium point of the differential inclusion. Under the assumption of *uniform monotonicity* of $\boldsymbol{\Xi}$, we prove that the system is asymptotically stable and that the unique equilibrium point is globally asymptotically stable. If additionally the $\boldsymbol{\Xi}$ is *strongly monotone*, we show that the variational equilibrium then becomes globally exponentially stable. In a second step, we present in Section 5.4 two algorithmic schemes: a subgradient-type method and a dual extrapolation method. The convergence analysis of both methods relies on the notion here introduced and called far-sighted Lipschitz continuity. Our dual extragradient method extends the method suggested by Nesterov and Scrimali (2011) for VIs to GVIs. In the same vein as their method, we define a strongly convex merit function and define the method able to iteratively generate points with decreasing merit function values. This method significantly outperforms the subgradient-type method as demonstrated theoretically and numerically in Section 5.4. Finally, Sections 5.5 and 5.6 discuss the results, propose avenues for further research on this subject, and conclude the paper.

5.2 Preliminaries and notation

Let $\mathcal{H} = \mathbb{R}^n$ be equipped with the Euclidean inner product $\langle \bullet , \bullet \rangle$. We denote by $\| \bullet \|$ the associated norm.

Topology

Definition 5.1. *Let $\mathcal{Y}$ be an Euclidean space and $\|\cdot\|$ the Euclidean norm. The* open ball *in $\mathcal{Y}$ of radius R centred at a point $\boldsymbol{y}_c$ denoted by $\mathcal{B}(\boldsymbol{y}\,;\,R)$ is defined by*

$$\mathcal{B}(\boldsymbol{y}\,;\,R) \coloneqq \{\boldsymbol{y} \in \mathcal{Y} \,|\, \|\boldsymbol{y} - \boldsymbol{y}_c\| < R\}\,. \tag{5.2}$$

The closed ball *in $\mathcal{Y}$ of radius R centred at a point $\boldsymbol{y}_c$ denoted by $\mathcal{B}[\boldsymbol{y}\,;\,R]$ is defined by*

$$\mathcal{B}[\boldsymbol{y}\,;\,R] \coloneqq \{\boldsymbol{y} \in \mathcal{Y} \,|\, \|\boldsymbol{y} - \boldsymbol{y}_c\| \leq R\}\,. \tag{5.3}$$

Convex analysis

Definition 5.2 (Normal cone). *Let $\mathcal{Y}$ be a closed, convex set of $\mathbb{R}^d$. The* normal cone *of $\mathcal{Y}$ is the set-valued map $\mathcal{N}_{\mathcal{Y}} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$, given by*

$$\mathcal{N}_{\mathcal{Y}}(\boldsymbol{y}) \coloneqq \begin{cases} \left\{\boldsymbol{n} \in \mathbb{R}^d \,\middle|\, \forall \boldsymbol{z} \in \mathcal{Y} : \langle \boldsymbol{n}\,,\, \boldsymbol{y} - \boldsymbol{z}\rangle \geq 0\right\} & \text{if } \boldsymbol{y} \in \mathcal{Y}, \\ \emptyset & \text{if } \boldsymbol{y} \notin \mathcal{Y}. \end{cases} \tag{5.4}$$



(a) (b)

FIGURE 5.1: Example of normal cones

For all $\boldsymbol{y} \in \mathcal{Y}$, the normal cone $\mathcal{N}_{\mathcal{Y}}(\boldsymbol{y})$ is a closed convex set. Any element of $\mathcal{N}_{\mathcal{Y}}(\boldsymbol{y})$ is said to be a normal vector *to $\mathcal{Y}$ at $\boldsymbol{y}$.*

DEFINITION 5.3 (MINIMAL NORM ELEMENT). *Given a closed convex set $\mathcal{Y}$, the* minimal norm element *$m(\mathcal{Y})$ of $\mathcal{Y}$ is the unique element of $\mathcal{Y}$ with minimal norm,* i.e.

$$m(\mathcal{Y}) = \arg\min_{\boldsymbol{z} \in \mathcal{Y}} \|\boldsymbol{z}\| . \tag{5.5}$$

DEFINITION 5.4. *The* orthogonal projection *of $\boldsymbol{y}$ onto a nonempty closed convex set $\mathcal{Y}$ is defined as*

$$\Pi_{\mathcal{Y}}(\boldsymbol{y}) = \arg\min_{\boldsymbol{z} \in \mathcal{Y}} \|\boldsymbol{z} - \boldsymbol{y}\| . \tag{5.6}$$

THEOREM 5.1 (FACCHINEI AND PANG (2003A, THEOREM 1.5.5)). *Let $\mathcal{Y}$ be a nonempty closed convex subset of $\mathbb{R}^d$. Then the following assertions hold:*

- *For each $\boldsymbol{y} \in \mathbb{R}^d$, $\Pi_{\mathcal{Y}}(\boldsymbol{y})$ exists and is unique.*
- *For each $\boldsymbol{y} \in \mathbb{R}^d$, $\Pi_{\mathcal{Y}}(\boldsymbol{y})$ is the unique vector $\bar{\boldsymbol{y}} \in \mathcal{Y}$ satisfying the inequality:*

$$\langle \boldsymbol{z} - \bar{\boldsymbol{y}} , \bar{\boldsymbol{y}} - \boldsymbol{y} \rangle \geq 0, \quad \forall \boldsymbol{z} \in \mathcal{Y}. \tag{5.7}$$

- *As a function of $\boldsymbol{y}$, $\Pi_{\mathcal{Y}}(\boldsymbol{y})$ is a non-expansive map,* i.e.

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d : \|\Pi_{\mathcal{Y}}(\boldsymbol{x}) - \Pi_{\mathcal{Y}}(\boldsymbol{y})\| \leq \|\boldsymbol{x} - \boldsymbol{y}\| . \tag{5.8}$$

Set-valued map

DEFINITION 5.5. *A* set-valued map *$\boldsymbol{\Psi}$ from $\mathcal{X}$ to $\mathcal{Y}$, also noted $\boldsymbol{\Psi} : \mathcal{X} \rightrightarrows \mathcal{Y}$, is a map that associates with any $\boldsymbol{x} \in \mathcal{X}$ a subset $\boldsymbol{\Psi}(\boldsymbol{x})$ of $\mathcal{Y}$. The domain of $\boldsymbol{\Psi}$ is noted $\mathbf{dom}(\boldsymbol{\Psi})$ and is defined as follows:*

$$\mathbf{dom}(\boldsymbol{\Psi}) \coloneqq \{\boldsymbol{x} \in \mathcal{X} \,|\, \boldsymbol{\Psi}(\boldsymbol{x}) \neq \emptyset\} . \tag{5.9}$$

The graph of $\boldsymbol{\Psi}$ is noted $\mathbf{Graph}(\boldsymbol{\Psi})$ and is defined as the subset of pairs $(\boldsymbol{x}, \boldsymbol{y})$ where $\boldsymbol{y} \in \boldsymbol{\Psi}(\boldsymbol{x})$:

$$\mathbf{Graph}(\boldsymbol{\Psi}) \coloneqq \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y} \,|\, \boldsymbol{y} \in \boldsymbol{\Psi}(\boldsymbol{x})\} . \tag{5.10}$$

DEFINITION 5.6 (MONOTICITY). *Let $\mathcal{X}$ be a Hilbert space, and let $\boldsymbol{\Psi}$ be a set-valued map from $\mathcal{X}$ into $\mathcal{X}$.*

- $\boldsymbol{\Psi}$ *is called* monotone *if and only if*

$$\forall \boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathbf{dom}(\boldsymbol{\Psi}), \forall \boldsymbol{v}_i \in \boldsymbol{\Psi}(\boldsymbol{x}_i), i \in \{1,2\} : \langle \boldsymbol{v}_1 - \boldsymbol{v}_2, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle \geq 0. \tag{5.11}$$

- $\boldsymbol{\Psi}$ *is called* strictly monotone *if and only if*

$$\forall \boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathbf{dom}(\boldsymbol{\Psi}), \forall \boldsymbol{v}_i \in \boldsymbol{\Psi}(\boldsymbol{x}_i), i \in \{1,2\} : \langle \boldsymbol{v}_1 - \boldsymbol{v}_2, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle > 0. \tag{5.12}$$

- $\boldsymbol{\Psi}$ *is called* uniformly monotone *with modulus* ψ*, where* $\psi : \mathbb{R}_+ \to [0, +\infty]$ *is an increasing function only vanishing at 0, if and only if*

$$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbf{dom}(\boldsymbol{\Psi}), \forall \boldsymbol{v}_i \in \boldsymbol{\Psi}(\boldsymbol{x}_i), i \in \{1,2\} : \langle \boldsymbol{v}_1 - \boldsymbol{v}_2, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle \geq \psi(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|). \tag{5.13}$$

- $\boldsymbol{\Psi}$ *is called* strongly monotone *if and only if there exists a constant* $\mu > 0$ *such that*

$$\forall \boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathbf{dom}(\boldsymbol{\Psi}), \forall \boldsymbol{v}_i \in \boldsymbol{\Psi}(\boldsymbol{x}_i), i \in \{1,2\} : \langle \boldsymbol{v}_1 - \boldsymbol{v}_2, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle \geq \mu \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2. \tag{5.14}$$

We say that $\boldsymbol{\Psi}$ *is* monotone on a set $\mathbf{K} \subseteq \mathbf{dom}(\boldsymbol{\Psi})$ *if Condition* (5.11) *holds for any* $\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathbf{K}$ *(and similarly for the other monotonicity definitions).*

It is clear from these definitions that strong monotonicity implies uniform monotonicity, which itself implies (strict) monotonicity.

Definition 5.7 (Lipschitz continuity). *Let* $\mathcal{X}$ *be a Hilbert space, and let* $\boldsymbol{\Psi}$ *be a set-valued map from* $\mathcal{X}$ *into* $\mathcal{X}$*.* $\boldsymbol{\Psi}$ *is said to be* Lipschitz continuous with constant L *if*

$$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbf{dom}(\boldsymbol{\Psi}), \forall \boldsymbol{v}_i \in \boldsymbol{\Psi}(\boldsymbol{x}_i), i \in \{1,2\} : \|\boldsymbol{v}_1 - \boldsymbol{v}_2\| \leq L \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|. \tag{5.15}$$

We say that $\boldsymbol{\Psi}$ *is* Lipschitz continuous with constant L on a set $\mathbf{K} \subseteq \mathbf{dom}(\boldsymbol{\Psi})$ *if Condition* (5.15) *holds for any* $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbf{K}$*.*

Definition 5.8 (Closed graph). *Let* $\mathcal{X} \subseteq \mathbb{R}^n$*,* $\mathcal{Y} \subseteq \mathbb{R}^m$ *and* $\boldsymbol{\Psi} : \mathcal{X} \rightrightarrows \mathcal{Y}$*.* $\boldsymbol{\Psi}$ *has* closed graph *if its graph is a closed subset of* $\mathcal{X} \times \mathcal{Y}$*.*

Definition 5.9 (Closed-valued). *A set-valued map* $\boldsymbol{\Psi} : \mathcal{X} \rightrightarrows \mathcal{Y}$ *is called* closed-valued *if* $\boldsymbol{\Psi}(\boldsymbol{x})$ *is a closed subset of* $\mathcal{Y}$ *for all* $\boldsymbol{x}$*.*

Definition 5.10 (Compact-valued). *A set-valued map* $\boldsymbol{\Psi} : \mathcal{X} \rightrightarrows \mathcal{Y}$ *is called* compact-valued *if* $\boldsymbol{\Psi}(\boldsymbol{x})$ *is a compact subset of* $\mathcal{Y}$ *for all* $\boldsymbol{x}$*.*

Definition 5.11 (Convex-valued). *A set-valued map* $\boldsymbol{\Psi} : \mathcal{X} \rightrightarrows \mathcal{Y}$ *is called* convex-valued *if* $\boldsymbol{\Psi}(\boldsymbol{x})$ *is a convex subset of* $\mathcal{Y}$ *for all* $\boldsymbol{x}$*.*

Definition 5.12 (Upper hemicontinuity). *Let $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathcal{Y} \subseteq \mathbb{R}^m$ and $\boldsymbol{\Psi} : \mathcal{X} \rightrightarrows \mathcal{Y}$. $\boldsymbol{\Psi}$ is* upper hemicontinuous *at $\boldsymbol{x}_0 \in \mathcal{X}$ if, for every open set $\mathcal{V} \supseteq \boldsymbol{\Psi}(\boldsymbol{x}_0)$, there is an open set $\mathcal{U}$ with $\boldsymbol{x}_0 \in \mathcal{U}$ such that*

$$\boldsymbol{\Psi}(\boldsymbol{x}) \subseteq \mathcal{V} \text{ for every } \boldsymbol{x} \in \mathcal{U}. \tag{5.16}$$

$\boldsymbol{\Psi}$ is upper hemicontinuous *if it is upper hemicontinuous at every $\boldsymbol{x} \in \mathcal{X}$.*

Proposition 5.1 (Aubin and Frankowska (2009, Proposition 1.4.8)). *Let $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathcal{Y} \subseteq \mathbb{R}^m$ and $\boldsymbol{\Psi} : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a set-valued map with closed domain. If $\boldsymbol{\Psi}$ is closed-valued and upper hemicontinuous, then $\boldsymbol{\Psi}$ has closed graph.*

Lemma 5.1. *Let $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathcal{Y} = \mathbb{R}^m$. Let $\boldsymbol{\Psi}_1 : \mathcal{X} \rightrightarrows \mathcal{Y}$ and $\boldsymbol{\Psi}_2 : \mathcal{X} \rightrightarrows \mathcal{Y}$ be two compact-valued and upper hemicontinuous set-valued maps. Then the set-valued map $\boldsymbol{\Psi} \coloneqq \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2$ is compact-valued and upper hemicontinuous.*

Proof. Let $\boldsymbol{x}_0 \in \mathbf{dom}(\boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2)$. By assumption, $\boldsymbol{\Psi}_1(\boldsymbol{x}_0)$ and $\boldsymbol{\Psi}_2(\boldsymbol{x}_0)$ are two compact sets, therefore $\boldsymbol{\Psi}(\boldsymbol{x}_0)$ which is defined as $\{\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2 \,|\, \boldsymbol{y}_1 \in \boldsymbol{\Psi}_1(\boldsymbol{x}_0) \text{ and } \boldsymbol{y}_2 \in \boldsymbol{\Psi}_2(\boldsymbol{x}_0)\}$ is a compact set in $\mathbb{R}^m$ as it results from the sum of two compact sets (Rudin, 1991); $\boldsymbol{\Psi}$ is consequently compact-valued. Let $\mathcal{V}$ be an open set such that $\boldsymbol{\Psi}(\boldsymbol{x}_0) \subseteq \mathcal{V}$. Then the complement $\mathcal{V}^c = \mathbb{R}^m \setminus \mathcal{V}$ of $\mathcal{V}$ is a closed set in $\mathbb{R}^m$, which is disjoint of $\boldsymbol{\Psi}(\boldsymbol{x}_0)$, *i.e.* $\mathcal{V}^c \cap \boldsymbol{\Psi}(\boldsymbol{x}_0) = \emptyset$. By the separation theorem (Rudin, 1991, Theorem 1.10), there exists $\varepsilon > 0$ such that

$$\left(\boldsymbol{\Psi}(\boldsymbol{x}_0) + \mathcal{B}\left(\mathbf{0}\,;\,\frac{\varepsilon}{2}\right)\right) \cap \left(\mathcal{V}^c + \mathcal{B}\left(\mathbf{0}\,;\,\frac{\varepsilon}{2}\right)\right) = \emptyset \Leftrightarrow (\boldsymbol{\Psi}(\boldsymbol{x}_0) + \mathcal{B}(\mathbf{0}\,;\,\varepsilon)) \cap \mathcal{V}^c = \emptyset, \tag{5.17}$$

where $\mathcal{B}(\boldsymbol{z}\,;\,\varepsilon)$ denotes the open ball of radius ε around $\boldsymbol{z}$, *i.e.* $\mathcal{B}(\boldsymbol{z}\,;\,\varepsilon) \coloneqq \{\boldsymbol{w} \,|\, \|\boldsymbol{w} - \boldsymbol{z}\|\} < \varepsilon$. However, we have that

$$\boldsymbol{\Psi}(\boldsymbol{x}_0) + \mathcal{B}(\mathbf{0}\,;\,\varepsilon) = \boldsymbol{\Psi}_1(\boldsymbol{x}_0) + \boldsymbol{\Psi}_2(\boldsymbol{x}_0) + \mathcal{B}\left(\mathbf{0}\,;\,\frac{\varepsilon}{2}\right) + \mathcal{B}\left(\mathbf{0}\,;\,\frac{\varepsilon}{2}\right). \tag{5.18}$$

Consequently, if we define the two open sets $\mathcal{V}_1 = \boldsymbol{\Psi}_1(\boldsymbol{x}_0) + \mathcal{B}(\mathbf{0}\,;\,\frac{\varepsilon}{2})$ and $\mathcal{V}_2 = \boldsymbol{\Psi}_2(\boldsymbol{x}_0) + \mathcal{B}(\mathbf{0}\,;\,\frac{\varepsilon}{2})$, we have that $\boldsymbol{\Psi}_1(\boldsymbol{x}_0) \subseteq \mathcal{V}_1$, $\boldsymbol{\Psi}_2(\boldsymbol{x}_0) \subseteq \mathcal{V}_2$ and $\mathcal{V}_1 + \mathcal{V}_2 \subseteq \mathcal{V}$. For $i \in \{1, 2\}$, the upper hemicontinuity of $\boldsymbol{\Psi}_i$ implies the existence of an open set $\mathcal{U}_i$ with $\boldsymbol{x}_0 \in \mathcal{U}_i$ such that $\forall \boldsymbol{x} \in \mathcal{U}_i$: $\boldsymbol{\Psi}_i(\boldsymbol{x}) \subseteq \mathcal{V}_i$. Therefore, if we define $\mathcal{U} = \mathcal{U}_1 \cap \mathcal{U}_2$, then $\forall \boldsymbol{x} \in \mathcal{U} : \boldsymbol{\Psi}_1(\boldsymbol{x}) + \boldsymbol{\Psi}_2(\boldsymbol{x}) \subseteq \mathcal{V}_1 + \mathcal{V}_2 \subseteq \mathcal{V}$, which concludes the proof. □

Lemma 5.2. *Let $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} = \mathbb{R}^m$. Assume that $\mathcal{X}$ is compact and that $\boldsymbol{\Psi} \colon \mathcal{X} \rightrightarrows \mathcal{Y}$ is an upper hemicontinuous set-valued map such that $\boldsymbol{\Psi}(\boldsymbol{x})$ is a bounded set for all $\boldsymbol{x} \in \mathcal{X}$. Then $\boldsymbol{\Psi}$ is bounded uniformly in $\mathcal{X}$, in the sense that there exists $R \geq 0$ such that*

$$\forall \boldsymbol{x} \in \mathcal{X}, \qquad \boldsymbol{\Psi}(\boldsymbol{x}) \subseteq \mathcal{B}(\mathbf{0}\,;\,R)\,. \tag{5.19}$$

Proof. Assume for contradiction that there does not exist R such that (5.19) holds. Then there is a sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ in $\mathcal{X}$ such that $\boldsymbol{\Psi}(\boldsymbol{x}_n) \not\subseteq \mathcal{B}(\mathbf{0}\,;\,n)$ for all $n \in \mathbb{N}$. By compactness of $\mathcal{X}$, there is a subsequence $\{\boldsymbol{x}_{n_k}\}_{k\in\mathbb{N}}$ such that $\boldsymbol{x}_{n_k} \to \boldsymbol{x}_\infty$ in the limit as $k \to \infty$, for some $\boldsymbol{x}_\infty \in \mathcal{X}$. Since $\boldsymbol{\Psi}$ takes values in bounded sets by assumption, there is $R_\infty > 0$ such that

$\boldsymbol{\Psi}(\boldsymbol{x}_\infty) \subseteq \mathcal{B}(\mathbf{0}\,;\, R_\infty)$, and by the upper hemicontinuity of $\boldsymbol{\Psi}$ there exists $\delta > 0$ such that $\boldsymbol{\Psi}(\boldsymbol{x}) \subset \mathcal{B}(\mathbf{0}\,;\, R_\infty)$ for all $\boldsymbol{x} \in \mathcal{X}$ within distance δ of $\boldsymbol{x}_\infty$. Therefore $\boldsymbol{\Psi}(\boldsymbol{x}_{n_k}) \subseteq \mathcal{B}(\mathbf{0}\,;\, R_\infty)$ for k sufficiently large since $\boldsymbol{x}_{n_k}$ converges to $\boldsymbol{x}_\infty$, but this contradicts the fact that $\boldsymbol{x}_{n_k} \not\subseteq \mathcal{B}(\mathbf{0}\,;\, n_k)$ for all k. □

PROPOSITION 5.2. *Let $\mathcal{X} \subseteq R^n$, and $\mathcal{Y} = \mathbb{R}^m$. Assume that $\mathcal{X}$ is compact and that $\boldsymbol{\Psi}_1 : \mathcal{X} \rightrightarrows \mathcal{Y}$ and $\boldsymbol{\Psi}_2 : \mathcal{X} \rightrightarrows \mathcal{Y}$ are upper hemicontinuous set-valued maps that are respectively compact-valued and closed-valued. Then the set-valued map $\boldsymbol{\Psi} \coloneqq \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2$ has closed graph.*

PROOF. Let $\{(\boldsymbol{x}_n, \boldsymbol{y}_n)\}_{n\in\mathbb{N}}$ be a sequence in the graph of $\boldsymbol{\Psi}$ that is convergent in $\mathcal{X} \times \mathcal{Y}$, *i.e.* $\boldsymbol{x}_n \in \mathcal{X}$ and $\boldsymbol{y}_n \in \boldsymbol{\Psi}(\boldsymbol{x}_n)$ for all n. As the sequence $\{(\boldsymbol{x}_n, \boldsymbol{y}_n)\}_{n\in\mathbb{N}}$ is convergent, there exists $K > 0$ such that $\|\boldsymbol{y}_n\| \leq K$ for all $n \in \mathbb{N}$. By definition of $\boldsymbol{\Psi}$, each $\boldsymbol{y}_n$ can be decomposed as $\boldsymbol{y}_n = \boldsymbol{y}_{n,1} + \boldsymbol{y}_{n,2}$ with $\boldsymbol{y}_{n,1} \in \boldsymbol{\Psi}_1(\boldsymbol{x}_n)$ and $\boldsymbol{y}_{n,2} \in \boldsymbol{\Psi}_2(\boldsymbol{x}_n)$. With Lemma 5.2, there exists $R_1 > 0$ such that $\boldsymbol{\Psi}_1$ is bounded uniformly in $\mathcal{X}$. Hence, $\|\boldsymbol{y}_{n,2}\| < K + R_1$ for all $n \in \mathbb{N}$. If we define the set-valued map $\boldsymbol{\Psi}_2^{R_2} \colon \mathcal{X} \rightrightarrows \mathcal{Y}$ as defined in Equation (5.20) for $R_2 = K + R_1$,

$$\boldsymbol{\Psi}_2^{R_2}(\boldsymbol{x}) \coloneqq \boldsymbol{\Psi}_2(\boldsymbol{x}) \cap \mathcal{B}[\mathbf{0}\,;\, R_2]\,, \tag{5.20}$$

then the sequence $\{(\boldsymbol{x}_n, \boldsymbol{y}_n)\}_{n\in\mathbb{N}}$ belongs to the graph of the set-valued map $\boldsymbol{\Psi}^{R_2} \coloneqq \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2^{R_2}$. However, $\boldsymbol{\Psi}_2^{R_2}$ is an upper hemicontinuous and compact-valued set-valued map, which implies with Lemma 5.1 that $\boldsymbol{\Psi}^{R_2}$ is upper hemicontinuous and compact-valued, which in turn implies with Proposition 5.1 that $\boldsymbol{\Psi}^{R_2}$ has closed graph. Consequently, the limit point $(\boldsymbol{x}_\infty, \boldsymbol{y}_\infty)$ of the sequence $\{(\boldsymbol{x}_n, \boldsymbol{y}_n)\}_{n\in\mathbb{N}}$ belongs to $\mathbf{Graph}\left(\boldsymbol{\Psi}^{R_2}\right)$, and *a fortiori* to $\mathbf{Graph}(\boldsymbol{\Psi})$ as $\mathbf{Graph}\left(\boldsymbol{\Psi}^{R_2}\right) \subseteq \mathbf{Graph}(\boldsymbol{\Psi})$. We have thus shown that any convergent sequence in $\mathbf{Graph}(\boldsymbol{\Psi})$ has its limit point that belongs to $\mathbf{Graph}(\boldsymbol{\Psi})$, which proves that the graph of $\boldsymbol{\Psi}$ is closed. □

PROPOSITION 5.3 (GREGORY (1980)). *Let $\mathcal{Y}$ be a closed, convex set of $\mathbb{R}^d$. The normal cone $\mathcal{N}_\mathcal{Y}$ of $\mathcal{Y}$ is an upper hemicontinuous set-valued map.*

DEFINITION 5.13 (NATURAL MAP). *Let $\mathcal{Y}$ be a closed, convex set of $\mathbb{R}^d$ and $\boldsymbol{\Psi} : \mathcal{Y} \rightrightarrows \mathbb{R}^d$ be an arbitrary set-valued map. The* natural map $\boldsymbol{\Psi}_\mathcal{Y}^{\mathbf{nat}} : \mathcal{Y} \rightrightarrows \mathbb{R}^d$ *associated to $\boldsymbol{\Psi}$ and $\mathcal{Y}$ is the set-valued map defined as follows:*

$$\boldsymbol{\Psi}_\mathcal{Y}^{\mathbf{nat}}(\boldsymbol{y}) \coloneqq \{\boldsymbol{y} - \Pi_\mathcal{Y}(\boldsymbol{y} - \boldsymbol{u}) \,|\, \boldsymbol{u} \in \boldsymbol{\Psi}(\boldsymbol{y})\}\,. \tag{5.21}$$

For a tuple $(\boldsymbol{y}, \alpha\boldsymbol{u})$ with $\boldsymbol{u} \in \boldsymbol{\Psi}(\boldsymbol{y})$ and $\alpha > 0$, we define the associated natural map as

$$\boldsymbol{\Psi}_\mathcal{Y}^{\mathbf{nat}}(\boldsymbol{y}, \alpha\boldsymbol{u}) \coloneqq \boldsymbol{y} - \Pi_\mathcal{Y}(\boldsymbol{y} - \alpha\boldsymbol{u})\,. \tag{5.22}$$

5.3 Generalised variational inequality

Throughout this chapter, we make the following assumption:

ASSUMPTION 5.1. *$\mathbf{K}$ is a compact convex set. $\boldsymbol{\Xi}$ is upper hemicontinuous, compact- and convex-valued, and strongly monotone with known constant μ on $\mathbf{K}$. In addition, it holds that $\mathbf{K} \subset \mathbf{dom}(\boldsymbol{\Xi})$.*

These assumptions imply the existence of a solution to (5.1) (Chan and Pang, 1982, Corollary 3.1). Similarly to the fact that the strict convexity of a function implies the uniqueness of its minimum if one exists, the strict monotonicity of the set-valued map $\boldsymbol{\Xi}$ implies that there is at most one solution to (5.1); this is shown in Proposition 5.4 for the sake of completeness.

PROPOSITION 5.4. *If $\boldsymbol{\Xi}$ is strictly monotone on $\mathbf{K}$, then* GVI$(\mathbf{K}, \boldsymbol{\Xi})$ *has at most one solution.*

PROOF. Assume the existence of two solutions $\boldsymbol{y}_1^\star$ and $\boldsymbol{y}_2^\star$, *i.e.* $\boldsymbol{y}_1^\star, \boldsymbol{y}_2^\star \in \mathrm{SOL}(\mathrm{GVI}(\mathbf{K}, \boldsymbol{\Xi}))$. Then with (5.1), for $i \in \{1, 2\}$, there exists a $\boldsymbol{u}_i^\star \in \boldsymbol{\Xi}(\boldsymbol{y}_i^\star)$ such that

$$\forall \boldsymbol{y} \in \mathbf{K} : \langle \boldsymbol{u}_i^\star \,,\, \boldsymbol{y} - \boldsymbol{y}_i^\star \rangle \geq 0, \tag{5.23}$$

which implies $\langle \boldsymbol{u}_1^\star \,,\, \boldsymbol{y}_2^\star - \boldsymbol{y}_1^\star \rangle \geq 0$ and $\langle \boldsymbol{u}_2^\star \,,\, \boldsymbol{y}_1^\star - \boldsymbol{y}_2^\star \rangle \geq 0$, which is equivalent to $\langle \boldsymbol{u}_1^\star \,,\, \boldsymbol{y}_1^\star - \boldsymbol{y}_2^\star \rangle \leq 0$ and $\langle -\boldsymbol{u}_2^\star \,,\, \boldsymbol{y}_1^\star - \boldsymbol{y}_2^\star \rangle \leq 0$. By summing these inequalities, we obtain $\langle \boldsymbol{u}_1^\star - \boldsymbol{u}_2^\star \,,\, \boldsymbol{y}_1^\star - \boldsymbol{y}_2^\star \rangle \leq 0$, which contradicts the strict monotonicity of $\boldsymbol{\Xi}$. □

In the rest of this chapter, we denote by $\boldsymbol{y}^\star$ the unique solution of GVI$(\mathbf{K}, \boldsymbol{\Xi})$.

5.3.1 Monotonicity property: interpretation

To gain intuition on how to solve a GVI, one should mention that the monotonicity property of a set-valued map can be seen as a generalisation of the subdifferential of a convex function. Indeed, if $f : \mathbb{R}^d \to \mathbb{R}$ is a proper convex function on a closed and convex set $\mathcal{X}$, then its subdifferential denoted by $\boldsymbol{\partial} f$ is a monotone set-valued map on $\mathcal{X}$ (Rockafellar, 1970a, Theorem A).

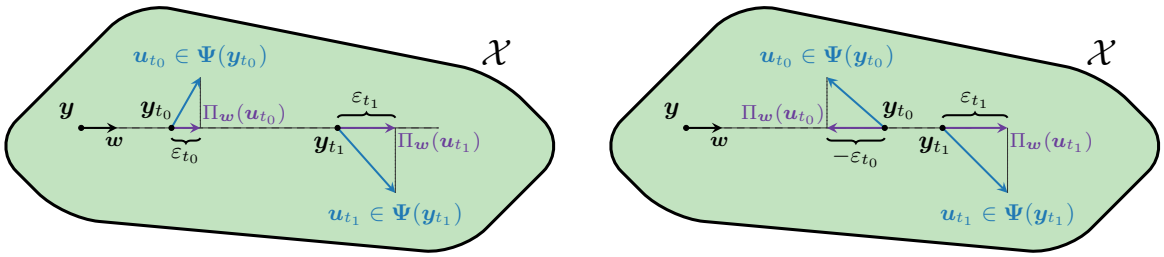
PROPOSITION 5.5. *A set-valued map $\boldsymbol{\Psi} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is monotone on a closed and convex set $\mathcal{X}$ if and only if $\boldsymbol{\Psi}$ is monotone over the set $\{\boldsymbol{y} + t\boldsymbol{w} \in \mathcal{X} \,|\, t \in \mathbb{R}\}$ for all $\boldsymbol{y} \in \mathcal{X}$ and $\boldsymbol{w} \in \mathbb{R}^d$.*

PROOF. The proof is straightforward. □

We are here interested in the graphical meaning of Proposition 5.5. For a given $\boldsymbol{y} \in \mathcal{X}$ and $\boldsymbol{w} \in \mathbb{R}^d$, we define the set-valued map $\boldsymbol{\varepsilon} : \mathbb{R} \rightrightarrows \mathbb{R}; t \mapsto \{\langle \boldsymbol{u}_t \,,\, \boldsymbol{w} \rangle \,|\, \boldsymbol{u}_t \in \boldsymbol{\Psi}(\boldsymbol{y}_t)\}$ with $\boldsymbol{y}_t = \boldsymbol{y} + t\boldsymbol{w}$. Then the monotonicity of $\boldsymbol{\Psi}$ implies that for all $0 \leq t_0 < t_1$ such that $\boldsymbol{y}_{t_i} \in \mathcal{X}$ and $\varepsilon_{t_i} \in \boldsymbol{\varepsilon}(t_i)$ for $i \in \{0, 1\}$, the following inequality holds:

$$\varepsilon_{t_0} \leq \varepsilon_{t_1}. \tag{5.24}$$

Formulated differently, for any point $\boldsymbol{y} \in \mathcal{X}$ and any direction $\boldsymbol{w}$, the further away from $\boldsymbol{y}$ is a point $\boldsymbol{y}_t$ in this direction $\boldsymbol{w}$, the larger is the scalar product of any element of $\boldsymbol{\Psi}(\boldsymbol{y}_t)$ with $\boldsymbol{w}$. This is clearly depicted in Figures 5.2a and 5.2b, where we denote by $\Pi_{\boldsymbol{w}}(\boldsymbol{u}_{t_i})$ the orthogonal projection of $\boldsymbol{u}_{t_i}$ onto the set $\{\boldsymbol{x} \,|\, \boldsymbol{x} = \boldsymbol{y} + t\boldsymbol{w} \text{ with } t \in \mathbb{R}\}$. Hence, if $\boldsymbol{\Xi}$ is strictly monotone on a compact convex set $\mathbf{K} \subseteq \mathbb{R}^d$ and if $\boldsymbol{y}^\star \in \mathbf{K}$ is the solution of SOL(GVI$(\mathbf{K}, \boldsymbol{\Xi})$), then for any direction $\boldsymbol{w} \in \mathbb{R}^d$ and any $t > 0$ such that $\boldsymbol{y}^\star + t\boldsymbol{w} \in \mathbf{K}$, every vector of $\boldsymbol{\Xi}(\boldsymbol{y}^\star + t\boldsymbol{w})$ has a positive component in the direction of $\boldsymbol{w}$. Moreover, the larger t, the greater this positive component. This partially explains why a *gradient-type* algorithm is able to solve GVI$(\mathbf{K}, \boldsymbol{\Xi})$ as proven later on in Section 5.4.1.



(a) $\varepsilon_0\,\varepsilon_1 > 0$ and $\varepsilon_0 < \varepsilon_1$ (b) $\varepsilon_0 < 0 < \varepsilon_1$

FIGURE 5.2

5.3.2 Theoretical approach: differential inclusion reformulation

In this section, we first show that a GVI($\mathbf{K}, \boldsymbol{\Xi}$) can be solved by finding an equilibrium point of a differential inclusion where at every point $\boldsymbol{y}(t)$ we move in the opposite direction of an element of $\boldsymbol{\Xi}(\boldsymbol{y}(t))$. We then prove that the equilibrium is globally asymptotically stable if $\boldsymbol{\Xi}$ is strictly monotone and globally exponentially stable if $\boldsymbol{\Xi}$ is strongly monotone.

Equation (5.1) tells us that a vector $\boldsymbol{y} \in \mathbf{K}$ solves GVI($\mathbf{K}, \boldsymbol{\Xi}$) if and only if there exists a vector $-\boldsymbol{u}$ with $\boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y})$ that is a normal vector to $\mathbf{K}$ at $\boldsymbol{y}$ (Facchinei and Pang, 2003a), *i.e.*

$$\mathbf{0} \in \boldsymbol{u} + \mathcal{N}_{\mathbf{K}}(\boldsymbol{y})\,, \tag{5.25}$$

where $\mathcal{N}_{\mathbf{K}}(\boldsymbol{y})$ is the normal cone to the closed convex set $\mathbf{K}$ at $\boldsymbol{y}$ (see Definition 5.2). We denote by $\widetilde{\boldsymbol{\Xi}}$ the set-valued map $\widetilde{\boldsymbol{\Xi}} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ that extends $\boldsymbol{\Xi}$ to $\mathbb{R}^d$, *i.e.*

$$\widetilde{\boldsymbol{\Xi}}(\boldsymbol{y}) = \begin{cases} \boldsymbol{\Xi}(\boldsymbol{y}) & \text{if } \boldsymbol{y} \in \mathbf{K}, \\ \emptyset & \text{if } \boldsymbol{y} \notin \mathbf{K}, \end{cases} \tag{5.26}$$

and we define the set-valued map $\boldsymbol{\Phi} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d; \boldsymbol{y} \mapsto \widetilde{\boldsymbol{\Xi}}(\boldsymbol{y}) + \mathcal{N}_{\mathbf{K}}(\boldsymbol{y})$, where by convention, for a set $\mathcal{S}$, $\emptyset + \mathcal{S} = \emptyset = \mathcal{S} + \emptyset$. Finally, we define the following differential inclusion

$$\dot{\boldsymbol{y}}(t) \in -\boldsymbol{\Phi}(\boldsymbol{y}) \Leftrightarrow \dot{\boldsymbol{y}}(t) \in -\Big(\widetilde{\boldsymbol{\Xi}}(\boldsymbol{y}) + \mathcal{N}_{\mathbf{K}}(\boldsymbol{y})\Big)\,. \tag{5.27}$$

An equilibrium point of (5.27) is a point $\boldsymbol{y}^\star \in \mathbf{K}$ such that

$$\mathbf{0} \in -\boldsymbol{\Phi}(\boldsymbol{y}^\star)\,. \tag{5.28}$$

PROPOSITION 5.6. *If $\boldsymbol{y}^\star \in \mathbf{K}$ is an equilibrium point of (5.27), then it is a solution of* GVI($\mathbf{K}, \boldsymbol{\Xi}$).

PROOF. If $\boldsymbol{y}^\star \in \mathbf{K}$ is an equilibrium point of (5.27), then there exist a $\boldsymbol{u}^\star \in \boldsymbol{\Xi}(\boldsymbol{y}^\star)$ and $\boldsymbol{n}^\star \in \mathcal{N}_{\mathbf{K}}(\boldsymbol{y}^\star)$, such that $\boldsymbol{u}^\star + \boldsymbol{n}^\star = \mathbf{0}$. This implies that for all $\boldsymbol{y} \in \mathbf{K}$

$$0 = \langle \boldsymbol{u}^\star + \boldsymbol{n}^\star\,,\, \boldsymbol{y} - \boldsymbol{y}^\star \rangle = \langle \boldsymbol{u}^\star\,,\, \boldsymbol{y} - \boldsymbol{y}^\star \rangle + \langle \boldsymbol{n}^\star\,,\, \boldsymbol{y} - \boldsymbol{y}^\star \rangle \leq \langle \boldsymbol{u}^\star\,,\, \boldsymbol{y} - \boldsymbol{y}^\star \rangle\,, \tag{5.29}$$

where the inequality arises from the definition of the normal cone (Definition 5.2). This shows that $\boldsymbol{y}^\star \in \mathrm{SOL}(\mathrm{GVI}(\mathbf{K}, \boldsymbol{\Xi}))$ since, according to (5.29), the tuple $(\boldsymbol{y}^\star, \boldsymbol{u}^\star)$ with $\boldsymbol{u}^\star \in \boldsymbol{\Xi}(\boldsymbol{y}^\star)$ is such that (5.1) holds for any $\boldsymbol{y} \in \mathbf{K}$. □

It is well known that if $\boldsymbol{\Phi}$ is strictly monotone on $\mathbf{K}$, then the differential inclusion (5.27) has at most one equilibrium point (Bauschke and Combettes, 2011, Proposition 23.35). This is in accordance with the previous results; if there were two distinct equilibrium points, it would imply with Proposition 5.6 that there are two solutions to GVI($\mathbf{K}, \boldsymbol{\Xi}$), which would contradict Proposition 5.4.

It therefore makes sense to analyse the solution of the differential inclusion (5.27) since any method able to find an equilibrium point of (5.27) can also be used to find a solution to GVI($\mathbf{K}, \boldsymbol{\Xi}$). To do so, we rely on the following definitions: A *(Carathéodory) solution* of (5.27) defined on $[t_0, t_1] \subset [0, \infty)$ is an absolutely continuous map $\boldsymbol{y} : [t_0, t_1] \to \mathbb{R}^d$ such that $\dot{\boldsymbol{y}}(t) \in -\boldsymbol{\Phi}(\boldsymbol{y}(t))$ for almost every $t \in [t_0, t_1]$ (Cortes, 2008). A solution of (5.27) is said to be a *slow solution* if, for almost all $t \in [t_0, t_1]$, $\dot{\boldsymbol{y}}(t) = -m(\boldsymbol{\Phi}(\boldsymbol{y}(t)))$, where $m(\boldsymbol{\Phi}(\boldsymbol{y}(t)))$ is the minimal norm element (see Definition 5.3). A monotone set-valued map $\boldsymbol{\Psi}$ is *maximal* if there is no other monotone set-valued map whose graph contains strictly the graph of $\boldsymbol{\Psi}$; a more formal definition can be found in (Aubin and Cellina, 1984, Proposition 3.1.1.).

The *maximal monotonicity* of a set-valued-map is a key property for the analysis of the solution of a differential inclusion. In order to prove that the set-valued map $\boldsymbol{\Phi}$ of (5.27) is maximal monotone, one relies on the next proposition.

PROPOSITION 5.7 (VOISEI (2012, THEOREM 20)). *A set-valued map $\boldsymbol{\Psi} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is maximal monotone if and only if*

(i) $\boldsymbol{\Psi}$ *is monotone on* $\mathbb{R}^d$,

(ii) $\mathbf{dom}(\boldsymbol{\Psi})$ *is nearly-convex, that is, there is a convex set* C *such that* $C \subseteq \mathbf{dom}(\boldsymbol{\Psi}) \subseteq \mathbf{cl}(C)$,

(iii) $\boldsymbol{\Psi}(\boldsymbol{y})$ *is convex-valued,*

(iv) $\boldsymbol{\Psi} = \boldsymbol{\Psi} + \mathcal{N}_{\mathbf{dom}(\boldsymbol{\Psi})}$,

(v) The graph of $\boldsymbol{\Psi}$ *is closed.*

PROPOSITION 5.8. *If Assumption 5.1 holds. Then the set-valued map $\boldsymbol{\Phi}$ defined as in (5.27) is maximal monotone.*

PROOF. We show that all necessary and sufficient conditions of Proposition 5.7 are verified.
• Proposition 5.7i: By definition of $\boldsymbol{\Phi}$, $\mathbf{dom}(\boldsymbol{\Phi}) = \mathbf{K}$. Hence, $\boldsymbol{\Phi}$ is monotone if $\forall \boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbf{K}$, $\boldsymbol{v}_1 \in \boldsymbol{\Phi}(\boldsymbol{y}_1)$ and $\boldsymbol{v}_2 \in \boldsymbol{\Phi}(\boldsymbol{y}_2)$ the following inequality holds: $\langle \boldsymbol{v}_2 - \boldsymbol{v}_1 \, , \, \boldsymbol{y}_2 - \boldsymbol{y}_1 \rangle \geq 0$. If $\boldsymbol{v}_i \in \boldsymbol{\Phi}(\boldsymbol{y}_i)$, there exists a tuple $(\boldsymbol{u}_i, \boldsymbol{n}_i)$ with $\boldsymbol{u}_i \in \boldsymbol{\Xi}(\boldsymbol{y}_i)$ and $\boldsymbol{n}_i \in \mathcal{N}_{\mathbf{K}}(\boldsymbol{y}_i)$ such that $\boldsymbol{v}_i = \boldsymbol{u}_i + \boldsymbol{n}_i$ for $i \in \{1, 2\}$. Hence,

$$\begin{aligned} \langle \boldsymbol{v}_2 - \boldsymbol{v}_1 \, , \, \boldsymbol{y}_2 - \boldsymbol{y}_1 \rangle &= \langle \boldsymbol{u}_2 + \boldsymbol{n}_2 - (\boldsymbol{u}_1 + \boldsymbol{n}_1) \, , \, \boldsymbol{y}_2 - \boldsymbol{y}_1 \rangle \, , \\ &= \langle \boldsymbol{u}_2 - \boldsymbol{u}_1 \, , \, \boldsymbol{y}_2 - \boldsymbol{y}_1 \rangle + \langle \boldsymbol{n}_2 \, , \, \boldsymbol{y}_2 - \boldsymbol{y}_1 \rangle + \langle \boldsymbol{n}_1 \, , \, \boldsymbol{y}_1 - \boldsymbol{y}_2 \rangle \geq 0, \end{aligned} \tag{5.30}$$

where the inequality holds by the monotonicity of $\boldsymbol{\Xi}$ on $\mathbf{K}$ and by definition of the normal cone at a point $\boldsymbol{y}_i \in \mathbf{K}$ (see Definition 5.2).

• Proposition 5.7ii: $\mathbf{dom}(\boldsymbol{\Phi}) = \mathbf{K}$ is a closed convex set by assumption and is thus also nearly-convex.

• Proposition 5.7iii: If $\boldsymbol{y} \notin \mathbf{K}$, then $\boldsymbol{\Phi}(\boldsymbol{y}) = \emptyset$, which is a convex set. If $\boldsymbol{y} \in \mathbf{K}$ and $\boldsymbol{\Phi}(\boldsymbol{y})$ is a singleton, then $\boldsymbol{\Phi}(\boldsymbol{y})$ is convex. If $\boldsymbol{y} \in \mathbf{K}$ and $\boldsymbol{\Phi}(\boldsymbol{y})$ is not a singleton, we must show that for any $\boldsymbol{v}_1 \neq \boldsymbol{v}_2$ in $\boldsymbol{\Phi}(\boldsymbol{y})$ and for any $t \in (0,1)$, $\boldsymbol{v}_t \coloneqq t\boldsymbol{v}_1 + (1-t)\,\boldsymbol{v}_2$ is an element of $\boldsymbol{\Phi}(\boldsymbol{y})$. Since $\boldsymbol{v}_i \in \boldsymbol{\Phi}(\boldsymbol{y}_i)$, there exists a tuple $(\boldsymbol{u}_i, \boldsymbol{n}_i)$ with $\boldsymbol{u}_i \in \boldsymbol{\Xi}(\boldsymbol{y})$ and $\boldsymbol{n}_i \in \mathcal{N}_{\mathbf{K}}(\boldsymbol{y})$ such that $\boldsymbol{v}_i = \boldsymbol{u}_i + \boldsymbol{n}_i$ for $i \in \{1,2\}$. Plugging this in the definition of $\boldsymbol{v}_t$ gives $\boldsymbol{v}_t = (t\boldsymbol{u}_1 + (1-t)\,\boldsymbol{u}_2) + (t\boldsymbol{n}_1 + (1-t)\,\boldsymbol{n}_2)$. By the convexity of $\boldsymbol{\Xi}(\boldsymbol{y})$ ($\boldsymbol{\Xi}$ being convex-valued) and $\mathcal{N}_{\mathbf{K}}(\boldsymbol{y})$, $\boldsymbol{u}_t \coloneqq t\boldsymbol{u}_1 + (1-t)\,\boldsymbol{u}_2 \in \boldsymbol{\Xi}(\boldsymbol{y})$ and $\boldsymbol{n}_t \coloneqq t\boldsymbol{n}_1 + (1-t)\,\boldsymbol{n}_2 \in \mathcal{N}_{\mathbf{K}}(\boldsymbol{y})$, which in turn implies by the definition of $\boldsymbol{\Phi}$ that $\boldsymbol{v}_t = \boldsymbol{u}_t + \boldsymbol{n}_t \in \boldsymbol{\Phi}(\boldsymbol{y})$.

• Proposition 5.7iv: Since $\mathcal{N}_{\mathbf{K}}$ is a convex cone, $\mathcal{N}_{\mathbf{K}} = \mathcal{N}_{\mathbf{K}} + \mathcal{N}_{\mathbf{K}}$. Therefore $\boldsymbol{\Phi} = \widetilde{\boldsymbol{\Xi}} + \mathcal{N}_{\mathbf{K}} = \widetilde{\boldsymbol{\Xi}} + \mathcal{N}_{\mathbf{K}} + \mathcal{N}_{\mathbf{K}} = \boldsymbol{\Phi} + \mathcal{N}_{\mathbf{K}}$.

• Proposition 5.7v: On the one hand, $\boldsymbol{\Xi}$ is by hypothesis compact-valued and upper hemicontinuous and so is $\widetilde{\boldsymbol{\Xi}}$. On the other hand, in light of Definition 5.2 and Proposition 5.3, the normal cone $\mathcal{N}_{\mathbf{K}}$ of $\mathbf{K}$ is an upper hemicontinuous and closed-valued set-valued map. Hence, $\boldsymbol{\Phi}$ has closed-graph using Proposition 5.2. □

Proposition 5.8 implies that if $\mathbf{K}$ is a compact convex set and if $\boldsymbol{\Xi}$ is upper hemicontinuous, compact- and convex-valued and (strongly) monotone on $\mathbf{K}$, then the set-valued map $\boldsymbol{\Phi}$ of (5.27) is maximal monotone, which in turns implies the following result:

THEOREM 5.2. *If $\boldsymbol{\Phi}$ is maximal monotone, then for any initial point $\boldsymbol{y}(0) = \boldsymbol{y}_0 \in \mathbf{dom}(\boldsymbol{\Phi}) = \mathbf{K}$ there exists a unique solution $\boldsymbol{y}(\cdot)$ to (5.27) defined on $[0,\infty)$, which is the slow solution,* i.e. *for almost all $t \in \mathbb{R}^+$ $\dot{\boldsymbol{y}}(t) = -m(\boldsymbol{\Phi}(\boldsymbol{y}(t)))$. Moreover,*

(i) $t \to \|\dot{\boldsymbol{y}}(t)\|$ is non-increasing.

(ii) Let $\boldsymbol{x}(\cdot)$ and $\boldsymbol{y}(\cdot)$ be the solutions issued from two distinct initial points $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ in $\mathbf{dom}(\boldsymbol{\Phi})$. Then for $t \geq 0$, $\|\boldsymbol{x}(t) - \boldsymbol{y}(t)\| \leq \|\boldsymbol{x}_0 - \boldsymbol{y}_0\|$.

(iii) $\forall t : \dot{\boldsymbol{y}}(t) \coloneqq \lim_{h \to 0^+} \frac{\boldsymbol{y}(t+h) - \boldsymbol{y}(t)}{h}$ and $\dot{\boldsymbol{y}}(\,\cdot\,)$ is continuous from the right.

If an equilibrium point $\boldsymbol{y}^\star$ exists then

(iv) If additionally $\boldsymbol{\Phi}$ is strictly monotone on $\mathbf{K}$, then $t \to \|\boldsymbol{y}(t) - \boldsymbol{y}^\star\|$ is decreasing for almost all t.

(v) If additionally $\boldsymbol{\Phi}$ is uniformly monotone on $\mathbf{K}$ with modulus ψ, then

$$\lim_{t \to \infty} \boldsymbol{y}(t) = \boldsymbol{y}^\star.$$

(vi) If additionally $\boldsymbol{\Phi}$ is strongly monotone on $\mathbf{K}$ with constant μ, then

$$\|\boldsymbol{y}(t) - \boldsymbol{y}^\star\| \leq \|\boldsymbol{y}(0) - \boldsymbol{y}^\star\| \exp(-\mu t)\,, \qquad \textit{for almost all } t \geq 0. \tag{5.31}$$

If any of the conditions iv, v or vi holds, then $\boldsymbol{y}^\star$ is the unique equilibrium point.

PROOF. The fact that the solution is the slow solution as well as statements Theorem 5.2i, 5.2ii or 5.2iii are shown in the proof of Aubin and Cellina (1984, Theorem 3.2.1.). If any of the conditions Theorem 5.2iv, 5.2v or 5.2vi holds, the uniqueness of the equilibrium point $\boldsymbol{y}^\star$ can be shown with Bauschke and Combettes (2011, Proposition 23.35).

• Theorem 5.2iv: By definition of a solution of a differential inclusion, $\dot{\boldsymbol{y}}(t)$ belongs to $-\boldsymbol{\Phi}(\boldsymbol{y}(t))$ for almost all $t \geq 0$. Therefore, if

$$\lambda(t) \coloneqq \frac{1}{2} \left\| \boldsymbol{y}(t) - \boldsymbol{y}^\star \right\|_2^2 , \tag{5.32}$$

then for almost all $t \geq 0$: $\dot{\lambda}(t) \in \boldsymbol{\Lambda}(t)$ with

$$\boldsymbol{\Lambda}(t) \coloneqq \left\{ \langle \boldsymbol{y}(t) - \boldsymbol{y}^\star \,,\, \dot{\boldsymbol{y}}(t) \rangle \,|\, \dot{\boldsymbol{y}}(t) \in -\boldsymbol{\Phi}(\boldsymbol{y}(t)) \right\} . \tag{5.33}$$

However, by hypothesis $\mathbf{0} \in -\boldsymbol{\Phi}(\boldsymbol{y}^\star)$, therefore, if $\boldsymbol{y}(t) \neq \boldsymbol{y}^\star$, then by the strict monotonicity of $\boldsymbol{\Phi}$

$$\langle \boldsymbol{y}(t) - \boldsymbol{y}^\star \,,\, \dot{\boldsymbol{y}}(t) \rangle < 0. \tag{5.34}$$

Consequently, every element of $\boldsymbol{\Lambda}$ is strictly negative, which implies Theorem 5.2iv.

• Theorem 5.2v: The proof follows the one of Cavazzuti et al. (2002, Theorem 5.1). If $\boldsymbol{\Phi}$ is uniformly monotone on $\mathbf{K}$ with modulus ψ, every element of $\boldsymbol{\Lambda}(t)$ can be upper-bounded as follows:

$$\langle \boldsymbol{y}(t) - \boldsymbol{y}^\star \,,\, \dot{\boldsymbol{y}}(t) \rangle \leq -\psi(\left\| \boldsymbol{y}(t) - \boldsymbol{y}^\star \right\|) < 0 \tag{5.35}$$

Thus $\lambda(t)$ is a decreasing function for almost all t, and $\lim_{t \to +\infty} \lambda(t) = l \geq 0$. By contradiction, we assume that $l > 0$, which implies by the definition of $\lambda(t)$ that $\left\| \boldsymbol{y}(t) - \boldsymbol{y}^\star \right\| > \sqrt{2l}$ for almost all $t \geq 0$. Since ψ is an increasing function, there exists a $m > 0$ such that $\psi(\left\| \boldsymbol{y}(t) - \boldsymbol{y}^\star \right\|) \geq m > 0$ for almost all $t \geq 0$. This implies with (5.35) that $\dot{\lambda}(t) \leq -m < 0$ and thus that $\lim_{t \to +\infty} \lambda(t) = -\infty$, which is absurd by the definition of $\lambda(t)$.

• Theorem 5.2vi: If $\boldsymbol{\Phi}$ is strongly monotone on $\mathbf{K}$ with constant μ, then for every $\dot{\boldsymbol{y}}(t) \in -\boldsymbol{\Phi}(\boldsymbol{y}(t))$

$$\langle \boldsymbol{y}(t) - \boldsymbol{y}^\star \,,\, \dot{\boldsymbol{y}}(t) \rangle = \langle \boldsymbol{y}(t) - \boldsymbol{y}^\star \,,\, \dot{\boldsymbol{y}}(t) - \mathbf{0} \rangle \leq -\mu \left\| \boldsymbol{y}(t) - \boldsymbol{y}^\star \right\|_2^2 = -2\mu\lambda(t) . \tag{5.36}$$

Therefore, for almost all t, $\dot{\lambda}(t) \leq -2\mu\lambda(t)$, which implies (5.31). □

Theorem 5.2 implies that if $\boldsymbol{\Xi}$ is uniformly monotone on $\mathbf{K}$, then the system described by (5.27) is asymptotically stable in $\mathbf{K}$ in the sense that for any initial point, the solution $\boldsymbol{y}(t)$ converges to an equilibrium point as $t \to +\infty$. Moreover, the unique equilibrium point $\boldsymbol{y}^\star$ of (5.27) is globally asymptotically stable, *i.e.* for every starting point $\boldsymbol{y}_0 \in \mathbf{K}$ the solution to (5.27) converges to $\boldsymbol{y}^\star$.

5.4 Methods and algorithms for solving strongly monotone GVI

As mentioned in the introduction, the *proximal point algorithm* is a fundamental algorithm to find an equilibrium point of (5.27) (and consequently a solution to GVI($\mathbf{K}, \boldsymbol{\Xi}$)) when $\boldsymbol{\Phi}$ is

maximal monotone (Martinet, 1970; Rockafellar, 1976). This algorithm relies on the *resolvent* operator, which, given a positive value $\lambda > 0$, is defined as $\mathbf{J}_\lambda^{\mathbf{\Phi}} := (\mathbf{I} + \lambda \mathbf{\Phi})^{-1}$ where $\mathbf{I}$ is the identity operator. Since $\mathbf{\Phi}$ is maximal monotone, the following properties hold (Ryu and Boyd, 2016): *(i)* $\mathbf{J}_\lambda^{\mathbf{\Phi}}$ is single-valued and is a non-expansive function, *i.e.* $\forall \boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbf{dom}\left(\mathbf{J}_\lambda^{\mathbf{\Phi}}\right), \boldsymbol{y}_1 \in \mathbf{J}_\lambda^{\mathbf{\Phi}}(\boldsymbol{u}_1), \boldsymbol{y}_2 \in \mathbf{J}_\lambda^{\mathbf{\Phi}}(\boldsymbol{u}_2) : \|\boldsymbol{u}_1 - \boldsymbol{u}_2\| \leq \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|$, *(ii)* $\mathbf{dom}\left(\mathbf{J}_\lambda^{\mathbf{\Phi}}\right) = \mathbb{R}^d$ (Minty, 1962, Corollary of Theorem 4), *(iii)* $\mathbf{0} \in \mathbf{\Phi}(\boldsymbol{y})$ if and only if $\boldsymbol{y} = \mathbf{J}_\lambda^{\mathbf{\Phi}}(\boldsymbol{y})$, or in other words, if and only if $\boldsymbol{y}$ is a fixed point of $\mathbf{J}_\lambda^{\mathbf{\Phi}}$. Given an initial point $\boldsymbol{y}_0 \in \mathbb{R}^d$ and the sequence of positive numbers $(\lambda_i)_{i \geq 0}$, the proximal point algorithm generates the sequence $(\boldsymbol{y}_i)_{i \geq 1} \in \mathbb{R}^d$ iteratively defined by $\boldsymbol{y}_i := \mathbf{J}_{\lambda_{i-1}}^{\mathbf{\Phi}}(\boldsymbol{y}_{i-1})$. If the resolvent can be computed exactly, then strong convergence is assured if λ_i is bounded away from zero and if $\mathbf{\Phi}$ is strongly monotone with constant $\mu > 0$ (Rockafellar, 1976). The drawback of this method is that the resolvent can be hard or impossible to compute (Eckstein and Bertsekas, 1992). Consequently, as previously mentioned, numerous of other techniques have then been proposed to solve VI$(\mathbf{K}, \mathbf{\Xi})$.

Among these different methods, we focus here on designing a *projected subgradient-type method* (for the sake of completeness) and an *extragradient method* to solve GVI$(\mathbf{K}, \mathbf{\Xi})$. For a variational equality problem VI$(\mathbf{K}, \mathbf{\Xi})$ for which the set-valued map $\mathbf{\Xi}$ is a Lipschitz continuous strongly monotone operator (and therefore single-valued), Nesterov and Scrimali (2011) show that a solution can be found using a standard gradient-type method. They then suggest a method based on dual extrapolation (Nesterov, 2007), which has better convergence properties while having a comparable computational cost.

As $\mathbf{\Xi}$ is a set-valued map for a generalised variational inequality, we cannot make the assumption that $\mathbf{\Xi}$ is Lipschitz continuous, which is central in the convergence analysis proposed by Nesterov and Scrimali (2011). However, to define a method able to handle a set-valued operator and for which we obtain similar convergence results as Nesterov and Scrimali (2011), we introduce the so-called *far-sighted Lipschitz continuity* property of a set-valued map, which we defined as follows:

Definition 5.14 (δ-far-sighted Lipschitz continuity). *Given a parameter $\delta > 0$, a set-valued map $\mathbf{\Psi} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is said to be* δ-far-sighted Lipschitz continuous *with constant $L_\delta^{\mathcal{Y}}$ on $\mathcal{Y} \subseteq \mathbf{dom}(\mathbf{\Psi})$ if and only if, for all $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathcal{Y}$ such that*

$$\boldsymbol{y}_2 \notin \mathcal{B}(\boldsymbol{y}_1 \,;\, \delta) := \{\boldsymbol{z} \,|\, \|\boldsymbol{y}_1 - \boldsymbol{z}\| < \delta\}\,, \tag{5.37}$$

we have

$$\forall \boldsymbol{u}_i \in \mathbf{\Psi}(\boldsymbol{y}_i)\,, i \in \{1, 2\} : \|\boldsymbol{u}_1 - \boldsymbol{u}_2\| \leq L_\delta \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|\,. \tag{5.38}$$

A set-valued map $\mathbf{\Psi} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is said to be δ-far-sighted Lipschitz continuous *with constant $L_{\boldsymbol{y}_1,\delta}^{\mathcal{Y}}$ on $\mathcal{Y} \subseteq \mathbf{dom}(\mathbf{\Psi})$ at a point $\boldsymbol{y}_1 \in \mathbf{dom}(\mathbf{\Psi})$ if and only if (5.38) holds $\forall \boldsymbol{y}_2 \in \mathcal{Y} \notin \mathcal{B}(\boldsymbol{y}_1 \,;\, \delta)$.*

The intuition behind Definition 5.14 is that the Lipschitz continuity condition (5.38) holds for every couple of points that are not in the same δ-neighbourhood. Graphically (see Figures 5.3 and 5.4), it means that there exists a double cone (in green in Figures 5.3 and 5.4), whose origin

can be moved along the graph so that every point of the graph always stays outside this double cone as long as it does not belong to the cylinder of diameter 2δ centred at the origin of the double cone. In accordance with intuition, the smaller δ, the larger the minimum value L_δ such that the δ-far-sighted Lipschitz continuity holds. Obviously, when δ tends to 0, we recover the Lipschitz continuity definition. Given a value $\delta > 0$, we denote by $\gamma_\delta := \frac{L_\delta}{\mu} \geq 1$ the *condition number* of $\boldsymbol{\Xi}$ related to δ. For readability purposes, we use in the following the notation L_δ as a shorthand for $L_\delta^{\mathbf{K}}$, *i.e.* the δ-far-sighted Lipschitz continuity constant on $\mathbf{K}$ of $\boldsymbol{\Xi}$. As shown later on, the *δ-far-sighted Lipschitz continuity* property of $\boldsymbol{\Xi}$ will allow us to prove the convergence towards a δ-neighbourhood of the solution $\boldsymbol{y}^\star$ of GVI$(\mathbf{K}, \boldsymbol{\Xi})$. The convergence rate towards this δ-neighbourhood will be dictated by the condition number γ_δ of $\boldsymbol{\Xi}$. For readability purposes, we use in the following the notation L_δ as a shorthand for $L_\delta^{\mathbf{K}}$, *i.e.* the δ-far-sighted Lipschitz continuity constant of $\boldsymbol{\Xi}$ on $\mathbf{K}$.

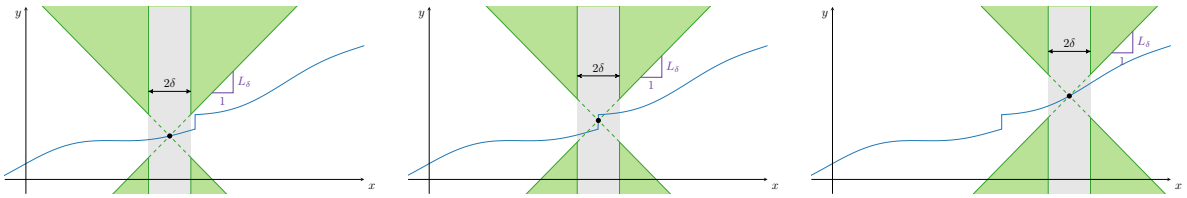


FIGURE 5.3

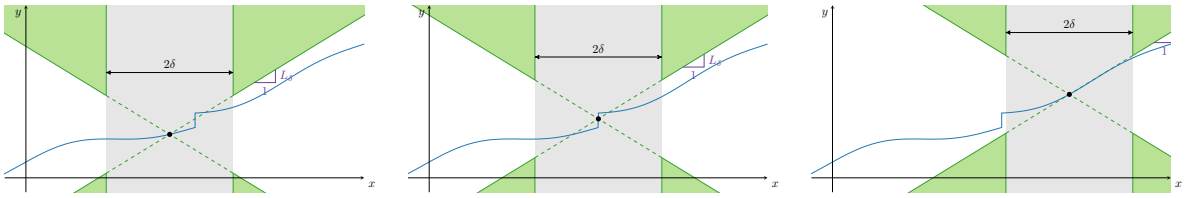


FIGURE 5.4

In the following, additionally to Assumption 5.1, we rely on the following assumption in order to design methods able to solve the generalised variational inequality GVI$(\mathbf{K}, \boldsymbol{\Xi})$ formulated in (5.1).

ASSUMPTION 5.2. *If $\boldsymbol{\Xi}$ is single-valued on $\mathcal{B}(\boldsymbol{y}\,;\,\epsilon) \cap \mathbf{K}$ for a given $\boldsymbol{y} \in \mathbf{K}$ and $\epsilon > 0$, then $\boldsymbol{\Xi}$ is Lipschitz continuous on $\mathcal{B}(\boldsymbol{y}\,;\,\epsilon) \cap \mathbf{K}$ with constant L. Moreover, for any $\delta > 0$, $\boldsymbol{\Xi}$ is δ-far-sighted Lipschitz continuous on $\mathbf{K}$ for any $\delta > 0$. We assume to be known the minimum value L_δ such that $\boldsymbol{\Xi}$ is δ-far-sighted Lipschitz continuous on $\mathbf{K}$, as well as the value $M_\delta^\star$ defined as follows*

$$M_\delta^\star := \max_{\boldsymbol{y} \in \mathcal{B}[\boldsymbol{y}^\star\,;\,\delta] \cap \mathbf{K} \setminus \{\boldsymbol{y}^\star\}} \|\boldsymbol{u} - \boldsymbol{u}^\star\| \;\; \textit{such that } \boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y})\,. \tag{5.39}$$

LEMMA 5.3. *If Assumptions 5.1 and 5.2 hold then*

$$\lim_{\delta \to 0} \frac{M_\delta^\star}{L_\delta^2} = 0. \tag{5.40}$$

PROOF. If there exists $\varepsilon > 0$ such that $\boldsymbol{\Xi}$ is single-valued on $\mathcal{B}(\boldsymbol{y}^\star\,;\,\epsilon) \cap \mathbf{K}$, then the upper-hemicontinuity of $\boldsymbol{\Xi}$ implies that $\lim_{\delta\to 0} M_\delta^\star = 0$ whereas $L_\delta \geq \mu$, which implies (5.40). We assume now that there does not exists such ε. If $\lim_{\delta\to 0} M_\delta^\star = 0$ then (5.40) holds with a similar argument as before. If $\lim_{\delta\to 0} M_\delta^\star > 0$ then $\exists c > 0$ such that $\forall \delta > 0$, $\exists \boldsymbol{y} \in \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta) \setminus \{\boldsymbol{y}^\star\}$ and $\boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y})$ such that $\|\boldsymbol{u} - \boldsymbol{u}^\star\| > c$. Therefore $\lim_{\delta\to 0} L_\delta = +\infty$. Moreover, with Lemma 5.2, there exist $\bar{M} > 0$ such that $\|\boldsymbol{u} - \boldsymbol{u}^\star\| \leq \bar{M}$ for all $\boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y})$ and $\boldsymbol{y} \in \mathbf{K}$. Hence $0 < M_\delta^\star \leq \bar{M}$, which together with $\lim_{\delta\to 0} L_\delta = +\infty$ implies (5.40). □

PROPOSITION 5.9. *If $\mathcal{B}[\boldsymbol{y}^\star\,;\,\delta] \subseteq \mathbf{K}$, then $M_\delta^\star$ is upper-bounded by δL_δ.*

PROOF. Let $\boldsymbol{y}$ be any element of $\mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$, $\boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y})$, and $\boldsymbol{d_u} := \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}$. Then there exists two points $\boldsymbol{y}_\delta^{(1)}$ and $\boldsymbol{y}_\delta^{(2)}$ on the boundary of $\mathcal{B}[\boldsymbol{y}^\star\,;\,\delta]$, such that

$$\boldsymbol{y}_\delta^{(2)} - \boldsymbol{y}_\delta^{(1)} = \kappa \boldsymbol{d_u} \text{ with } \kappa > 0, \quad \textit{and} \quad \boldsymbol{y} = \lambda \boldsymbol{y}_\delta^{(1)} + (1-\lambda)\, \boldsymbol{y}_\delta^{(2)} \text{ with } \lambda \in [0,1]. \tag{5.41}$$

Then, by the strong monotonicity of $\boldsymbol{\Xi}$ we have with (5.24) that for all $\boldsymbol{v}^{(2)} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(2)}\right)$:

$$\|\boldsymbol{u}\| = \langle \boldsymbol{u}\,,\, \boldsymbol{d_u} \rangle < \left\langle \boldsymbol{v}^{(2)}\,,\, \boldsymbol{d_u} \right\rangle \leq \left\| \boldsymbol{v}^{(2)} \right\|. \tag{5.42}$$

Then, since $\mathcal{B}[\boldsymbol{y}^\star\,;\,\delta] \subseteq \mathbf{K}$, we have that $\boldsymbol{u}^\star = \mathbf{0}$. This implies with the δ-far-sighted Lipschitz continuity that $\left\|\boldsymbol{v}^{(2)}\right\| = \left\|\boldsymbol{v}^{(2)} - \boldsymbol{u}^\star\right\| \leq L_\delta \left\|\boldsymbol{y}_\delta^{(2)} - \boldsymbol{y}^\star\right\| = \delta L_\delta$, which in turn implies with (5.42) that $\|\boldsymbol{u}\| < \delta L_\delta$. □

5.4.1 Projected subgradient-type (PS) method

We first study the rate of convergence towards a δ-neighbourhood of $\boldsymbol{y}^\star$ of the subgradient-type method suggested in Algorithm 5.1, where given a set $\mathcal{S}$, the notation $s \in_R \mathcal{S}$ is used to denote that s is chosen uniformly randomly from the set $\mathcal{S}$. Some proofs for such type of method can be found in the literature (see for instance Nesterov and Scrimali (2011) for VI and Chan and Pang (1982) for quasi-variational inequalities (QVI)), however the corresponding convergence rate depends on the Lipschitz continuity of $\boldsymbol{\Xi}$, which indirectly imposes that $\boldsymbol{\Xi}$ is single-valued. The convergence analysis presented here-in after is more general.

ALGORITHM 5.1 (PROJECTED SUBGRADIENT-TYPE (PS) METHOD).

Input: *Choose* $\boldsymbol{y}^{(0)} \in \mathbf{K}$ *and* $\alpha > 0$.

Iteration ($k \geq 0$)**:**

$$\boldsymbol{u}^{(k)} \in_R \boldsymbol{\Xi}\left(\boldsymbol{y}^{(k)}\right),$$
$$\boldsymbol{y}^{(k+1)} = \Pi_{\mathbf{K}}\left(\boldsymbol{y}^{(k)} - \alpha \boldsymbol{u}^{(k)}\right).$$

Output: $\boldsymbol{y}^{(k+1)}$.

THEOREM 5.3. *Suppose that Assumption 5.1 holds, and that, for a given $\delta > 0$, $\boldsymbol{\Xi}$ is μ-strongly monotone and δ-far-sighted Lipschitz continuous with constant L_δ on $\mathbf{K}$. Then, there exists a*

fixed step length $\alpha > 0$ such that the sequence $(\boldsymbol{y}_k)_{k\geq 0}$ generated by Algorithm 5.1 converges linearly to and then stays in a neighbourhood of size $\delta + \frac{\mu M_\delta^\star}{L_\delta^2}$ of $\boldsymbol{y}^\star$, i.e. $\mathcal{B}\left[\boldsymbol{y}^\star\,;\,\delta + \frac{\mu M_\delta^\star}{L_\delta^2}\right]$. Moreover, the step length ensuring the fastest convergence is given by $\alpha_\delta^* := \frac{\mu}{L_\delta^2}$.

PROOF. The equilibrium condition (5.28) implies that there exists a $\boldsymbol{u}^\star \in \boldsymbol{\Xi}(\boldsymbol{y}^\star)$ such that $\boldsymbol{y}^\star = \Pi_{\mathbf{K}}(\boldsymbol{y}^\star - \alpha \boldsymbol{u}^\star)$. First, we show that if the current iterate point $\boldsymbol{y}^{(k)}$ is outside the δ-neighbourhood of $\boldsymbol{y}^\star$ then we can find a step length that guarantees that we get closer to $\boldsymbol{y}^\star$: If $\boldsymbol{y}^{(k)} \notin \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$ then

$$\begin{aligned}
\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^\star\right\|^2 &= \left\|\Pi_{\mathbf{K}}\left(\boldsymbol{y}^{(k)} - \alpha \boldsymbol{u}^{(k)}\right) - \Pi_{\mathbf{K}}(\boldsymbol{y}^\star - \alpha \boldsymbol{u}^\star)\right\|^2, \\
&\overset{(5.8)}{\leq} \left\|\left(\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right) - \alpha\left(\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right)\right\|^2, \\
&= \left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\|^2 - 2\alpha\left\langle \boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\,,\,\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right\rangle + \alpha^2\left\|\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right\|^2, \qquad (5.43a) \\
&\overset{(5.14),(5.38)}{\leq} \left(1 - 2\alpha\mu + \alpha^2 L_\delta^2\right)\left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\|^2. \qquad (5.43b)
\end{aligned}$$

To ensure convergence, we must have that $(1 - 2\alpha\mu + \alpha^2 L_\delta^2) < 1 \Rightarrow \alpha < \frac{2\mu}{L_\delta^2}$. The step length $\alpha_\delta^* = \frac{\mu}{L_\delta^2}$, which minimises $(1 - 2\alpha\mu + \alpha^2 L_\delta^2)$, leads to the best convergence factor guarantee. Using the inequality

$$(1 - x) \leq \exp(-x) \qquad (5.44)$$

and the optimal step α_δ^*, we obtain with (5.43b) that

$$\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^\star\right\|^2 \leq \left\|\boldsymbol{y}^{(0)} - \boldsymbol{y}^\star\right\|^2 \exp\left(\frac{-k}{\gamma_\delta^2}\right), \qquad (5.45)$$

with γ_δ being the condition number of $\boldsymbol{\Xi}$ related to the δ. Now, if $\boldsymbol{y}^{(k)} \in \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$, then by hypothesis $\left\|\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right\| \leq M_\delta^\star$, which implies that

$$\begin{aligned}
\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^\star\right\| \leq \left\|\Pi_{\mathbf{K}}\left(\boldsymbol{y}^{(k)} - \alpha \boldsymbol{u}^{(k)}\right) - \Pi_{\mathbf{K}}(\boldsymbol{y}^\star - \alpha \boldsymbol{u}^\star)\right\| &\overset{(5.8)}{\leq} \left\|\left(\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right) - \alpha\left(\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right)\right\| \\
\leq \delta + \alpha\left\|\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right\| \leq \delta + \alpha M_\delta^\star. \qquad (5.46)&
\end{aligned}$$

Taking α equal to α_δ^* leads to the desired result. □

REMARK 5.1. *According to Theorem 5.3, if the optimal step length $\alpha_\delta^* = \frac{\mu}{L_\delta^2}$ is used, then the sequence $(\boldsymbol{y}_k)_{k\geq 0}$ stays in a neighbourhood of size $\delta + \frac{\mu M_\delta^\star}{L_\delta^2}$ of $\boldsymbol{y}^\star$ once it is reached. By definition, the smaller δ, the greater L_δ and the smaller $M_\delta^\star$, thus the smaller $\frac{\mu M_\delta^\star}{L_\delta^2}$, and consequently the closer $(\boldsymbol{y}_k)_{k\geq 0}$ stays from $\mathcal{B}[\boldsymbol{y}^\star\,;\,\delta]$ once it is reached. This may seem counter-intuitive as the smaller δ the larger the condition number γ_δ. However this makes sense as to guarantee convergence, the smaller δ, the smaller the step length required.*

REMARK 5.2. *Using the optimal step length α_δ^* suggested by Theorem 5.3 might be too conservative. Indeed, in the proof of Theorem 5.3, the constant L_δ could be replaced by the δ-far-sighted Lipschitz continuity constant $L_{\boldsymbol{y}^\star,\delta}$ at the solution point $\boldsymbol{y}^\star$ in order to bound (5.43a)*

with (5.43b). *Convergence would still be guaranteed if we were to use the step length* $\alpha^*_{\star,\delta} = \frac{\mu}{L^2_{\boldsymbol{y}^\star,\delta}}$, *which is greater or equal than* α^*_δ *since* $L_{\boldsymbol{y}^\star,\delta} \leq L_\delta$. *If* $L_{\boldsymbol{y}^\star,\delta} < L_\delta$, *then using the step length* $\alpha^*_{\star,\delta}$ *leads to a better convergence guarantee:*

$$\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^\star\right\|^2 \leq \left\|\boldsymbol{y}^{(0)} - \boldsymbol{y}^\star\right\|^2 \exp\left(\frac{-k}{\gamma^2_{\star,\delta}}\right), \tag{5.47}$$

with $\gamma_{\star,\delta} \coloneqq \frac{L_{\boldsymbol{y}^\star,\delta}}{\mu} \leq \frac{L_\delta}{\mu} = \gamma_\delta$.

REMARK 5.3. *The constant* $L_{\boldsymbol{y}^\star,\delta}$ *is not know* a priori *which mitigates the convergence benefit suggested in Remark 5.2. However, it is possible that, in certain situations, an upper-bound on* $L_{\boldsymbol{y}^\star,\delta}$ *could be derived and used in practice.*

COROLLARY 5.1. *If Assumptions 5.1 and 5.2 hold then*

$$\lim_{\alpha\to 0} \limsup_{k\to\infty} \left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\| = 0. \tag{5.48}$$

PROOF. The proof follows from Lemma 5.3 and Theorem 5.3. □

PROPOSITION 5.10. *If Assumptions 5.1 and 5.2 hold and if there exists* $\epsilon > 0$ *such that* $\boldsymbol{\Xi}$ *is Lipschitz continuous on* $\mathcal{B}(\boldsymbol{y}^\star\,;\,\epsilon)\cap\mathbf{K}$ *with constant* L, *then there exists a step length* $\alpha > 0$ *such that* $\lim_{k\to\infty}\left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\| = 0$.

PROOF. Given the assumptions, for any $\delta < \epsilon$, $\boldsymbol{\Xi}$ is δ-far-sighted Lipschitz continuous with constant $L^\star \coloneqq \max(L, L_\epsilon)$, where L_ϵ is the ϵ-far-sighted Lipschitz constant of $\boldsymbol{\Xi}$ on $\mathbf{K}$. Since by assumption $\boldsymbol{\Xi}$ is Lipschitz continuous and upper hemicontinuous on $\mathcal{B}(\boldsymbol{y}^\star\,;\,\epsilon)\cap\mathbf{K}$, it is single-valued on $\mathcal{B}(\boldsymbol{y}^\star\,;\,\epsilon)\cap\mathbf{K}$ and $M^\star_\delta \to 0$ as $\delta \to 0$. Therefore the results holds by taking $\alpha = \frac{2\mu}{L^\star}$ and using Theorem 5.3 with $\delta \to 0$. □

5.4.2 Original dual extrapolation (ODE) method

As proved in Theorem 5.3, it is known that the convergence of Algorithm 5.1 might be very slow when the condition number γ_δ is large. In their paper Nesterov and Scrimali (2011) provide an alternative method with better convergence while having a similar computational cost that can be seen as the combination of the dual extrapolation method (Nesterov, 2007) with the technique of estimate functions (Nesterov, 2018). We will show here that this method can be adapted to solve GVIs where $\boldsymbol{\Xi}$ is not necessarily single-valued.

When using the subgradient-type method to minimise a convex function f, if $\left(\boldsymbol{y}^{(k)}\right)_{0\leq k\leq \bar{k}}$ denotes the sequence of points generated after $\bar{k}$ iterations, then the best estimate of the minimizer $\boldsymbol{y}^{best}$ is simply obtained by taking the element of the sequence that achieves the lowest value function, *i.e.*

$$f(\boldsymbol{y}_{best}) = \min\left\{f\left(\boldsymbol{y}^{(1)}\right), \ldots, f\left(\boldsymbol{y}^{(\bar{k})}\right)\right\}. \tag{5.49}$$

In the case of finding a solution to (5.1), we do not have a consensual function f that allows to evaluate the *quality* of a point in the sequence $\left(\boldsymbol{y}^{(k)}\right)_{k\geq 0}$ generated by Algorithm 5.1. We

can however certify with Theorem 5.3 that, provided that the step length α is small enough, the sequence $\left(\boldsymbol{y}^{(k)}\right)_{k\geq 0}$ converges to and stays in a neighbourhood of $\boldsymbol{y}^\star$. Some papers measure the accuracy of the solution in terms of *weak gap*, *i.e.* $\mathrm{gap}(\boldsymbol{y}) \coloneqq \max_{\boldsymbol{z}\in\mathbf{K},\,\boldsymbol{u}\in\boldsymbol{\Xi}(\boldsymbol{z})} \langle \boldsymbol{u}\,,\,\boldsymbol{y}-\boldsymbol{z}\rangle$ (*e.g.* Kotsalis et al. (2020)). In this work, following the ideas of Fukushima (1992), we define the following *merit function* in order to estimate the quality of approximate solutions to (5.1):

$$f_{\mathbf{K}}(\boldsymbol{y}) \coloneqq \max_{\boldsymbol{z}\in\mathbf{K}} \left\{ \max_{\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})} \{\langle \boldsymbol{v}\,,\,\boldsymbol{y}-\boldsymbol{z}\rangle\} + \frac{1}{2}\mu\left\|\boldsymbol{y}-\boldsymbol{z}\right\|^2 \right\}, \tag{5.50}$$

where μ is the strong monotonicity parameter.

THEOREM 5.4. *The merit function defined in (5.50) is well defined, is strongly convex on* $\mathbf{K}$ *with parameter* μ*, is non-negative on* $\mathbf{K}$*, and vanishes only at the unique solution* $\boldsymbol{y}^\star$ *of (5.1).*

PROOF. The proof follows the one of Nesterov and Scrimali (2011, Theorem 2.1). Given any two points $\boldsymbol{y}\in\mathbf{K}$ and $\boldsymbol{z}\in\mathbf{K}$, the following inequalities holds for any $\boldsymbol{u}\in\boldsymbol{\Xi}(\boldsymbol{y})$ and $\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})$ based on the strong convexity of $\boldsymbol{\Xi}$:

$$\langle \boldsymbol{v}\,,\,\boldsymbol{y}-\boldsymbol{z}\rangle + \frac{1}{2}\mu\left\|\boldsymbol{y}-\boldsymbol{z}\right\|^2 \overset{(5.14)}{\leq} \langle \boldsymbol{u}\,,\,\boldsymbol{y}-\boldsymbol{z}\rangle - \frac{1}{2}\mu\left\|\boldsymbol{y}-\boldsymbol{z}\right\|^2 \tag{5.51a}$$

$$\leq \left\|\boldsymbol{u}\right\|\left\|\boldsymbol{y}-\boldsymbol{z}\right\| - \frac{1}{2}\mu\left\|\boldsymbol{y}-\boldsymbol{z}\right\|^2 \leq \frac{1}{2\mu}\left\|\boldsymbol{u}\right\|^2, \tag{5.51b}$$

where the last bound is obtained by finding the maximum value of the parabola $g:\mathbb{R}\to\mathbb{R}; x\mapsto \|\boldsymbol{u}\|\,x - \frac{1}{2}\mu x^2$. With the fact that $\mathbf{K}$ and $\boldsymbol{\Xi}(\boldsymbol{z})$ for $\boldsymbol{z}\in\mathbf{K}$ are closed sets, the value $f_{\mathbf{K}}(\boldsymbol{y})$ is well defined. This merit function $f_{\mathbf{K}}$ is strongly convex in $\boldsymbol{y}$ on $\mathbf{K}$ with parameter μ, as a maximum of parametric family of strongly convex functions, which themselves can be expressed as a maximum of parametric family of strongly convex quadratic functions:

$$f_{\mathbf{K}}(\boldsymbol{y}) = \max_{\boldsymbol{z}\in\mathbf{K}} \phi_{\boldsymbol{z}}(\boldsymbol{y}), \tag{5.52a}$$

$$\phi_{\boldsymbol{z}}(\boldsymbol{y}) = \max_{\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})} \phi_{\boldsymbol{z},\boldsymbol{v}}(\boldsymbol{y}), \tag{5.52b}$$

$$\phi_{\boldsymbol{z},\boldsymbol{v}}(\boldsymbol{y}) = \langle \boldsymbol{v}\,,\,\boldsymbol{y}-\boldsymbol{z}\rangle + \frac{1}{2}\mu\left\|\boldsymbol{y}-\boldsymbol{z}\right\|^2. \tag{5.52c}$$

For any $\boldsymbol{y}\in\mathbf{K}$, $f_{\mathbf{K}}(\boldsymbol{y})$ is lower-bounded by 0. Indeed, taking $\boldsymbol{z}=\boldsymbol{y}$ in (5.50) gives the desired result. If $\boldsymbol{y}^\star$ is the solution to GVI$(\mathbf{K},\boldsymbol{\Xi})$, then in view of (5.51a) and (5.1) we have for all $\boldsymbol{u}^\star\in\boldsymbol{\Xi}(\boldsymbol{y}^\star)$, any $\boldsymbol{z}\in\mathbf{K}$ and $\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})$ that:

$$\langle \boldsymbol{v}\,,\,\boldsymbol{y}^\star-\boldsymbol{z}\rangle + \frac{1}{2}\mu\left\|\boldsymbol{y}^\star-\boldsymbol{z}\right\|^2 \overset{(5.51a)}{\leq} \langle \boldsymbol{u}^\star\,,\,\boldsymbol{y}^\star-\boldsymbol{z}\rangle - \frac{1}{2}\mu\left\|\boldsymbol{y}^\star-\boldsymbol{z}\right\|^2 \overset{(5.1)}{\leq} 0. \tag{5.53}$$

Hence, with the non-negativity of $f_{\mathbf{K}}$ and (5.53), we have that $f_{\mathbf{K}}(\boldsymbol{y}^\star)=0$. Suppose now that there exists a $\boldsymbol{y}^\dagger$ such that $f_{\mathbf{K}}\left(\boldsymbol{y}^\dagger\right)=0$ and $\boldsymbol{y}^\dagger\neq\boldsymbol{y}^\star$. Since $\boldsymbol{y}^\dagger$ is such that $f_{\mathbf{K}}\left(\boldsymbol{y}^\dagger\right)=0$, then with (5.52) we have that $\phi_{\boldsymbol{z},\boldsymbol{v}}\left(\boldsymbol{y}^\dagger\right)\leq 0$ for all $\boldsymbol{z}\in\mathbf{K}$ and $\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})$, which implies that $\boldsymbol{y}^\dagger$ is a solution to the Minty variational inequality (Minty, 1967), *i.e.*

$$\forall \boldsymbol{z}\in\mathbf{K}, \boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z}) : \left\langle \boldsymbol{v}\,,\,\boldsymbol{z}-\boldsymbol{y}^\dagger\right\rangle \geq 0. \tag{5.54}$$

In particular, for $\boldsymbol{z} = \boldsymbol{y}^\star$ and $\boldsymbol{v} = \boldsymbol{u}^\star$ the second term of $\phi_{\boldsymbol{y}^\star,\boldsymbol{u}^\star}\left(\boldsymbol{y}^\dagger\right)$ is positive since $\boldsymbol{y}^\dagger \neq \boldsymbol{y}^\star$, which implies that $\left\langle \boldsymbol{u}^\star\,,\,\boldsymbol{y}^\star - \boldsymbol{y}^\dagger \right\rangle > 0$. By definition, since $\boldsymbol{y}^\star$ is the unique solution to GVI$(\mathbf{K},\boldsymbol{\Xi})$, there exists a $\boldsymbol{u}^\star \in \boldsymbol{\Xi}(\boldsymbol{y}^\star)$ such that

$$\forall \boldsymbol{z} \in \mathbf{K} : \langle \boldsymbol{u}^\star\,,\,\boldsymbol{z} - \boldsymbol{y}^\star \rangle \geq 0. \tag{5.55}$$

However, taking $\boldsymbol{z} = \boldsymbol{y}^\dagger$ in (5.55) leads to $\left\langle \boldsymbol{u}^\star\,,\,\boldsymbol{y}^\dagger - \boldsymbol{y}^\star \right\rangle \geq 0$ and contradicts the fact that $\left\langle \boldsymbol{u}^\star\,,\,\boldsymbol{y}^\star - \boldsymbol{y}^\dagger \right\rangle > 0$. □

In the following, we will show that the accelerated method suggested by Nesterov and Scrimali (2011) can be adapted in order to generate a sequence of points with iteratively smaller value for the merit function $f_{\mathbf{K}}$ defined in (5.50). Given a sequence of step lengths $\left(\alpha^{(i)}\right)_{i=0}^{k}$, the method relies on the following quantities:

$$S^{(k)} := \sum_{i=0}^{k} \alpha^{(i)}, \tag{5.56a}$$

$$\bar{\boldsymbol{y}}^{(k)} := \frac{1}{S^{(k)}} \sum_{i=0}^{k} \alpha^{(i)} \boldsymbol{y}^{(i)}, \tag{5.56b}$$

$$\widetilde{\Delta}_{\mathbf{K}}^{(k)} := \max_{\boldsymbol{z}\in\mathbf{K}} \left\{ \sum_{i=0}^{k} \alpha^{(i)} \left(\min_{\boldsymbol{u}^{(i)} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(i)}\right)} \left\{ \left\langle \boldsymbol{u}^{(i)}\,,\,\boldsymbol{y}^{(i)} - \boldsymbol{z} \right\rangle \right\} - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(i)} \right\|^2 \right) \right\}. \tag{5.56c}$$

Following the same proof technique as Nesterov and Scrimali (2011, Lemma 2.2), we can show the following result.

LEMMA 5.4. *It holds that*

$$f_{\mathbf{K}}\left(\bar{\boldsymbol{y}}^{(k)}\right) \leq \frac{1}{S^{(k)}} \widetilde{\Delta}_{\mathbf{K}}^{(k)}. \tag{5.57}$$

PROOF. First, we note that the following inequalities hold:

$$\left\| \boldsymbol{x} - \bar{\boldsymbol{y}}^{(k)} \right\|^2 \leq \frac{1}{S^{(k)}} \sum_{i=0}^{k} \alpha^{(i)} \left\| \boldsymbol{x} - \boldsymbol{y}^{(i)} \right\|^2, \tag{5.58}$$

and $\forall \boldsymbol{v} \in \boldsymbol{\Xi}(\boldsymbol{z})\,, \boldsymbol{u}^{(i)} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(i)}\right)$:

$$\left\langle \boldsymbol{v}\,,\,\boldsymbol{y}^{(i)} - \boldsymbol{z} \right\rangle + \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(i)} \right\|^2 \overset{(5.14)}{\leq} \left\langle \boldsymbol{u}^{(i)}\,,\,\boldsymbol{y}^{(i)} - \boldsymbol{z} \right\rangle - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(i)} \right\|^2. \tag{5.59}$$

By definition of $f_{\mathbf{K}}\left(\bar{\boldsymbol{y}}^{(k)}\right)$ and using the two previous inequalities we obtain:

$$f_{\mathbf{K}}\left(\bar{\boldsymbol{y}}^{(k)}\right) = \max_{\boldsymbol{z}\in\mathbf{K}} \left\{ \max_{\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})} \left\{ \left\langle \boldsymbol{v}\,,\,\bar{\boldsymbol{y}}^{(k)} - \boldsymbol{z} \right\rangle \right\} + \frac{1}{2}\mu \left\| \boldsymbol{z} - \bar{\boldsymbol{y}}^{(k)} \right\|^2 \right\} \tag{5.60a}$$

$$\overset{(5.56b)}{=} \frac{1}{S^{(k)}} \max_{\boldsymbol{z}\in\mathbf{K}} \left\{ \max_{\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})} \left\{ \sum_{i=0}^{k} \alpha^{(i)} \left\langle \boldsymbol{v}\,,\,\boldsymbol{y}^{(i)} - \boldsymbol{z} \right\rangle + \frac{1}{2}\mu S^{(k)} \left\| \boldsymbol{z} - \bar{\boldsymbol{y}}^{(k)} \right\|^2 \right\} \right\} \tag{5.60b}$$

$$\overset{(5.58)}{\leq} \frac{1}{S^{(k)}} \max_{\boldsymbol{z}\in\mathbf{K}} \left\{ \max_{\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})} \left\{ \sum_{i=0}^{k} \alpha^{(i)} \left(\left\langle \boldsymbol{v}\,,\,\boldsymbol{y}^{(i)} - \boldsymbol{z} \right\rangle + \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(i)} \right\|^2 \right) \right\} \right\} \tag{5.60c}$$

$$\leq \frac{1}{S^{(k)}} \max_{\boldsymbol{z}\in\mathbf{K}} \left\{ \sum_{i=0}^{k} \alpha^{(i)} \left(\max_{\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})} \left\{ \left\langle \boldsymbol{v}\,,\,\boldsymbol{y}^{(i)} - \boldsymbol{z} \right\rangle \right\} + \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(i)} \right\|^2 \right) \right\} \tag{5.60d}$$

$$\overset{(5.59)}{\leq} \frac{1}{S^{(k)}} \max_{\boldsymbol{z}\in\mathbf{K}} \left\{ \sum_{i=0}^{k} \alpha^{(i)} \left(\min_{\boldsymbol{u}^{(i)}\in\boldsymbol{\Xi}\left(\boldsymbol{y}^{(i)}\right)} \left\{ \left\langle \boldsymbol{u}^{(i)} \,,\, \boldsymbol{y}^{(i)} - \boldsymbol{z} \right\rangle \right\} - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(i)} \right\|^2 \right) \right\} \tag{5.60e}$$

$$= \frac{1}{S^{(k)}} \widetilde{\Delta}_{\mathbf{K}}^{(k)}. \tag{5.60f}$$

□

Hence, to generate a sequence $\left(\boldsymbol{y}^{(k)}\right)_{k\geq 0}$ with decreasing values of the merit function $f_{\mathbf{K}}$, we have to control the growth of $\widetilde{\Delta}_{\mathbf{K}}^{(k)}$ compared to that of $S^{(k)}$. To this end, given $\beta > 0$, we define

$$\widetilde{\psi}_{\boldsymbol{y}}^{\beta}(\boldsymbol{z}) \coloneqq \min_{\boldsymbol{u}\in\boldsymbol{\Xi}(\boldsymbol{y})} \{\langle \boldsymbol{u} \,,\, \boldsymbol{y} - \boldsymbol{z} \rangle\} - \frac{1}{2}\beta \left\| \boldsymbol{z} - \boldsymbol{y} \right\|^2, \tag{5.61a}$$

$$\widetilde{\Psi}_k(\boldsymbol{z}) \coloneqq \sum_{i=0}^{k} \alpha^{(i)} \widetilde{\psi}_{\boldsymbol{y}^{(i)}}^{\mu}(\boldsymbol{z}). \tag{5.61b}$$

From the definition of $\widetilde{\Delta}_{\mathbf{K}}^{(k)}$, we observe that $\widetilde{\Delta}_{\mathbf{K}}^{(k)} = \max_{\boldsymbol{z}\in\mathbf{K}} \widetilde{\Psi}_k(\boldsymbol{z})$, which leads us to define the following iterations given $\left(\boldsymbol{y}^{(i)}\right)_{i=0}^{k}$:

$$\boldsymbol{x}^{(k)} \coloneqq \arg\max_{\boldsymbol{z}\in\mathbf{K}} \widetilde{\Psi}_k(\boldsymbol{z}), \tag{5.62a}$$

$$\boldsymbol{y}^{(k+1)} \coloneqq \arg\max_{\boldsymbol{z}\in\mathbf{K}} \widetilde{\psi}_{\boldsymbol{x}^{(k)}}^{\beta}(\boldsymbol{z}). \tag{5.62b}$$

Iterations (5.62a) and (5.62b) are easily computable after noticing that both functions $\widetilde{\psi}_{\boldsymbol{x}^{(k)}}^{\beta}$ and $\widetilde{\Psi}_k$ are strongly concave piecewise quadratic functions with parameters β and $\mu S^{(k)}$, respectively. Given a concave function $f_{\mathbf{K}}$, let $\bar{\boldsymbol{\partial}} f_{\mathbf{K}}(\boldsymbol{z})$ denote the *superdifferential* of $f_{\mathbf{K}}$ at the point $\boldsymbol{z}$, *i.e.* $\bar{\boldsymbol{\partial}} f_{\mathbf{K}}(\boldsymbol{z}) = -\boldsymbol{\partial}(-f_{\mathbf{K}}(\boldsymbol{z}))$. If $f_{\mathbf{K}}$ is strongly concave on $\mathbf{K}$ with parameter μ, then for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbf{K}$ and $\boldsymbol{v} \in \bar{\boldsymbol{\partial}} f_{\mathbf{K}}(\boldsymbol{x})$, the following inequality holds:

$$f_{\mathbf{K}}(\boldsymbol{z}) \leq f_{\mathbf{K}}(\boldsymbol{x}) + \langle \boldsymbol{v} \,,\, \boldsymbol{z} - \boldsymbol{x} \rangle - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{x} \right\|^2. \tag{5.63}$$

Since by assumption $\mathbf{K}$ is a compact convex set, a strongly concave function $f_{\mathbf{K}}$ on $\mathbf{K}$ with parameter μ has a unique maximum $\boldsymbol{x}^\star$ and there exists a $\boldsymbol{v}^\star \in \bar{\boldsymbol{\partial}} f_{\mathbf{K}}(\boldsymbol{x})$ such that $\langle \boldsymbol{v}^\star \,,\, \boldsymbol{z} - \boldsymbol{x}^\star \rangle \leq 0$ for all $\boldsymbol{z} \in \mathbf{K}$. Together with (5.63), this implies that the following inequality holds:

$$f_{\mathbf{K}}(\boldsymbol{z}) \leq f_{\mathbf{K}}(\boldsymbol{x}^\star) + \langle \boldsymbol{v}^\star \,,\, \boldsymbol{z} - \boldsymbol{x}^\star \rangle - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{x}^\star \right\|^2 \leq f_{\mathbf{K}}(\boldsymbol{x}^\star) - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{x}^\star \right\|^2. \tag{5.64}$$

Since by assumption $\boldsymbol{\Xi}(\boldsymbol{y})$ is a compact set for all $\boldsymbol{y} \in \mathbf{K}$, the superdifferential $\bar{\boldsymbol{\partial}}\widetilde{\psi}_{\boldsymbol{y}}^{\beta}(\bar{\boldsymbol{z}})$ of $\widetilde{\psi}_{\boldsymbol{y}}^{\beta}$ in $\boldsymbol{z}$ at a point $\bar{\boldsymbol{z}}$ satisfies the following inclusion:

$$\begin{aligned} \bar{\boldsymbol{\partial}}\widetilde{\psi}_{\boldsymbol{y}}^{\beta}(\bar{\boldsymbol{z}}) &= \mathbf{conv}\left\{ \bigcup \left\{ -\boldsymbol{u} - \beta(\bar{\boldsymbol{z}} - \boldsymbol{y}) \,\middle|\, \boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y}) \text{ and } \langle \boldsymbol{u} \,,\, \boldsymbol{y} - \bar{\boldsymbol{z}} \rangle - \frac{1}{2}\beta \left\| \bar{\boldsymbol{z}} - \boldsymbol{y} \right\|^2 = \widetilde{\psi}_{\boldsymbol{y}}^{\beta}(\bar{\boldsymbol{z}}) \right\} \right\} \\ &\subseteq \left\{ -\boldsymbol{u} - \beta(\bar{\boldsymbol{z}} - \boldsymbol{y}) \,\middle|\, \boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y}) \right\}. \end{aligned} \tag{5.65}$$

This gives us the tools to prove the following result.

THEOREM 5.5. *If* $\alpha^{(k+1)} \leq \frac{\mu}{\beta} S^{(k)}$ *then* $\exists \boldsymbol{v}^{(k)} \in \boldsymbol{\Xi}\left(\boldsymbol{x}^{(k)}\right)$ *such that for all* $\boldsymbol{u} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(k+1)}\right)$*:*

$$\widetilde{\Delta}_{\mathbf{K}}^{(k+1)} \leq \widetilde{\Delta}_{\mathbf{K}}^{(k)} + \frac{1}{2}\alpha^{(k+1)} \left[\frac{1}{\mu + \beta} \left\| \boldsymbol{u} - \boldsymbol{v}^{(k)} \right\|^2 - \beta \left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|^2 \right]. \tag{5.66}$$

PROOF. The proof follows the one of Nesterov and Scrimali (2011, Theorem 2.3). First, we note that $\widetilde{\Psi}_k\left(\boldsymbol{x}^{(k)}\right) = \widetilde{\Delta}_{\mathbf{K}}^{(k)}$. With this in mind and after noticing that $\widetilde{\Psi}_{k+1}(\boldsymbol{z}) = \widetilde{\Psi}_k(\boldsymbol{z}) + \alpha^{(k+1)}\widetilde{\psi}^{\mu}_{\boldsymbol{y}^{(k+1)}}(\boldsymbol{z})$, we can bound $\widetilde{\Delta}_{\mathbf{K}}^{(k+1)}$ as follows:

$$\begin{aligned}
\widetilde{\Delta}_{\mathbf{K}}^{(k+1)} &= \max_{\boldsymbol{z}\in\mathbf{K}}\left\{\widetilde{\Psi}_k(\boldsymbol{z}) + \alpha^{(k+1)}\widetilde{\psi}^{\mu}_{\boldsymbol{y}^{(k+1)}}(\boldsymbol{z})\right\} \\
&\overset{(5.64)}{\leq} \max_{\boldsymbol{z}\in\mathbf{K}}\left\{\widetilde{\Delta}_{\mathbf{K}}^{(k)} - \frac{1}{2}\mu S^{(k)}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2 + \alpha^{(k+1)}\widetilde{\psi}^{\mu}_{\boldsymbol{y}^{(k+1)}}(\boldsymbol{z})\right\} \\
&\leq \widetilde{\Delta}_{\mathbf{K}}^{(k)} + \max_{\boldsymbol{z}\in\mathbf{K}}\Bigg\{ - \frac{1}{2}\mu S^{(k)}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2 \\
&\qquad + \alpha^{(k+1)}\left(\min_{\boldsymbol{u}\in\boldsymbol{\Xi}\left(\boldsymbol{y}^{(k+1)}\right)}\left\{\left\langle \boldsymbol{u}\,,\,\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\rangle\right\} - \frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2\right)\Bigg\}. \qquad (5.67)
\end{aligned}$$

Then, as $\boldsymbol{y}^{(k+1)}$ maximises $\widetilde{\psi}^{\beta}_{\boldsymbol{x}^{(k)}}$, there exists a $\boldsymbol{w}^{(k)} \in \bar{\boldsymbol{\partial}}\widetilde{\psi}^{\beta}_{\boldsymbol{x}^{(k)}}\left(\boldsymbol{y}^{(k+1)}\right)$ such that:

$$\left\langle \boldsymbol{w}^{(k)}\,,\,\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\rangle \leq 0, \quad \forall \boldsymbol{z}\in\mathbf{K}\,. \qquad (5.68)$$

This implies with (5.65) that there exists a $\boldsymbol{v}^{(k)} \in \boldsymbol{\Xi}\left(\boldsymbol{x}^{(k)}\right)$ such that:

$$\left\langle -\boldsymbol{v}^{(k)} - \beta\left(\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right)\,,\,\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\rangle \leq 0, \quad \forall \boldsymbol{z}\in\mathbf{K}\,. \qquad (5.69)$$

Using (5.69) and the following equality

$$2\left\langle \boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\,,\,\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\rangle = \left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2 - \left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2 - \left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2, \qquad (5.70)$$

we obtain for any $\boldsymbol{u} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(k+1)}\right)$ that:

$$\begin{aligned}
&\left\langle \boldsymbol{u}\,,\,\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2 && (5.71a)\\
&= \left\langle \boldsymbol{u}-\boldsymbol{v}^{(k)}\,,\,\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2 + \left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\rangle && (5.71b)\\
&\overset{(5.69)}{\leq} \left\|\boldsymbol{u}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\| - \frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2 + \beta\left\langle \boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\,,\,\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\rangle && (5.71c)\\
&\overset{(5.70)}{=} \left\|\boldsymbol{u}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\| - \frac{1}{2}(\mu+\beta)\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2 + \frac{\beta}{2}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2 - \frac{\beta}{2}\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2. && (5.71d)
\end{aligned}$$

The first two terms of (5.71d) can be seen as the polynomial $c_1 r - \frac{1}{2}c_2 r^2$ with $c_1 = \left\|\boldsymbol{u}-\boldsymbol{v}^{(k)}\right\|$, $c_2 = \mu+\beta$, and $r = \left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|$. This polynomial reaches its maximum at $r^\star = c_1/c_2$ and can be upper-bounded by $(c_1^2/2c_2)$. Hence

$$\begin{aligned}
&\left\langle \boldsymbol{u}\,,\,\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2 \\
&\qquad\qquad \leq \frac{1}{2(\mu+\beta)}\left\|\boldsymbol{u}-\boldsymbol{v}^{(k)}\right\|^2 + \frac{\beta}{2}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2 - \frac{\beta}{2}\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2. \qquad (5.72)
\end{aligned}$$

By combining (5.67) with (5.72) and using the fact that $\alpha^{(k+1)} \leq \frac{\mu}{\beta}S^{(k)}$, we obtain (5.66). □

COROLLARY 5.2. *Assume that $\boldsymbol{\Xi}$ is μ-strongly monotone and δ-far-sighted Lipschitz continuous with constant L_δ on $\mathbf{K}$ for a given $\delta > 0$. With the choice*

$$\beta = L_\delta, \quad \alpha^{(k+1)} = \frac{1}{\gamma_\delta} S^{(k)}, \tag{5.73}$$

where $\gamma_\delta = \frac{L_\delta}{\mu}$ is the condition number of $\boldsymbol{\Xi}$ related to δ, we obtain that $\widetilde{\Delta}_{\mathbf{K}}^{(k+1)} \leq \widetilde{\Delta}_{\mathbf{K}}^{(k)}$ while $\left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\| \geq \delta$.

PROOF. With Theorem 5.5, Inequality (5.66) holds true provided that $\alpha^{(k+1)} \leq \frac{\mu}{\beta} S^{(k)}$, which is the case given the parameter values provided in (5.73). Plugging in the values for $\alpha^{(k+1)}$ and β and using the definition of δ-far-sighted Lipschitz continuity in (5.66) gives the following upper-bound:

$$\widetilde{\Delta}_{\mathbf{K}}^{(k+1)} \leq \widetilde{\Delta}_{\mathbf{K}}^{(k)} + \frac{\mu S^{(k)}}{2L_\delta} \left(\frac{L_\delta^2}{\mu + L_\delta} - L_\delta \right) \left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|^2 \tag{5.74a}$$

$$= \widetilde{\Delta}_{\mathbf{K}}^{(k)} - \frac{\mu^2 S^{(k)}}{2(\mu + L_\delta)} \left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|^2 \tag{5.74b}$$

$$\leq \widetilde{\Delta}_{\mathbf{K}}^{(k)}. \tag{5.74c}$$

□

Corollary 5.2 leads us to the method stated in Algorithm 5.2 that, given a tolerance value $\delta > 0$, generates a sequence of points $\left(\bar{\boldsymbol{y}}^{(k)} \right)_{k \geq 0}$ converging to the generalised variational solution $\boldsymbol{y}^\star$ as long as the points $\boldsymbol{y}^{(k+1)}$ and $\boldsymbol{x}^{(k)}$ do not belong to the same δ-neighbourhood (see subsequent Theorem 5.6).

ALGORITHM 5.2 (ORIGINAL DUAL EXTRAPOLATION (ODE) METHOD).

Input: *Choose $\boldsymbol{y}^{(0)} \in \mathbf{K}$ and $\beta > 0$.*

Initialisation: *Set $\alpha^{(0)} = 1$, and $\gamma = \frac{\beta}{\mu}$.*

Iteration ($k \geq 0$):

$$S^{(k)} = \sum_{i=0}^{k} \alpha^{(i)} \text{ and } \alpha^{(k+1)} = \tfrac{1}{\gamma} S^{(k)},$$
$$\boldsymbol{x}^{(k)} = \arg\max_{\boldsymbol{z} \in \mathbf{K}} \widetilde{\Psi}_k(\boldsymbol{z}),$$
$$\boldsymbol{y}^{(k+1)} = \arg\max_{\boldsymbol{z} \in \mathbf{K}} \widetilde{\psi}^{\beta}_{\boldsymbol{x}^{(k)}}(\boldsymbol{z}).$$

Output: $\bar{\boldsymbol{y}}^{(k)} = \frac{1}{S^{(k)}} \sum_{i=0}^{k} \alpha^{(i)} \boldsymbol{y}^{(i)}$.

THEOREM 5.6. *Let $\boldsymbol{\Xi}$ be a μ-strongly monotone and δ-far-sighted Lipschitz continuous with parameter L_δ on $\mathbf{K}$. Let $\left\{ \bar{\boldsymbol{y}}^{(k)} \right\}_{k \in [\![0, \bar{k}]\!]}$ be the sequence generated by Algorithm 5.2 with $\beta = L_\delta$, where $\bar{k}$ is the first iteration number such that $\left\| \boldsymbol{y}^{(\bar{k}+1)} - \boldsymbol{x}^{(\bar{k})} \right\| < \delta$ holds true. If $\left\| \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\| > \delta$, then for any $k \in [\![0, \bar{k}]\!]$:*

$$\frac{1}{2}\mu \left\| \bar{\boldsymbol{y}}^{(k)} - \boldsymbol{y}^\star \right\|^2 \leq f_{\mathbf{K}}\left(\bar{\boldsymbol{y}}^{(k)} \right) \leq \left[f_{\mathbf{K}}\left(\boldsymbol{y}^{(0)} \right) + \frac{\mu(\gamma_\delta^2 - 1)}{2} \left\| \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\|^2 \right] \exp\left(-\frac{k}{\gamma_\delta + 1} \right)$$
$$\leq f_{\mathbf{K}}\left(\boldsymbol{y}^{(0)} \right) \cdot \gamma_\delta^2 \cdot \exp\left(-\frac{k}{\gamma_\delta + 1} \right). \tag{5.75}$$

PROOF. The proof follows from Nesterov and Scrimali (2011, Theorem 2.4) and Corollary 5.2 and is given here for completeness. We denote by $(\boldsymbol{y}^\star, \boldsymbol{u}^\star)$ with $\boldsymbol{u}^\star \in \boldsymbol{\Xi}(\boldsymbol{y}^\star)$ the unique solution of GVI$(\mathbf{K}, \boldsymbol{\Xi})$. This implies with Theorem 5.4 that $f_{\mathbf{K}}(\boldsymbol{y}^\star) = 0$, that $\boldsymbol{y}^\star$ is the minimizer of $f_{\mathbf{K}}$ over $\mathbf{K}$, and consequently that there exists a $\boldsymbol{v}^\star \in \boldsymbol{\partial} f_{\mathbf{K}}(\boldsymbol{y}^\star)$ such that $\langle \boldsymbol{v}^\star \,,\, \boldsymbol{z} - \boldsymbol{y}^\star \rangle \geq 0$ for all every $\boldsymbol{z} \in \mathbf{K}$. With the strong convexity of the function $f_{\mathbf{K}}$ with parameter μ, we have that

$$f_{\mathbf{K}}\left(\bar{\boldsymbol{y}}^{(k)}\right) \geq \underbrace{f_{\mathbf{K}}(\boldsymbol{y}^\star)}_{=0} + \left\langle \boldsymbol{v} \,,\, \bar{\boldsymbol{y}}^{(k)} - \boldsymbol{y}^\star \right\rangle + \frac{1}{2}\mu \left\| \bar{\boldsymbol{y}}^{(k)} - \boldsymbol{y}^\star \right\|^2 \tag{5.76}$$

for any $\boldsymbol{v} \in \boldsymbol{\partial} f_{\mathbf{K}}(\boldsymbol{y}^\star)$. In particular, (5.76) holds for $\boldsymbol{v} = \boldsymbol{v}^\star$, which proves to the first inequality of (5.75). Using a similar reasoning, we obtain that $\frac{1}{2}\mu \left\| \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\|^2 \leq f_{\mathbf{K}}\left(\boldsymbol{y}^{(0)}\right)$, which proves the last inequality.

In view of Corollary 5.2, $\widetilde{\Delta}_{\mathbf{K}}^{(k)} \leq \widetilde{\Delta}_{\mathbf{K}}^{(k-1)}$ for any $k \in [\![1, \bar{k}]\!]$. We also have that

$$S^{(k+1)} = S^{(k)} + \alpha^{(k+1)} = \left(1 + \frac{1}{\gamma_\delta}\right) S^{(k)} = \frac{1}{1 - \frac{1}{\gamma_\delta + 1}} S^{(k)}. \tag{5.77}$$

This leads with the inequality $\exp(x) \geq 1 + x$ to

$$f_{\mathbf{K}}\left(\bar{\boldsymbol{y}}^{(k)}\right) \overset{(5.85)}{\leq} \frac{\widetilde{\Delta}_{\mathbf{K}}^{(k)}}{S^{(k)}} \leq \widetilde{\Delta}_{\mathbf{K}}^{(0)} \left(1 - \frac{1}{\gamma_\delta + 1}\right)^k \leq \widetilde{\Delta}_{\mathbf{K}}^{(0)} \exp\left(-\frac{k}{\gamma_\delta + 1}\right). \tag{5.78}$$

Finally, we can bound $\widetilde{\Delta}_{\mathbf{K}}^{(0)}$ as follows:

$$\begin{aligned}
\widetilde{\Delta}_{\mathbf{K}}^{(0)} &= \max_{\boldsymbol{z} \in \mathbf{K}} \widetilde{\Psi}_0(\boldsymbol{z}) = \max_{\boldsymbol{z} \in \mathbf{K}} \widetilde{\psi}^{\mu}_{\boldsymbol{y}^{(0)}}(\boldsymbol{z}) \\
&= \max_{\boldsymbol{z} \in \mathbf{K}} \left\{ \min_{\boldsymbol{u} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(0)}\right)} \left\{ \left\langle \boldsymbol{u} \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{z} \right\rangle \right\} - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(0)} \right\|^2 \right\} \\
&= \max_{\boldsymbol{z} \in \mathbf{K}} \left\{ \min_{\boldsymbol{u} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(0)}\right)} \left\{ \left\langle \boldsymbol{u} - \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{z} \right\rangle \right\} + \left\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{z} \right\rangle - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(0)} \right\|^2 \right\},
\end{aligned}$$

where $\boldsymbol{u}^\star \in \boldsymbol{\Xi}(\boldsymbol{y}^\star)$ is such that $(\boldsymbol{y}^\star, \boldsymbol{u}^\star)$ is the solution to GVI$(\mathbf{K}, \boldsymbol{\Xi})$, which implies that (5.1) holds. Hence

$$\begin{aligned}
\widetilde{\Delta}_{\mathbf{K}}^{(0)} &= \max_{\boldsymbol{z} \in \mathbf{K}} \left\{ \min_{\boldsymbol{u} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(0)}\right)} \left\{ \left\langle \boldsymbol{u} - \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{z} \right\rangle \right\} + \left\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star + \boldsymbol{y}^\star - \boldsymbol{z} \right\rangle - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(0)} \right\|^2 \right\} \\
&\overset{(5.1)}{\leq} \left\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\rangle + \max_{\boldsymbol{z} \in \mathbf{K}} \left\{ \min_{\boldsymbol{u} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(0)}\right)} \left\{ \left\langle \boldsymbol{u} - \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{z} \right\rangle \right\} - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(0)} \right\|^2 \right\} \\
&\leq \left\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\rangle + \max_{\boldsymbol{z} \in \mathbf{K}} \left\{ \max_{\boldsymbol{u} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(0)}\right)} \left\{ \left\langle \boldsymbol{u} - \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{z} \right\rangle \right\} - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(0)} \right\|^2 \right\} \\
&\leq \left\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\rangle + \max_{\boldsymbol{u} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(0)}\right)} \left\{ \max_{\boldsymbol{z} \in \mathbf{K}} \left\{ \left\langle \boldsymbol{u} - \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{z} \right\rangle - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(0)} \right\|^2 \right\} \right\}.
\end{aligned}$$

The innermost optimisation problem if not constrained attains its maximum for $\boldsymbol{y}^{(0)} - \boldsymbol{z} = \frac{\boldsymbol{u} - \boldsymbol{u}^\star}{\mu}$. Its objective value can thus be upper-bounded by $\frac{1}{2\mu} \left\| \boldsymbol{u} - \boldsymbol{u}^\star \right\|^2$, which implies that

$$\widetilde{\Delta}_{\mathbf{K}}^{(0)} \leq \left\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\rangle + \max_{\boldsymbol{u} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(0)}\right)} \left\{ \frac{1}{2\mu} \left\| \boldsymbol{u} - \boldsymbol{u}^\star \right\|^2 \right\} \tag{5.79a}$$

$$\overset{(5.38)}{\leq} \left\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\rangle + \frac{L_\delta^2}{2\mu} \left\| \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\|^2 \tag{5.79b}$$

$$= \left\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\rangle + \frac{\mu\gamma_\delta^2}{2} \left\| \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\|^2 . \tag{5.79c}$$

By the definition of $f_{\mathbf{K}}$, *i.e.* Equation (5.50), we have that

$$\left\langle \boldsymbol{u}^\star \,,\, \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\rangle + \frac{1}{2}\mu \left\| \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\|^2 \leq f_{\mathbf{K}}\left(\boldsymbol{y}^{(0)}\right), \tag{5.80}$$

which implies that

$$\widetilde{\Delta}_{\mathbf{K}}^{(0)} \leq f_{\mathbf{K}}\left(\boldsymbol{y}^{(0)}\right) + \mu\left(\frac{\gamma_\delta^2 - 1}{2}\right) \left\| \boldsymbol{y}^{(0)} - \boldsymbol{y}^\star \right\|^2 . \tag{5.81}$$

This proves the middle inequality of (5.75) when combined with (5.78) and completes the proof. □

While the convergence remains linear, we observe a gain in the convergence rate. In the projected subgradient-type method, the convergence is driven by the exponential term in which figures the square of the condition number γ_δ (see (5.45)), whereas the convergence of the dual extrapolation method is driven by the exponential, which is only influenced by γ_δ (see (5.75)). More precisely, the convergence of the dual extrapolation method is driven by $\exp\left(-\frac{k}{\gamma_\delta+1}\right)$ instead of $\exp\left(\frac{-k}{\gamma_\delta^2}\right)$ as for the projected subgradient-type method. This can lead to a significant difference when the condition number is large.

5.4.3 Simplified dual extrapolation (SDE) method

In the previous subsection, we have seen that the dual extrapolation method has better convergence properties compared to the projected subgradient-type method. However, from the implementation point of view, finding the next iterate $\boldsymbol{y}^{(k+1)}$ requires the solve a maximisation problem with an embedded minimisation problem over the convex and compact set $\boldsymbol{\Xi}\left(\boldsymbol{x}^{(k)}\right)$ (and similarly when computing $\boldsymbol{x}^{(k)}$). This can be computationally challenging, and consequently mitigate the gain previously obtained in the convergence rate bound. Here, we show that we can get a similar convergence result as the one of Theorem 5.6 while avoiding this minimisation subproblem plaguing the computational complexity by taking a random element of $\boldsymbol{\Xi}(\boldsymbol{y})$ instead of minimising over the entire set $\boldsymbol{\Xi}(\boldsymbol{y})$ in (5.61a). In this regard, given $\beta > 0$, we define

$$\psi_{\boldsymbol{y},\boldsymbol{u}}^{\beta}(\boldsymbol{z}) \coloneqq \langle \boldsymbol{u} \,,\, \boldsymbol{y} - \boldsymbol{z} \rangle - \frac{1}{2}\beta \left\| \boldsymbol{z} - \boldsymbol{y} \right\|^2 , \tag{5.82a}$$

$$\Psi_k(\boldsymbol{z}) \coloneqq \sum_{i=0}^{k} \alpha^{(i)} \psi_{\boldsymbol{y}^{(i)},\boldsymbol{u}^{(i)}}^{\mu}(\boldsymbol{z}) . \tag{5.82b}$$

The simplified method also relies on $S^{(k)}$ and $\bar{\boldsymbol{y}}^{(k)}$ respectively defined in (5.56a) and (5.56b). Given the sequences $\left(\alpha^{(i)}\right)_{i=0}^{k}$, $\left(\boldsymbol{y}^{(i)}\right)_{i=0}^{k}$ and $\left(\boldsymbol{u}^{(i)}\right)_{i=0}^{k}$ with $\boldsymbol{u}^{(i)} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(i)}\right)$, we define $\Delta_{\mathbf{K}}^{(k)}$ as follows:

$$\Delta_{\mathbf{K}}^{(k)} \coloneqq \max_{\boldsymbol{z}\in\mathbf{K}} \left\{ \sum_{i=0}^{k} \alpha^{(i)} \left(\left\langle \boldsymbol{u}^{(i)} \,,\, \boldsymbol{y}^{(i)} - \boldsymbol{z} \right\rangle - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(i)} \right\|^2 \right) \right\} . \tag{5.83}$$

From the definition of $\Delta_{\mathbf{K}}^{(k)}$, we observe that $\Delta_{\mathbf{K}}^{(k)} = \max_{\boldsymbol{z}\in\mathbf{K}} \Psi_k(\boldsymbol{z})$. We now define the following iterations given $\left(\boldsymbol{y}^{(i)}\right)_{i=0}^{k}$ and $\left(\boldsymbol{u}^{(i)}\right)_{i=0}^{k}$ with $\boldsymbol{u}^{(i)} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(i)}\right)$:

$$\boldsymbol{x}^{(k)} := \arg\max_{\boldsymbol{z}\in\mathbf{K}} \Psi_k(\boldsymbol{z})\,, \tag{5.84a}$$

$$\boldsymbol{v}^{(k)} \in_R \boldsymbol{\Xi}\left(\boldsymbol{x}^{(k)}\right), \tag{5.84b}$$

$$\boldsymbol{y}^{(k+1)} := \arg\max_{\boldsymbol{z}\in\mathbf{K}} \psi^{\beta}_{\boldsymbol{x}^{(k)},\boldsymbol{v}^{(k)}}(\boldsymbol{z})\,, \tag{5.84c}$$

$$\boldsymbol{u}^{(k+1)} \in_R \boldsymbol{\Xi}\left(\boldsymbol{y}^{(k+1)}\right). \tag{5.84d}$$

We should note that computing (5.84a) and (5.84c) can be reduced to an orthogonal projection on $\mathbf{K}$ (see Proposition C.2). Similarly to Lemma 5.4, we can show that the following result holds.

LEMMA 5.5.

$$f_{\mathbf{K}}\left(\bar{\boldsymbol{y}}^{(k)}\right) \leq \frac{1}{S^{(k)}}\Delta_{\mathbf{K}}^{(k)}. \tag{5.85}$$

PROOF. By comparing (5.56c) with (5.83), it is straightforward that $\widetilde{\Delta}_{\mathbf{K}}^{(k)} \leq \Delta_{\mathbf{K}}^{(k)}$. The result then holds based on Lemma 5.4. □

THEOREM 5.7. *If* $\alpha^{(k+1)} \leq \frac{\mu}{\beta} S^{(k)}$ *then*

$$\Delta_{\mathbf{K}}^{(k+1)} \leq \Delta_{\mathbf{K}}^{(k)} + \frac{1}{2}\alpha^{(k+1)}\left[\frac{1}{\mu+\beta}\left\|\boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)}\right\|^2 - \beta\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right\|^2\right]. \tag{5.86}$$

PROOF. The same steps as for the proof of Theorem 5.5 can be followed. We note that $\Psi_k\left(\boldsymbol{x}^{(k)}\right) = \Delta_{\mathbf{K}}^{(k)}$ and $\Psi_{k+1}(\boldsymbol{z}) = \Psi_k(\boldsymbol{z}) + \alpha^{(k+1)}\psi^{\mu}_{\boldsymbol{y}^{(k+1)},\boldsymbol{u}^{(k+1)}}(\boldsymbol{z})$, which allow us to bound $\Delta_{\mathbf{K}}^{(k+1)}$ as follows:

$$\begin{aligned}
\Delta_{\mathbf{K}}^{(k+1)} &= \max_{\boldsymbol{z}\in\mathbf{K}}\left\{\Psi_k(\boldsymbol{z}) + \alpha^{(k+1)}\psi^{\mu}_{\boldsymbol{y}^{(k+1)},\boldsymbol{u}^{(k+1)}}(\boldsymbol{z})\right\} \\
&\overset{(5.64)}{\leq} \max_{\boldsymbol{z}\in\mathbf{K}}\left\{\Delta_{\mathbf{K}}^{(k)} - \frac{1}{2}\mu S^{(k)}\left\|\boldsymbol{z} - \boldsymbol{x}^{(k)}\right\|^2 + \alpha^{(k+1)}\psi^{\mu}_{\boldsymbol{y}^{(k+1)},\boldsymbol{u}^{(k+1)}}(\boldsymbol{z})\right\} \\
&\leq \Delta_{\mathbf{K}}^{(k)} + \max_{\boldsymbol{z}\in\mathbf{K}}\left\{-\frac{1}{2}\mu S^{(k)}\left\|\boldsymbol{z} - \boldsymbol{x}^{(k)}\right\|^2 + \alpha^{(k+1)}\left(\left\langle \boldsymbol{u}^{(k+1)}, \boldsymbol{y}^{(k+1)} - \boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z} - \boldsymbol{y}^{(k+1)}\right\|^2\right)\right\}.
\end{aligned} \tag{5.87}$$

Then, as $\boldsymbol{y}^{(k+1)}$ maximises $\psi^{\beta}_{\boldsymbol{x}^{(k)},\boldsymbol{v}^{(k)}}$, we have that

$$\left\langle -\boldsymbol{v}^{(k)} - \beta\left(\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right), \boldsymbol{z} - \boldsymbol{y}^{(k+1)}\right\rangle \leq 0, \quad \forall \boldsymbol{z} \in \mathbf{K}\,. \tag{5.88}$$

For the remaining part of this proof, we refer the reader to the proof of Theorem 5.5 as they are identical. □

COROLLARY 5.3. *Assume that* $\boldsymbol{\Xi}$ *is* μ*-strongly monotone and* δ*-far-sighted Lipschitz continuous with constant* L_δ *on* $\mathbf{K}$ *for a given* $\delta > 0$*. With the choice*

$$\beta = L_\delta, \quad \alpha^{(k+1)} = \frac{1}{\gamma_\delta} S^{(k)}, \tag{5.89}$$

where $\gamma_\delta = \frac{L_\delta}{\mu}$ *is the condition number of* $\boldsymbol{\Xi}$ *related to* δ*, we obtain that* $\Delta_{\mathbf{K}}^{(k+1)} \leq \Delta_{\mathbf{K}}^{(k)}$ *while* $\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right\| \geq \delta$.

PROOF. The proof is identical as the one of Corollary 5.2. □

With Corollary 5.3, we have everything needed to prove the convergence of the simplified dual extrapolation method stated in Algorithm 5.3. Given a value $\delta > 0$, the simplified dual extrapolation method is formulated as follows

ALGORITHM 5.3 (SIMPLIFIED DUAL EXTRAPOLATION (SDE) METHOD).

Input: *Choose* $\boldsymbol{y}^{(0)} \in \mathbf{K}$ *and* $\beta > 0$.

Initialisation: *Set* $\alpha^{(0)} = 1$, *and* $\gamma = \frac{\beta}{\mu}$.

Iteration ($k \geq 0$)**:**

$$
\begin{aligned}
&S^{(k)} = \textstyle\sum_{i=0}^{k} \alpha^{(i)} \text{ and } \alpha^{(k+1)} = \frac{1}{\gamma} S^{(k)},\\
&\boldsymbol{u}^{(k)} \in_R \boldsymbol{\Xi}\Big(\boldsymbol{y}^{(k)}\Big),\\
&\boldsymbol{x}^{(k)} = \arg\max_{\boldsymbol{z} \in \mathbf{K}} \Psi_k(\boldsymbol{z}),\\
&\boldsymbol{v}^{(k)} \in_R \boldsymbol{\Xi}\Big(\boldsymbol{x}^{(k)}\Big),\\
&\boldsymbol{y}^{(k+1)} = \arg\max_{\boldsymbol{z} \in \mathbf{K}} \psi^{\beta}_{\boldsymbol{x}^{(k)}, \boldsymbol{v}^{(k)}}(\boldsymbol{z}).
\end{aligned}
$$

Output: $\bar{\boldsymbol{y}}^{(k)} = \frac{1}{S^{(k)}} \sum_{i=0}^{k} \alpha^{(i)} \boldsymbol{y}^{(i)}$.

Again, a similar theorem as Theorem 5.6 holds true:

THEOREM 5.8. *Let* $\boldsymbol{\Xi}$ *be a* μ*-strongly monotone and* δ*-far-sighted Lipschitz continuous with parameter* L_δ *on* $\mathbf{K}$. *Let* $\left\{\bar{\boldsymbol{y}}^{(k)}\right\}_{k \in [\![0, \bar{k}]\!]}$ *be the sequence generated by Algorithm 5.3 with* $\beta = L_\delta$, *where* $\bar{k}$ *is the first iteration number such that* $\left\|\boldsymbol{y}^{(\bar{k}+1)} - \boldsymbol{x}^{(\bar{k})}\right\| < \delta$ *holds true. If* $\left\|\boldsymbol{y}^{(0)} - \boldsymbol{y}^{\star}\right\| > \delta$, *then for any* $k \in [\![0, \bar{k}]\!]$*:*

$$
\begin{aligned}
\frac{1}{2}\mu \left\|\bar{\boldsymbol{y}}^{(k)} - \boldsymbol{y}^{\star}\right\|^2 \leq f_{\mathbf{K}}\Big(\bar{\boldsymbol{y}}^{(k)}\Big) &\leq \left[f_{\mathbf{K}}\Big(\boldsymbol{y}^{(0)}\Big) + \frac{\mu(\gamma_\delta^2 - 1)}{2}\left\|\boldsymbol{y}^{(0)} - \boldsymbol{y}^{\star}\right\|^2\right] \exp\left(-\frac{k}{\gamma_\delta + 1}\right)\\
&\leq f_{\mathbf{K}}\Big(\boldsymbol{y}^{(0)}\Big) \cdot \gamma_\delta^2 \cdot \exp\left(-\frac{k}{\gamma_\delta + 1}\right). \qquad (5.90)
\end{aligned}
$$

PROOF. The proof follows from Nesterov and Scrimali (2011, Theorem 2.4) and Corollary 5.3 and follows the same steps as for Theorem 5.6. □

As for Algorithm 5.2, the convergence of Algorithm 5.3 is linear with a convergence rate that is driven by the condition number γ_δ rather than γ_δ^2 as for the projected subgradient-type method. Additionally, Algorithm 5.3 avoids the inner minimisation problem in (5.61a) and (5.61b), which was mitigating the gain in the convergence rate obtained with Algorithm 5.2. Hence, unlike Algorithm 5.2, Algorithm 5.3 has a computational cost per iteration that is comparable to the projected subgradient-type method, *i.e.* Algorithm 5.3 requires two projections and two computations of $\boldsymbol{\Xi}$ instead of one as it is the case for Algorithm 5.1 (see Proposition C.2).

5.4.3.1 Convergence discussion

Unlike the projected subgradient-type method whose convergence depends only on the far-sighted Lipschitz continuity constant at $\boldsymbol{y}^{\star}$. Theorem 5.8 tells us that convergence is guaranteed

as long as $\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\| > \delta$. This condition condition is only necessary to certify that $\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\| \leq L_\delta \left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|$ holds, which then implies with Corollary 5.3 that $\Delta_{\mathbf{K}}^{(k+1)} \leq \Delta_{\mathbf{K}}^{(k)}$. As long as this last inequality is verified, we have that (5.90) holds. We could consequently replace the stopping criteria of Algorithm 5.3, *i.e.* $\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\| < \delta$, by $\Delta_{\mathbf{K}}^{(k+1)} \geq \Delta_{\mathbf{K}}^{(k)}$. However, if $\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\| < \delta$, we can derive an upper-bound on the distance between $\boldsymbol{x}^{(k)}$ to $\boldsymbol{y}^\star$ using the following result:

LEMMA 5.6. *Let $\boldsymbol{x}^{(k)}$ and $\boldsymbol{y}^{(k+1)}$ be some iterates generated with Algorithm 5.3, and $L_{\boldsymbol{y}^\star,\delta}$ be the δ-far-sighted Lipschitz continuity constant on $\mathbf{K}$ at $\boldsymbol{y}^\star$. If $\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}^{(k+1)}\right\| < \delta$, then*

$$\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}^\star\right\| \leq \left(\frac{L_{\boldsymbol{y}^\star,\delta}+\beta}{\mu}\right)\delta. \tag{5.91}$$

PROOF. By the definition of $\boldsymbol{y}^{(k+1)}$ in (5.84c) and Proposition C.2, we have that

$$\begin{aligned}\boldsymbol{y}^{(k+1)} &= \arg\max_{\boldsymbol{z}\in\mathbf{K}} \psi^{\beta}_{\boldsymbol{x}^{(k)},\boldsymbol{v}^{(k)}}(\boldsymbol{z}) \\ &\overset{(5.82a)}{=} \arg\max_{\boldsymbol{z}\in\mathbf{K}} \left\langle \boldsymbol{v}^{(k)}, \boldsymbol{x}^{(k)}-\boldsymbol{z}\right\rangle - \frac{1}{2}\beta\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2 \\ &\overset{(C.2)}{=} \Pi_{\mathbf{K}}\left(\boldsymbol{x}^{(k)}-\frac{1}{\beta}\boldsymbol{v}^{(k)}\right) \overset{(5.22)}{=} \boldsymbol{x}^{(k)} - \boldsymbol{\Xi}^{\mathbf{nat}}_{\mathbf{K}}\left(\boldsymbol{x}^{(k)}, \frac{1}{\beta}\boldsymbol{v}^{(k)}\right).\end{aligned} \tag{5.92}$$

If $\boldsymbol{x}^{(k)} \notin \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$, we consequently obtain with Proposition C.1 that

$$\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}^\star\right\| \leq \left(\frac{L_{\boldsymbol{y}^\star,\delta}+\beta}{\mu}\right)\left\|\boldsymbol{\Xi}^{\mathbf{nat}}_{\mathbf{K}}\left(\boldsymbol{y}, \frac{1}{\beta}\boldsymbol{v}^{(k)}\right)\right\| \overset{(5.92)}{=} \left(\frac{L_{\boldsymbol{y}^\star,\delta}+\beta}{\mu}\right)\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}^{(k+1)}\right\| \tag{5.93}$$

$$\leq \left(\frac{L_{\boldsymbol{y}^\star,\delta}+\beta}{\mu}\right)\delta, \tag{5.94}$$

which proves (5.91). If $\boldsymbol{x}^{(k)} \in \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$, then (5.91) holds since $\left(\frac{L_{\boldsymbol{y}^\star,\delta}+\beta}{\mu}\right) \geq 1$. □

With Lemma 5.6, we notice that to obtain iterates $\boldsymbol{x}^{(k)}$ and $\bar{\boldsymbol{y}}^{(k)}$ that converges to $\boldsymbol{y}^\star$, we should to take $\delta \to 0$ such that $\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}^{(k+1)}\right\| < \delta$ holds as long as possible and such that the right-hand side of (5.91) tends to 0. This is however not as straightforward. Indeed, L_δ increases as δ decreases and β is chosen equal to L_δ in Algorithm 5.3, which implies that (5.91) can be replaced by:

$$\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}^\star\right\| \leq \left(\frac{L_{\boldsymbol{y}^\star,\delta}+L_\delta}{\mu}\right)\delta \leq \left(\frac{2L_\delta}{\mu}\right)\delta. \tag{5.95}$$

Hence, it is interesting to study how L_δ varies with δ. To do so, we define the *δ-neighbourhood* of a compact and convex set $\mathbf{K} \subset \mathbb{R}^d$ as follows:

$$\mathbf{K}^{[\delta]} := \left\{\boldsymbol{y} \in \mathbb{R}^d \,\middle|\, \|\boldsymbol{y}-\Pi_{\mathbf{K}}(\boldsymbol{y})\| \leq \delta\right\}. \tag{5.96}$$

Based on this definition, the following result hold:

PROPOSITION 5.11. *Let $r \geq 1$, $\delta > 0$ and $\mathbf{K}$ be a compact convex set. If the set-valued map $\boldsymbol{\Xi}$ is μ-strongly monotone on $\mathbf{K}^{[\delta]}$ and δ-far-sighted Lipschitz continuous with constant L on $\mathbf{K}^{[\delta]}$ at any point $\boldsymbol{x} \in \mathbf{K}$. Then $\boldsymbol{\Xi}$ is $\frac{\delta}{r}$-far-sighted Lipschitz continuous with constant rL on $\mathbf{K}^{[\delta]}$ at any point $\boldsymbol{x} \in \mathbf{K}$.*

PROOF. The proof is provided in Appendix C.2 at page 189. □

Proposition 5.11 implies that, even if rL is not the optimal constant, *i.e.* the smallest constant such that $\boldsymbol{\Xi}$ is $\frac{\delta}{r}$-far-sighted Lipschitz continuous on $\mathbf{K}^{[\delta]}$ at any point $\boldsymbol{x} \in \mathbf{K}$, it provides a rule of thumb for updating the far-sighted Lipschitz continuity constant when decreasing the value of δ. It can be shown that, if there exists a $\boldsymbol{y} \in \mathbf{K}$ such that $\boldsymbol{\Xi}(\boldsymbol{y})$ is not a singleton, then there exists a threshold value $\bar{\delta}$ with associated optimal constant $L_{\bar{\delta}}$ such that the far-sighted Lipschitz continuous constant for any $\delta = \frac{\bar{\delta}}{r}$ with $r \geq 1$ is given by $L_\delta = rL_{\bar{\delta}}$. This is for example illustrated in Figure 5.5, which represents the relationship of $L_\delta^{\mathbf{K}}$ with δ for $\mathbf{K}$ and $\boldsymbol{\Xi}$ defined as in (5.98).

This observation has two implications when there exists a $\boldsymbol{y} \in \mathbf{K}$ such that $\boldsymbol{\Xi}(\boldsymbol{y})$ is not a singleton: Firstly, according to Lemma 5.6, if $\left\|\boldsymbol{x}^{(k)} - \boldsymbol{y}^{(k+1)}\right\| < \delta$, then (5.95) holds true. This implies that for any $\delta < \bar{\delta}$, (5.95) can be rewritten as

$$\left\|\boldsymbol{x}^{(k)} - \boldsymbol{y}^\star\right\| \leq \left(\frac{2L_{\bar{\delta}}}{\mu}\right)\bar{\delta}. \tag{5.97}$$

Hence, there exist threshold value $\bar{\delta}$ such taking a smaller value for δ does not lead to any better convergence guarantee. Secondly, we could think that taking a smaller value for δ would imply that the stopping criteria $\left\|\boldsymbol{x}^{(k)} - \boldsymbol{y}^{(k+1)}\right\| < \delta$ is harder to be respected which in turn would lead to a better convergence. This is unfortunately not the case either. Indeed, the iterate $\boldsymbol{y}^{(k+1)}$ is defined as the arg max of $\psi^{\beta}_{\boldsymbol{x}^{(k)},\boldsymbol{v}^{(k)}}$ over $\mathbf{K}$, which is equal to $\Pi_{\mathbf{K}}\left(\boldsymbol{x}^{(k)} - \frac{1}{\beta}\boldsymbol{v}^{(k)}\right)$ in view of Proposition C.2. This means that, if $\left\|\boldsymbol{x}^{(k)} - \boldsymbol{y}^{(k+1)}\right\| < \bar{\delta}$, where $\beta = L_{\bar{\delta}}^{\mathbf{K}}$, then if δ is divided by a factor $r > 1$, then the step length $\frac{1}{\beta}$ is also divided by r since $\beta = L_\delta^{\mathbf{K}}$.

In conclusion, if for each neighbourhood of $\boldsymbol{y}^\star$, there exist a point such that the set-valued map $\boldsymbol{\Xi}$ is multi-valued, then the progress of Algorithm 5.3 towards $\boldsymbol{y}^\star$ is probable to stagnate. However, if we can find a convex neighbourhood $\mathcal{W}$ of $\boldsymbol{y}^\star$ on which $\boldsymbol{\Xi}$ is Lipschitz continuous, then there exists a value $L^{\mathcal{W}}$ such that $\boldsymbol{\Xi}$ is δ-far-sighted Lipschitz continuous on $\mathcal{W}$ for any $\delta > 0$. We could consequently use Algorithm 5.3 on $\mathcal{W}$ with $\beta = L^{\mathcal{W}}$ and $\delta = 0$, which guarantees the convergence towards $\boldsymbol{y}^\star$.

5.4.4 Numerical results (Part I)

In this section, we illustrate the performance and convergence results of the PS and SDE methods; the ODE method produces essentially the same results compared to those of the SDE method. To this end, we consider the generalised variational inequality problem defined as follows: let $\boldsymbol{s} \in \mathbb{R}^2$, $\mathbf{K} = [-3,3] \times [-3,3] \subset \mathbb{R}^2$ and $\boldsymbol{\Xi}$ be the set-valued map defined as

$$\boldsymbol{\Xi}(\boldsymbol{y}) = \begin{cases} [y_1 + y_2 + s_1,\ y_1 + 2y_2 + s_2]^{\mathrm{T}}, & \text{if } y_1 y_2 > 0, \\ [y_1 - y_2 + s_1,\ -y_1 + 2y_2 + s_2]^{\mathrm{T}}, & \text{if } y_1 y_2 < 0, \\ \mathbf{conv}\left\{ \begin{array}{l} [y_1 - y_2 + s_1,\ -y_1 + 2y_2 + s_2]^{\mathrm{T}}, \\ [y_1 - y_2 + s_1,\ -y_1 + 2y_2 + s_2]^{\mathrm{T}} \end{array} \right\}, & \text{if } y_1 y_2 = 0, \end{cases} \tag{5.98}$$

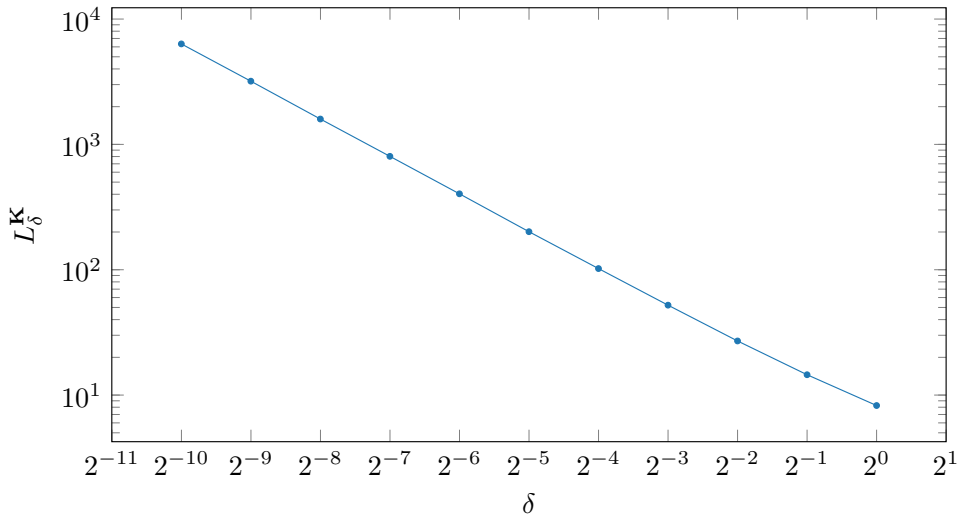


FIGURE 5.5: Relationship between $L_\delta^{\mathbf{K}}$ and δ for $\mathbf{K}$ and $\boldsymbol{\Xi}$ defined in (5.98).

where $\mathbf{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ is the convex hull of the vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. For any vector $\boldsymbol{s} \in \mathbb{R}^2$, $\boldsymbol{\Xi}$ is piecewise single-valued: the only points for which $\boldsymbol{\Xi}$ is set-valued is on the lines $y_1 = 0$ and $y_2 = 0$.

It can be shown that, if $\boldsymbol{\Xi}$ is monotone, then the parameter μ of the strong monotonicity of $\boldsymbol{\Xi}$ on $\mathbf{K}$ corresponds to the minimum strongly-monotone parameter value μ over each subset of the domain that is single-valued. We estimate now the strong-monotonicity parameter μ on $\mathbb{R}^2_{++} := \left\{ \boldsymbol{y} \in \mathbb{R}^2 \,\middle|\, y_1 > 0 \text{ and } y_2 > 0 \right\}$; a similar approach can be used for the other pieces. We use an approach similar to that of Rosen (1965, Theorem 6): let $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(2)}$ be two points in $\mathbb{R}^2_{++}$, $\boldsymbol{u}^{(i)} = \boldsymbol{\Xi}\left(\boldsymbol{y}^{(i)}\right)$ for $i \in \{1, 2\}$, $\boldsymbol{y}(\theta) = \theta \boldsymbol{y}^{(2)} + (1-\theta)\, \boldsymbol{y}^{(1)}$, $\mathbf{G}(\theta) = \mathbf{G}$ be the Jacobian of $\boldsymbol{\Xi}(\boldsymbol{y}(\theta))$, *i.e.*

$$\mathbf{G} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \tag{5.99}$$

and $\mathbf{C}$ be the Cholesky decomposition of $\mathbf{G}$ which exists as it is positive definite, *i.e.* $\mathbf{G} = \mathbf{C}^T\, \mathbf{C}$. Then

$$\frac{\mathrm{d}\boldsymbol{\Xi}(\boldsymbol{y}(\theta))}{\mathrm{d}\theta} = \mathbf{G}\, \frac{\mathrm{d}\boldsymbol{y}(\theta)}{\mathrm{d}\theta} = \mathbf{G}\left(\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)}\right). \tag{5.100}$$

This implies that

$$\boldsymbol{u}^{(2)} - \boldsymbol{u}^{(1)} = \int_0^1 \frac{\mathrm{d}\boldsymbol{\Xi}(\boldsymbol{y}(\theta))}{\mathrm{d}\theta}\, \mathrm{d}\theta = \mathbf{G}\left(\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)}\right). \tag{5.101}$$

which in turn implies when taking the dot product with $\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)}$

$$\begin{aligned} \left\langle \boldsymbol{u}^{(2)} - \boldsymbol{u}^{(1)}\,,\, \boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)} \right\rangle &= \left\langle \mathbf{G}\left(\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)}\right)\,,\, \boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)} \right\rangle \\ &= \left\langle \mathbf{C}^{\mathrm{T}}\, \mathbf{C}\left(\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)}\right)\,,\, \boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)} \right\rangle \\ &= \left\| \mathbf{C}\left(\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)}\right) \right\|_2^2 \geq (\sigma_{\min}(\mathbf{C}))^2 \left\| \left(\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)}\right) \right\|_2^2, \end{aligned} \tag{5.102}$$

where $\sigma_{\min}(\mathbf{C})$ is the smallest singular value of $\mathbf{C}$. Hence, given the GVI problem formulated in (5.98), we obtain $\mu \simeq 0.382$. This value corresponds to the square of the minimum singular value over each subset of the domain that is single-valued.

In view of the concluding remark of Section 5.4.3.1, we know that we can certify the convergence towards $\boldsymbol{y}^\star$ if we can find a neighbourhood $\mathcal{W}$ of $\boldsymbol{y}^\star$ on which $\boldsymbol{\Xi}$ is Lipschitz continuous. Therefore, we suggest to use as parameter $\beta = L^{con} \simeq 2.618$, where L^{con} is defined as the largest Lipschitz continuity constant over any subset of the domain on which $\boldsymbol{\Xi}$ is single-valued. For each of these subsets, the Lipschitz continuity constant corresponds to the largest eigenvalue $\lambda_{\max}(\mathbf{G})$ of the corresponding Jacobian matrix $\mathbf{G}$. Indeed, with (B.16) we have that

$$\left\|\boldsymbol{u}^{(2)} - \boldsymbol{u}^{(1)}\right\| = \left\|\mathbf{G}\left(\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)}\right)\right\| \leq \lambda_{\max}(\mathbf{G}) \left\|\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)}\right\|. \tag{5.103}$$

We should note that, while $\mu \simeq 0.382$ is the exact strong-monotonicity constant of $\boldsymbol{\Xi}$, taking $\beta = L^{con}$ is not the Lipschitz continuity constant of $\boldsymbol{\Xi}$ on $\mathbf{K}$; it is however a Lipschitz continuity constant for any subset on which $\boldsymbol{\Xi}$ is single-valued. Coupled with the notion of L^{con}, we denote by γ^{con} the upper-bound defined as $\gamma^{con} = L^{con}\mu^{-1}$ on largest condition number over any subset on which $\boldsymbol{\Xi}$ is single-valued.

In the numerical experiments, we consequently apply the PS method with $\alpha = \mu\beta^{-2} = 0.056$ and the SDE method with $\beta \simeq 2.618$. We apply these methods on two different cases:

(a) $\boldsymbol{s} = [1, -2]^{\mathrm{T}}$, which has as solution $\boldsymbol{y}^\star_{(a)} = [1, -2]^{\mathrm{T}}$;

(b) $\boldsymbol{s} = [2, 0]^{\mathrm{T}}$, which has as solution $\boldsymbol{y}^\star_{(b)} = [-2, 0]^{\mathrm{T}}$.

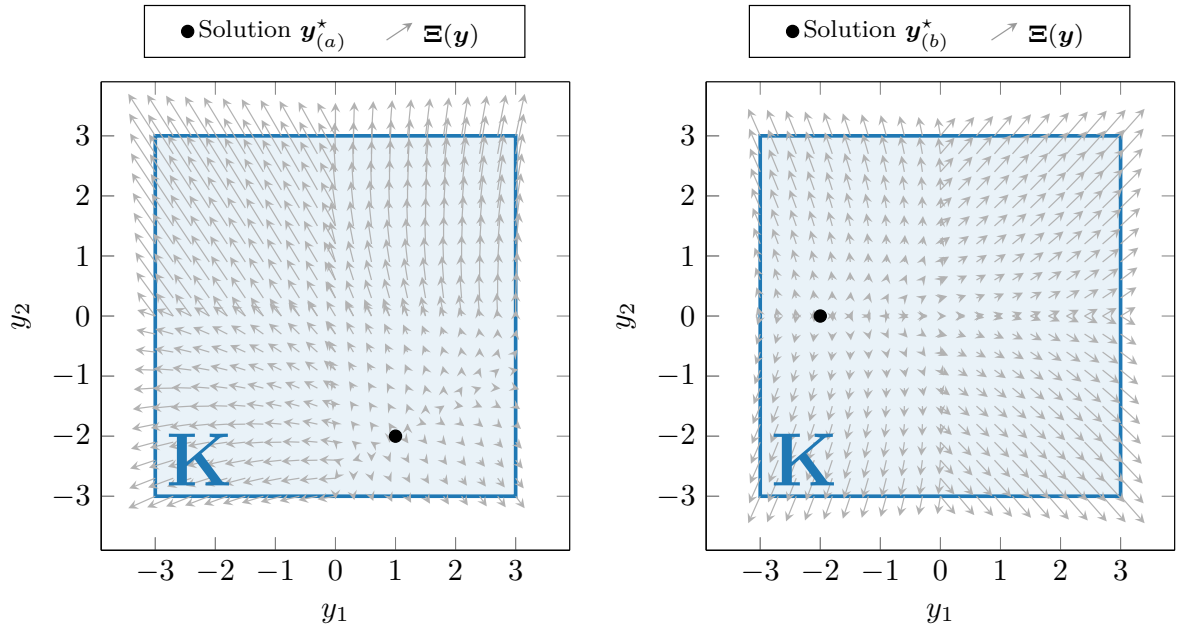
The set-valued map for both cases are represented in Figures 5.6a and 5.6b respectively.

In Figures 5.7 and 5.8, are illustrated the numerical results related to the PS and SDE methods respectively. For both methods, we represent the trajectories and the error of the iterates, *i.e.* $\left(\boldsymbol{y}^{(k)}\right)_{k\geq 0}$ for the PS method and $\left(\bar{\boldsymbol{y}}^{(k)}\right)_{k\geq 0}$ for the SDE method.

First, by comparing the error plots of both methods for Case (a) (see Figures 5.7c and 5.8c), we observe that the SDE method converges faster compared to the PS method. This is also observed in Table 5.1a by comparing the average number of iterations to reach any precision level. This observation remains valid if we take into account that an iteration of the SDE method is twice as expensive as an iteration of the PS method.

For Case (b) however, both methods fail to converge with high precision as represented in Figures 5.7d and 5.8d and reported in Table 5.1b. This difference of behaviour between both test cases is explained as follows: in the situation (a), there exists a relatively large neighbourhood $\mathcal{W}$ of $\boldsymbol{y}^\star_{(a)}$, such as for example $\mathcal{B}\left(\boldsymbol{y}^\star_{(a)}\,;\, 0.5\right)$, such that $\boldsymbol{\Xi}$ is Lipschitz continuous with constant β, whereas this is not the case for situation (b). It is the discontinuity of the set-valued map $\boldsymbol{\Xi}$ that prevents the convergence of the method.

To obtain a closer convergence to the solution $\boldsymbol{y}^\star_{(b)}$, we have with Theorem 5.3 that we can decrease the step length α in Algorithm 5.1 for the PS method. In Figure 5.9, we represent the convergence for the different step lengths, *i.e.* α, $\frac{1}{2}\alpha$ and $\frac{1}{10}\alpha$, where $\alpha = \mu\beta^{-2} = 0.056$



(a) Case (a): $\boldsymbol{s} = [1, -2]^{\mathrm{T}}$ and $\boldsymbol{y}_\star^{(a)} = [1, -2]^{\mathrm{T}}$.

(b) Case (b): $\boldsymbol{s} = [2, 0]^{\mathrm{T}}$ and $\boldsymbol{y}_\star^{(b)} = [-2, 0]^{\mathrm{T}}$.

FIGURE 5.6: Illustration of $\boldsymbol{\Xi}$ defined in (5.98) for both cases.

as previously defined. We observe the linear convergence of the method: by dividing the step length α by a factor $r > 1$, we have a final error that is divided by roughly the same factor but the number of iterates needed to reach convergence is also multiplied by r. For the SDE method, we can theoretically converge closer to the solution if we increase the parameter β in Algorithm 5.3. Similarly to the SDE method, we can observe in Figure 5.10 the linear convergence of the method: by multiplying the parameter value β by a factor $r > 1$, we have a final error that is divided by roughly the same factor. Note that, based on Figure 5.5, the value $\beta \simeq 2.618$ corresponds approximately to the far-sighted Lipschitz continuity constant for $\delta \simeq 1.447$ and we observe that the final error is of the same order of magnitude. With Proposition 5.11, we have approximately that $\delta \simeq 1.447 r^{-1}$ for $r\beta$. We observe that the error is of the order of the resulting δ. For instance if $r = 10$, with Figure 5.5 we have approximately that $\delta \simeq 0.1447$ for 10β and observe in Figure 5.10b that the error after convergence is of the same order of magnitude.

We have just seen that the convergence depends on how well behaved the set-valued map $\boldsymbol{\Xi}$ is around the solution $\boldsymbol{y}^\star$. Using constantly a large (*resp.* small) value for the parameter β (*resp.* α) in the SDE (*resp.* PS) method in order to minimise the final error after convergence is not a viable option as it is paired with a large number of iterations to reach convergence. Iteratively adapting the step size or the parameter value β is not as straightforward as for the corresponding convex optimisation method since we do not have an objective function to compare the quality of the different iterates. A different approach is therefore required. In the rest of this work, we devise a versatile method able on the one hand to take advantage of well-conditioned problems, and on the other hand to converge even for ill-conditioned problems.

To do so, we ameliorate Algorithm 5.3 in two ways: First, we use localisation sets, the idea of iteratively restricting the search of the optimal solution to a smaller set. By doing so, we hope to reach a neighbourhood $\mathcal{W}$ of $\boldsymbol{y}^\star$ on which $\boldsymbol{\Xi}$ is Lipschitz continuous, which would then imply that there exists a value $L^{\mathcal{W}}$ such that $\boldsymbol{\Xi}$ is δ-far-sighted Lipschitz continuous on $\mathcal{W}$ for any $\delta > 0$. We could consequently use Algorithm 5.3 restricted to $\mathcal{W}$ with $\beta = L^{\mathcal{W}}$ and $\delta = 0$, which guarantees the convergence towards $\boldsymbol{y}^\star$. Secondly, in the case where such neighbourhood $\mathcal{W}$ is not reached, we combine the SDE method with another method, the *cutting plane* method, that certifies the convergence towards $\boldsymbol{y}^\star$.

TABLE 5.1: Average number of iterations required by the PS, ODE and SDE methods to reach different levels of precision for both test cases. The entry "-" means that the level of precision has not been reached.

(a) Case (a)

Error	PS	ODE	SDE
1.0e0	18.0	10.0	10.0
1.0e-1	99.6	36.0	36.0
1.0e-2	201.7	57.0	57.0
1.0e-4	415.9	94.2	94.2
1.0e-6	629.9	129.9	129.9
1.0e-8	843.9	165.1	165.1

(b) Case (b)

Error	PS	ODE	SDE
1.0e0	17.1	14.0	14.0
1.0e-1	69.9	-	-
1.0e-2	-	-	-
1.0e-4	-	-	-
1.0e-6	-	-	-
1.0e-8	-	-	-

5.4.5 Cutting plane (CP) method

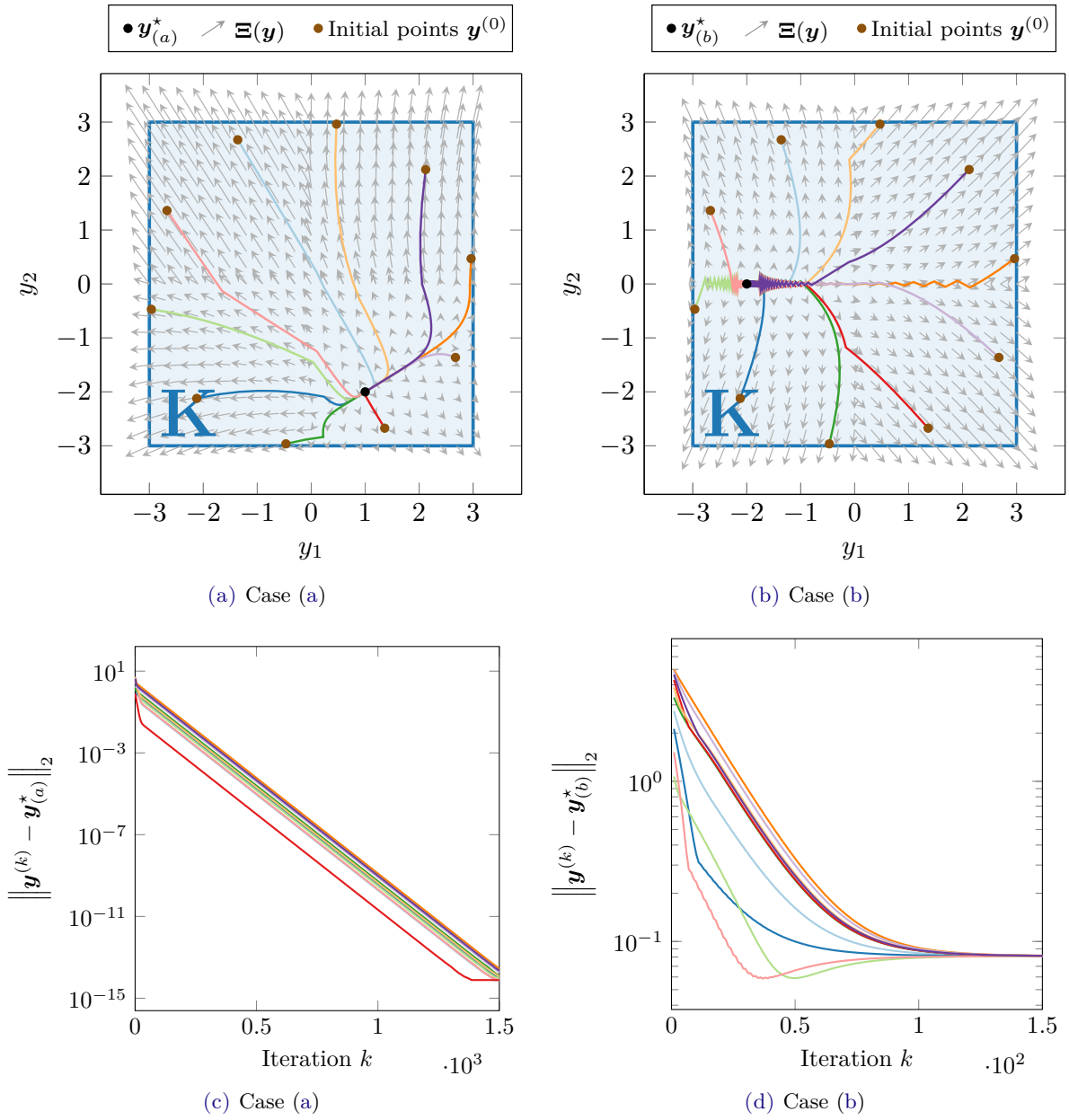
All the methods that we have presented so far converges only towards a neighbourhood of the optimal solution $\boldsymbol{y}^\star$. Once attained, the iterates oscillate in this neighbourhood around $\boldsymbol{y}^\star$, which is mainly due to the non-Lipschitz continuity of $\boldsymbol{\Xi}$. In the literature, a general scheme, the so-called *cutting plane scheme*, has been proposed to circumvent this issue of non-Lipschitz continuity and enables to construct methods that converge to $\boldsymbol{y}^\star$. This scheme relies on the interpretation of the monotonicity property made in Section 5.3.1 in order to iteratively shrink the set in which the optimal solution $\boldsymbol{y}^\star$ is located, which we define as the *localisation set* $\mathcal{L}$. If $\boldsymbol{\Xi}$ is monotone, then for any direction $\boldsymbol{w} \in \mathbb{R}^d$ and any $t > 0$ such that $\boldsymbol{y}^\star + t\boldsymbol{w} \in \mathbf{K}$, every element of $\boldsymbol{\Xi}(\boldsymbol{y}^\star + t\boldsymbol{w})$ has a non-negative component in the direction of $\boldsymbol{w}$. Formulated differently it means that

$$\forall \boldsymbol{y} \in \mathbf{K} \setminus \{\boldsymbol{y}^\star\} \text{ and } \forall \boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y}) : \quad \langle \boldsymbol{u}\,,\, \boldsymbol{y} - \boldsymbol{y}^\star \rangle \geq 0, \tag{5.104}$$

which implies that

$$\forall \boldsymbol{y} \in \mathbf{K} \setminus \{\boldsymbol{y}^\star\} \text{ and } \forall \boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y}) : \quad \boldsymbol{y}^\star \in \left\{ \boldsymbol{z} \in \mathbb{R}^d \,\middle|\, \langle \boldsymbol{u}\,,\, \boldsymbol{z} - \boldsymbol{y} \rangle \leq 0 \right\}. \tag{5.105}$$

If $\boldsymbol{\Xi}$ is strictly monotone, then the inequalities in Equations (5.104) and (5.105) hold in the strict sense. This means that, given a method that generates a sequence of iterates $\left\{\boldsymbol{y}^{(i)}\right\}_{i \in [\![0,k]\!]}$



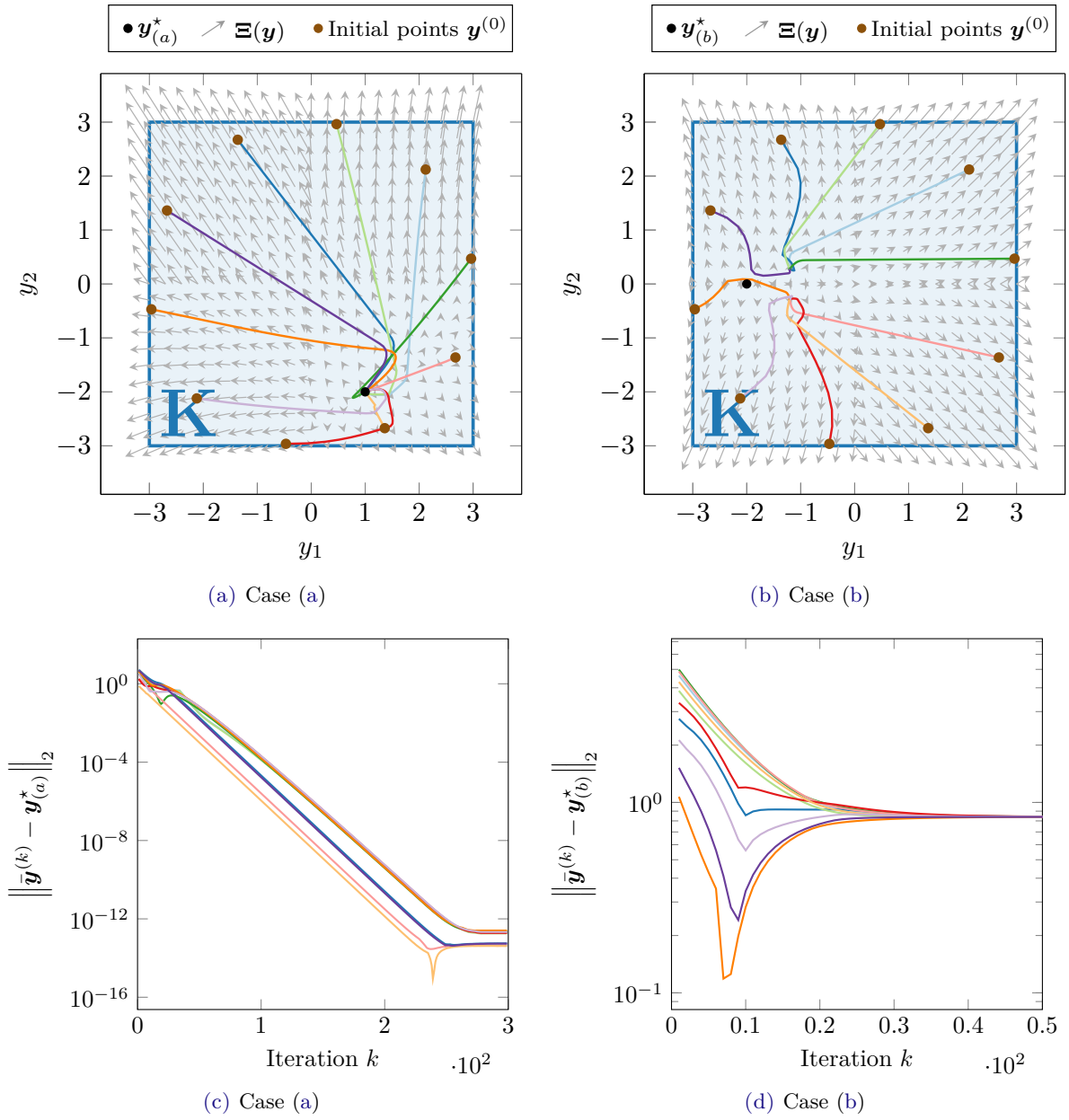
(a) Case (a)

(b) Case (b)

(c) Case (a)

(d) Case (b)

FIGURE 5.7: **PS method.** Figures 5.7a and 5.7b: Trajectories of the iterates $\left(\boldsymbol{y}^{(k)}\right)_{k\geq 0}$ for different initial points $\boldsymbol{y}^{(0)}$. Figures 5.7a and 5.7b: Convergence of $\left(\boldsymbol{y}^{(k)}\right)_{k\geq 0}$ towards the solution $\boldsymbol{y}^\star$ of the GVI.



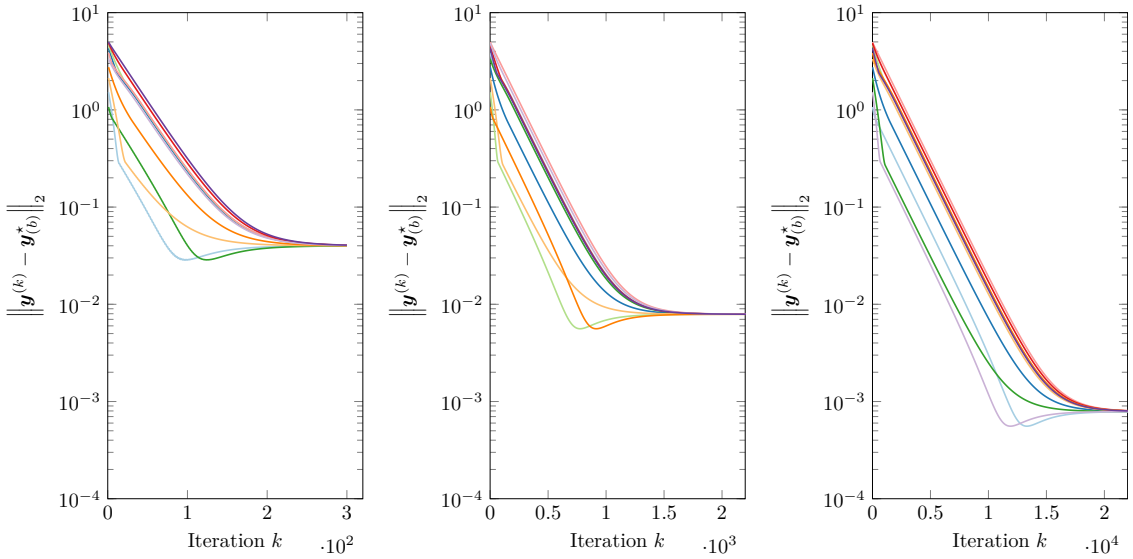
(a) Case (a)

(b) Case (b)

(c) Case (a)

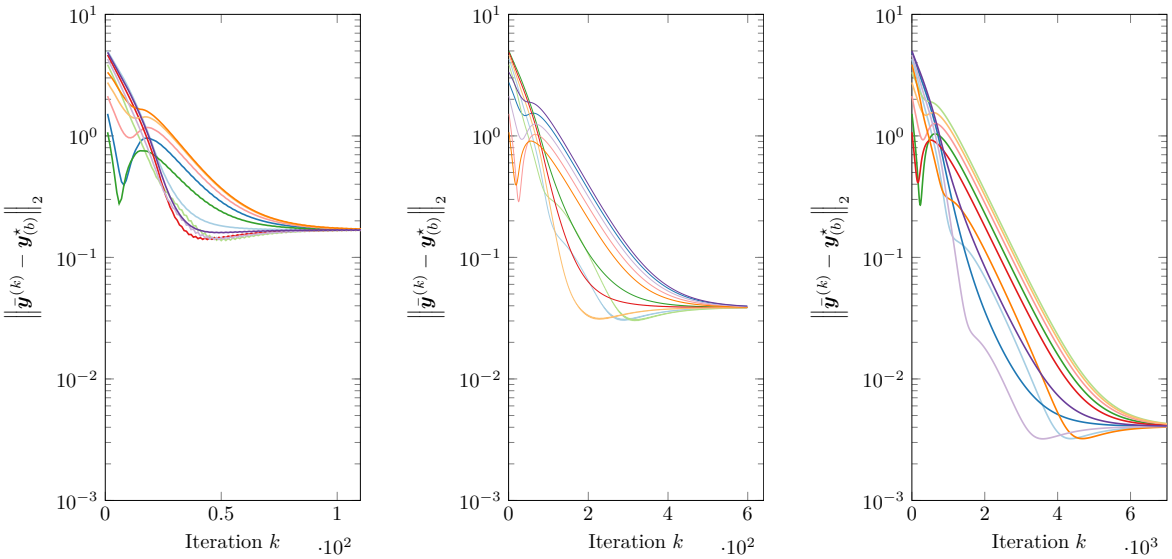
(d) Case (b)

FIGURE 5.8: **SDE method.** Figures 5.8a and 5.8b: Trajectories of the iterates $\left(\bar{\boldsymbol{y}}^{(k)}\right)_{k\geq 0}$ for different initial points $\boldsymbol{y}^{(0)}$. Figures 5.8c and 5.8d: Convergence of $\left(\bar{\boldsymbol{y}}^{(k)}\right)_{k\geq 0}$ towards the solution $\boldsymbol{y}^\star$ of the GVI.



(a) $\frac{1}{2}\alpha$ (b) $\frac{1}{10}\alpha$ (c) $\frac{1}{100}\alpha$

FIGURE 5.9: **PS method:** Convergence in situation (b) of $\left(\boldsymbol{y}^{(k)}\right)_{k\geq 0}$ towards the solution $\boldsymbol{y}^{\star}_{(b)}$ for different step length α in Algorithm 5.1.



(a) 2β (b) 10β (c) 100β

FIGURE 5.10: **SDE method:** Convergence in situation (b) of $\left(\bar{\boldsymbol{y}}^{(k)}\right)_{k\geq 0}$ towards the solution $\boldsymbol{y}^{\star}_{(b)}$ for different parameter values of β in Algorithm 5.3.

and $\left\{\boldsymbol{u}^{(i)}\right\}_{i\in[\![0,k]\!]}$ with $\boldsymbol{u}^{(i)} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(i)}\right)$ for all $i \in [\![0,k]\!]$, we can restrict the localisation set to $\mathcal{L}^{(k)}$, *i.e.*

$$\mathcal{L}^{(k)} \coloneqq \mathbf{K} \cap \left\{\boldsymbol{z} \in \mathbb{R}^d \,\middle|\, \forall i \in [\![0,k]\!] : \left\langle \boldsymbol{u}^{(i)}, \boldsymbol{z} - \boldsymbol{y}^{(i)} \right\rangle \leq 0\right\}. \tag{5.106}$$

In Figure 5.11, we illustrate the iterative shrinkage of the localisation set as the number of iteration grows for the function $\boldsymbol{x} \mapsto \max(2x_1 + x_2^2, x_1^2 + 2x_2^2)$. The idea of restricting the localisation set by iteratively adding a half-space constraint to get closer to the optimal solution is the central idea of the cutting plane scheme in convex optimisation (Nesterov, 2018), which is transposed as follows for (general) variational inequalities:

Algorithm 5.4 (General cutting plane scheme for GVIs).

Input: *Choose* $\boldsymbol{y}^{(0)} \in \mathbf{K}$.

Initialisation: *Set* $\mathcal{L}^{(-1)} = \mathbf{K}$.

Iteration $(k \geq 0)$**:**

$\boldsymbol{u}^{(k)} \in_R \boldsymbol{\Xi}\left(\boldsymbol{y}^{(k)}\right)$,

$\mathcal{L}^{(k)} = \mathcal{L}^{(k-1)} \cap \left\{\boldsymbol{z} \,\middle|\, \left\langle \boldsymbol{u}^{(k)}, \boldsymbol{z} - \boldsymbol{y}^{(k)} \right\rangle \leq 0\right\}$,

Find next iterate $\boldsymbol{y}^{(k+1)} \in \mathcal{L}^{(k)}$.

Output: $\boldsymbol{y}^{(k+1)}$.

The critical step of a cutting plane method depends on how the next point $\boldsymbol{y}^{(k+1)}$ is generated inside the current localisation set $\mathcal{L}^{(k)}$. Several methods are suggested in the literature such as the *ellipsoid method*, the *method of centres of gravity*, the *inscribed ellipsoid method*, or the *analytic centre method*, which are all discussed in Nesterov (2018).

5.4.6 Partial projected subgradient-type method with cutting planes (PPS-CP)

In this section, we present a cutting method that relies on a quadratic optimisation problem to find the next iterate. Given the parameters values $\alpha > 0$ and $0 < \kappa_- \leq \kappa_+ \leq 1$, we define the next point $\boldsymbol{y}^{(k+1)}$ as the randomised partial step from $\boldsymbol{y}^{(k)}$ towards the maximizer of $\psi^{1/\alpha}_{\boldsymbol{y}^{(k)},\boldsymbol{u}^{(k)}}$, *i.e.*

$$\boldsymbol{y}_\star^{(k+1)} \coloneqq \arg\min_{\boldsymbol{z}\in\mathcal{L}^{(k)}} \psi^{1/\alpha}_{\boldsymbol{y}^{(k)},\boldsymbol{u}^{(k)}}(\boldsymbol{z}), \tag{5.107a}$$

$$\boldsymbol{y}^{(k+1)} \coloneqq (1-\kappa)\,\boldsymbol{y}^{(k)} + \kappa \boldsymbol{y}_\star^{(k+1)}, \tag{5.107b}$$

with κ chosen uniformly randomly from the set $[\kappa_-, \kappa_+]$, *i.e.* $\kappa \in_R [\kappa_-, \kappa_+]$. Since we have assumed that the set $\mathbf{K}$ is compact, (5.107a) is well defined and $\boldsymbol{y}_\star^{(k+1)}$ is the unique maximizer of $\psi^{1/\alpha}_{\boldsymbol{y}^{(k)},\boldsymbol{u}^{(k)}}$. With Proposition C.2, (5.107a) is equivalent to

$$\boldsymbol{y}_\star^{(k+1)} = \Pi_{\mathcal{L}^{(k)}}\left(\boldsymbol{y}^{(k)} - \alpha \boldsymbol{u}^{(k)}\right). \tag{5.108}$$

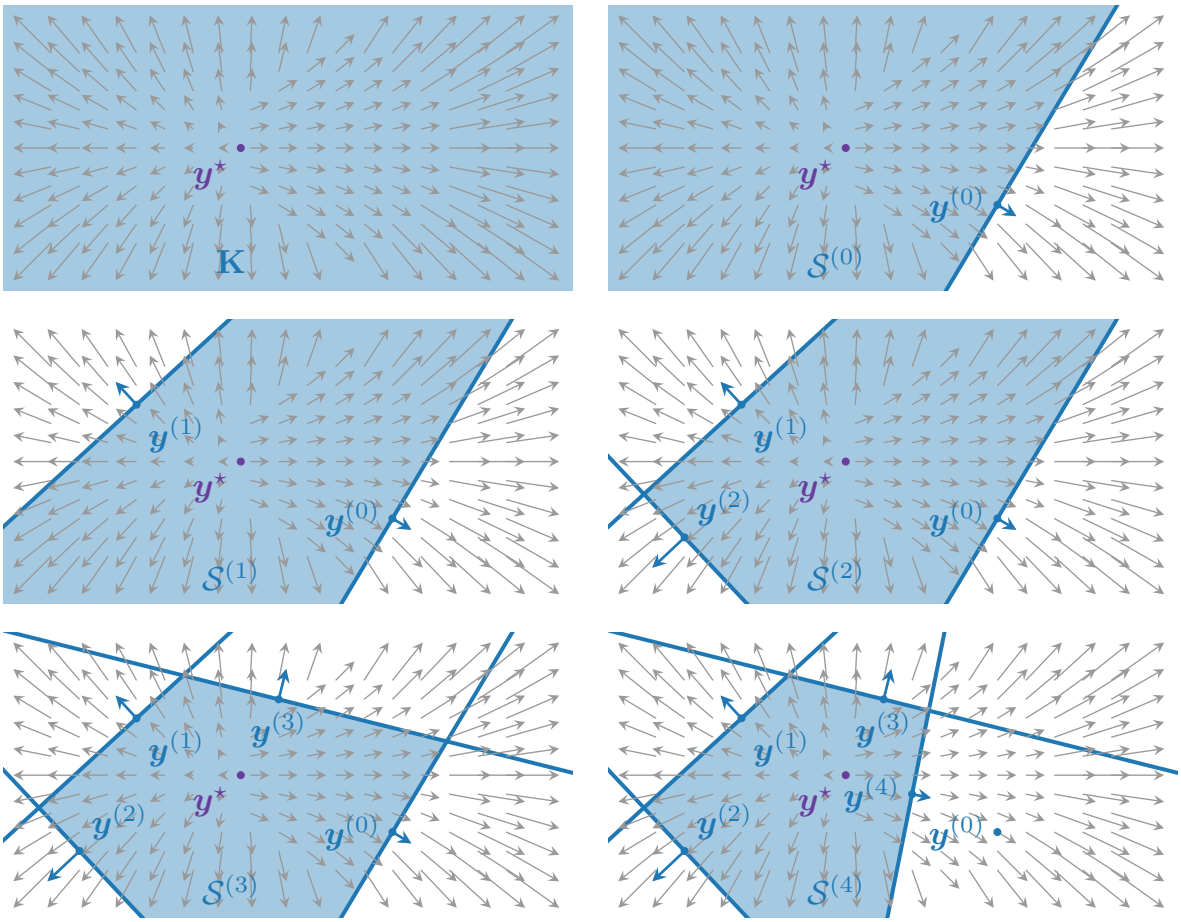


FIGURE 5.11: Illustration of the iterative shrinkage of the localisation set for a method that would generate the sequence of iterates $\left\{\boldsymbol{y}^{(k)}\right\}_{k\in[\![0,4]\!]}$. The vector field represented corresponds to a random element of the subdifferential of the function $\boldsymbol{x} \mapsto \max(2x_1 + x_2^2,\, x_1^2 + 2x_2^2)$.

Hence, if $\kappa_- = \kappa_+ = 1$, we recover the projected subgradient-type method with the difference that we take the orthogonal projection on the localisation set $\mathcal{L}^{(k)}$ rather than $\mathbf{K}$. This suggests the following projected subgradient-type method with cutting planes and randomised partial step to solve a GVI:

ALGORITHM 5.5 (PARTIAL PROJECTED SUBGRADIENT-TYPE METHOD WITH CUTTING PLANES (PPS-CP)).

Input: *Choose* $\boldsymbol{y}^{(0)} \in \mathbf{K}$, $0 < \kappa_- < \kappa_+ \leq 1$ *and* $\alpha > 0$.

Initialisation: *Set* $\mathcal{L}^{(-1)} = \mathbf{K}$.

Iteration $(k \geq 0)$**:**

$$\begin{aligned}
&\boldsymbol{u}^{(k)} \in_R \boldsymbol{\Xi}\Big(\boldsymbol{y}^{(k)}\Big),\\
&\mathcal{L}^{(k)} = \mathcal{L}^{(k-1)} \cap \Big\{\boldsymbol{z} \,\Big|\, \Big\langle \boldsymbol{u}^{(k)}\,,\, \boldsymbol{z} - \boldsymbol{y}^{(k)} \Big\rangle \leq 0\Big\},\\
&\boldsymbol{y}_\star^{(k+1)} = \Pi_{\mathcal{L}^{(k)}}\Big(\boldsymbol{y}^{(k)} - \alpha \boldsymbol{u}^{(k)}\Big),\\
&\kappa \in_R [\kappa_-, \kappa_+],\\
&\boldsymbol{y}^{(k+1)} = (1-\kappa)\, \boldsymbol{y}^{(k)} + \kappa \boldsymbol{y}_\star^{(k+1)}.
\end{aligned}$$

Output: $\boldsymbol{y}^{(k+1)}$.

One way to prove the convergence of a cutting plane method, is to prove that the volume of the localisation sets $\mathcal{L}^{(k)}$ are decreasing and tending to 0 as k grows. In the next lemma, we prove that the volume of the localisation sets generated by Algorithm 5.5 is decreasing almost surely with k:

LEMMA 5.7. *If* $0 < \kappa_- < \kappa_+ \leq 1$, *then for any point* $\boldsymbol{y}^{(k)} \in \mathcal{L}^{(k)}$, *we have that either* $\boldsymbol{y}^{(k)} = \boldsymbol{y}^\star$, *either the volume of the localisation set decreases almost surely,* i.e.

$$\mathbf{vol}\Big(\mathcal{L}^{(k+1)}\Big) \overset{\text{a.s.}}{<} \mathbf{vol}\Big(\mathcal{L}^{(k)}\Big), \tag{5.109}$$

where $\mathcal{L}^{(k+1)} = \mathcal{L}^{(k)} \cap \Big\{\boldsymbol{z} \,\Big|\, \Big\langle \boldsymbol{u}^{(k+1)}\,,\, \boldsymbol{z} - \boldsymbol{y}^{(k+1)} \Big\rangle \leq 0\Big\}$ *with* $\boldsymbol{y}^{(k+1)}$ *defined as in* (5.107b) *and* $\boldsymbol{u}^{(k+1)} \in_R \boldsymbol{\Xi}\Big(\boldsymbol{y}^{(k+1)}\Big)$.

PROOF. If $\Big\|\boldsymbol{y}_\star^{(k+1)} - \boldsymbol{y}^{(k)}\Big\|_2 = 0$, then $\mathbf{0} \in (\boldsymbol{\Xi} + \mathcal{N}_{\mathcal{L}^{(k)}})\Big(\boldsymbol{y}^{(k)}\Big)$, which implies that $\boldsymbol{y}^{(k)} = \boldsymbol{y}^\star$. We now assume that $\Big\|\boldsymbol{y}_\star^{(k+1)} - \boldsymbol{y}^{(k)}\Big\|_2 > 0$. By the strict monotonicity of the set-valued map, there is at most one point $\tilde{\boldsymbol{y}}$ on the line segment $\Big[\boldsymbol{y}^{(k)}, \boldsymbol{y}_\star^{(k+1)}\Big]$ such that $\Big\langle \boldsymbol{y}^{(k)} - \boldsymbol{y}_\star^{(k+1)}\,,\, \tilde{\boldsymbol{u}}\Big\rangle = 0$ with $\tilde{\boldsymbol{u}} \in \boldsymbol{\Xi}(\tilde{\boldsymbol{y}})$. Hence, $\Big\langle \boldsymbol{y}^{(k)} - \boldsymbol{y}^{(k+1)}\,,\, \boldsymbol{u}^{(k+1)}\Big\rangle \neq 0$ almost surely, which implies that we have almost surely that either $\boldsymbol{y}^{(k)} \notin \mathcal{L}^{(k+1)}$ or $\boldsymbol{y}_\star^{(k+1)} \notin \mathcal{L}^{(k+1)}$. Hence

$$\mathbf{vol}\Big(\mathcal{L}^{(k)} \cap \Big\{\boldsymbol{z} \,\Big|\, \Big\langle \boldsymbol{u}^{(k+1)}\,,\, \boldsymbol{z} - \boldsymbol{y}^{(k+1)} \Big\rangle > 0\Big\}\Big) > 0 \tag{5.110}$$

almost surely, which proves (5.109). □

Lemma 5.7 constitutes only a partial result since we do not provide any guarantee on the rate of decrease of the volume of the localisation set nor that the volume of the localisation set

convergences to 0 as k grows. Nevertheless, we show subsequently in Theorem 5.9 that, if the step length is fixed appropriately, then we have a convergence rate guarantee towards neighbourhood of $\boldsymbol{y}^\star$ that is similar to that for the PS method (*i.e.* Algorithm 5.1 and Theorem 5.3). Additionally, numerical simulations provide evidence that the method does converge to $\boldsymbol{y}^\star$ but we do not provide any convergence rate guarantee.

THEOREM 5.9. *Suppose that Assumption 5.1 holds. Then, for a given $\delta > 0$, there exists a fixed step length $\alpha > 0$ such that the sequence $(\boldsymbol{y}_k)_{k\geq 0}$ generated by Algorithm 5.5 converges linearly to the ball of radius $\delta + \frac{\mu M_\delta^\star}{L_\delta^2}$ around $\boldsymbol{y}^\star$,* i.e. $\mathcal{B}\left[\boldsymbol{y}^\star\,;\,\delta + \frac{\mu M_\delta^\star}{L_\delta^2}\right]$. *It then stays in that neighbourhood. Moreover, the step length ensuring the fastest convergence is given by $\alpha_\delta^* \coloneqq \frac{\mu}{L_\delta^2}$.*

PROOF. The proof follows the main lines of Theorem 5.3. By definition we have that $\kappa \in [\kappa_-, \kappa_+]$. If we use the same reasoning that leads to (5.43b) in Theorem 5.3, we obtain:

$$\left\|\Pi_{\mathcal{L}^{(k)}}\left(\boldsymbol{y}^{(k)} - \alpha\boldsymbol{u}^{(k)}\right) - \Pi_{\mathcal{L}^{(k)}}(\boldsymbol{y}^\star - \alpha\boldsymbol{u}^\star)\right\| \leq \left(\sqrt{1 - 2\alpha\mu + \alpha^2 L_\delta^2}\right)\left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\|. \tag{5.111}$$

If $\boldsymbol{y}^{(k)} \notin \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$, we have that:

$$\begin{aligned}
\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^\star\right\| &= \left\|(1-\kappa)\,\boldsymbol{y}^{(k)} + \kappa\,\Pi_{\mathcal{L}^{(k)}}\left(\boldsymbol{y}^{(k)} - \alpha\boldsymbol{u}^{(k)}\right) - \boldsymbol{y}^\star\right\| \\
&= \left\|(1-\kappa)\left(\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right) + \kappa\left(\Pi_{\mathcal{L}^{(k)}}\left(\boldsymbol{y}^{(k)} - \alpha\boldsymbol{u}^{(k)}\right) - \Pi_{\mathcal{L}^{(k)}}(\boldsymbol{y}^\star - \alpha\boldsymbol{u}^\star)\right)\right\| \\
&\leq (1-\kappa)\left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\| + \kappa\left\|\Pi_{\mathcal{L}^{(k)}}\left(\boldsymbol{y}^{(k)} - \alpha\boldsymbol{u}^{(k)}\right) - \Pi_{\mathcal{L}^{(k)}}(\boldsymbol{y}^\star - \alpha\boldsymbol{u}^\star)\right\| \qquad (5.112) \\
&\overset{(5.111)}{\leq} \left((1-\kappa) + \kappa\left(\sqrt{1 - 2\alpha\mu + \alpha^2 L_\delta^2}\right)\right)\left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\|. \qquad (5.113)
\end{aligned}$$

To ensure convergence, we must have that $(1-\kappa) + \kappa\left(\sqrt{1 - 2\alpha\mu + \alpha^2 L_\delta^2}\right) < 1$ and consequently that $(1 - 2\alpha\mu + \alpha^2 L_\delta^2) < 1 \Rightarrow \alpha < \frac{2\mu}{L_\delta^2}$. The step length $\alpha_\delta^* = \frac{\mu}{L_\delta^2}$, which minimises $(1 - 2\alpha\mu + \alpha^2 L_\delta^2)$, leads to the best convergence factor guarantee. Since $\sqrt{1-x} \leq 1 - \frac{x}{2}$ for all $x < 1$, we have that

$$(1-\kappa) + \kappa\left(\sqrt{1 - 2\alpha_\delta^*\mu + \left(\alpha_\delta^*\right)^2 L_\delta^2}\right) = (1-\kappa) + \kappa\left(\sqrt{1 - \frac{1}{\gamma_\delta^2}}\right) \leq 1 - \frac{\kappa}{2\gamma_\delta^2} \leq 1 - \frac{\kappa_-}{2\gamma_\delta^2}, \tag{5.114}$$

where we used the fact that $\kappa \in [\kappa_-, \kappa_+]$. This implies with Inequalities (5.44) and (5.113) that

$$\begin{aligned}
\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^\star\right\| &\leq \left(1 - \frac{\kappa_-}{2\gamma_\delta^2}\right)\left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\| \leq \left(1 - \frac{\kappa_-}{2\gamma_\delta^2}\right)^k \left\|\boldsymbol{y}^{(0)} - \boldsymbol{y}^\star\right\| \\
&\leq \exp\left(\frac{-k\kappa_-}{2\gamma_\delta^2}\right)\left\|\boldsymbol{y}^{(0)} - \boldsymbol{y}^\star\right\|, \qquad (5.115)
\end{aligned}$$

which in turns implies that

$$\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^\star\right\|^2 \leq \exp\left(\frac{-k\kappa_-}{\gamma_\delta^2}\right)\left\|\boldsymbol{y}^{(0)} - \boldsymbol{y}^\star\right\|^2. \tag{5.116}$$

If $\boldsymbol{y}^{(k)} \in \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$, then by hypothesis $\left\|\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right\| \leq M_\delta^\star$, which implies that

$$\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^\star\right\| \overset{(5.112),(5.8)}{\leq} (1-\kappa)\left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\| + \kappa\left\|\left(\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right) - \alpha\left(\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right)\right\|$$

$$\leq \left\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^\star\right\| + \kappa\alpha \left\|\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right\| \leq \delta + \kappa\alpha \left\|\boldsymbol{u}^{(k)} - \boldsymbol{u}^\star\right\|$$
$$\leq \delta + \kappa\alpha M_\delta^\star \leq \delta + \alpha M_\delta^\star. \tag{5.117}$$

Taking α equal to α_δ^* leads to the desired result. □

5.4.7 Cutting plane method and dual extrapolation method

In Section 5.4.5, we mentioned that using a cutting plane scheme can lead to methods that converges towards the solution $\boldsymbol{y}^\star$ of GVI($\mathbf{K}, \boldsymbol{\Xi}$). In this section, we design a method which combines the dual extrapolation and cutting plane methods. The resulting method, that we denote as the *partial dual extrapolation method with cutting planes* (PDE-CP), has the advantage of having a convergence rate to a neighbourhood of the solution $\boldsymbol{y}^\star$ that is similar the SDE method, all the while converging towards $\boldsymbol{y}^\star$ by exploiting its cutting plane scheme. Similarly to the PPS-CP method (see Algorithm 5.5), the PDE-CP method relies on the following two keys features: Firstly, it exploits the strong monotonicity of the set-valued map $\boldsymbol{\Xi}$ to iteratively shrink the localisation set via cutting planes. Secondly, it takes a partial step of randomised relative length when computing the new iterate $\boldsymbol{y}^{(k+1)}$ based on $\left(\boldsymbol{x}^{(k)}, \boldsymbol{v}^{(k)}\right)$. In the following, we show in a first step that altering the SDE method by iteratively restricting the search to the current localisation set does affect its convergence rate. Then, in a second step, we show a similar convergence rate still holds if we define the iterate $\boldsymbol{y}^{(k+1)}$ not as the maximizer $\boldsymbol{y}_\star^{(k+1)}$ of $\psi^\beta_{\boldsymbol{x}^{(k)}, \boldsymbol{v}^{(k)}}$ as done in (5.84c), but rather as a random point on the line between $\boldsymbol{x}^{(k)}$ and $\boldsymbol{y}_\star^{(k+1)}$.

5.4.7.1 SDE method with cutting planes (SDE-CP)

For any sequences of iterates $\left\{\boldsymbol{y}^{(i)}\right\}_{i\in[\![0,k]\!]}$ and $\left\{\boldsymbol{u}^{(i)}\right\}_{i\in[\![0,k]\!]}$ with $\boldsymbol{u}^{(i)} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(i)}\right)$ for all $i \in [\![0, k]\!]$, we have seen in Section 5.4.5 that we could define a localisation set which contains the optimal solution $\boldsymbol{y}^\star$ (see Equation (5.106)). We investigate here the convergence rate of the SDE method when we compute the next iterates $\boldsymbol{x}^{(k)}$ and $\boldsymbol{y}^{(k+1)}$ by imposing that they must belong the set $\boldsymbol{\mathcal{L}}^{(k)}$ instead of $\mathbf{K}$ as previously done in the Optimisation Problems (5.84a) and (5.84c). We would then have the following iterations

$$\boldsymbol{\mathcal{L}}^{(k)} \coloneqq \boldsymbol{\mathcal{L}}^{(k-1)} \cap \left\{\boldsymbol{z} \,\middle|\, \left\langle \boldsymbol{u}^{(k)}\,,\, \boldsymbol{z} - \boldsymbol{y}^{(k)} \right\rangle \leq 0\right\}, \tag{5.118a}$$
$$\boldsymbol{x}^{(k)} \coloneqq \arg\max_{\boldsymbol{z}\in\boldsymbol{\mathcal{L}}^{(k)}} \Psi_k(\boldsymbol{z}), \tag{5.118b}$$
$$\boldsymbol{v}^{(k)} \in_R \boldsymbol{\Xi}\left(\boldsymbol{x}^{(k)}\right), \tag{5.118c}$$
$$\boldsymbol{y}^{(k+1)} \coloneqq \arg\max_{\boldsymbol{z}\in\boldsymbol{\mathcal{L}}^{(k)}} \psi^\beta_{\boldsymbol{x}^{(k)}, \boldsymbol{v}^{(k)}}(\boldsymbol{z}), \tag{5.118d}$$
$$\boldsymbol{u}^{(k+1)} \in_R \boldsymbol{\Xi}\left(\boldsymbol{y}^{(k+1)}\right), \tag{5.118e}$$

where $\psi^\beta_{\boldsymbol{x},\boldsymbol{u}}$ and Ψ_k are defined as in Equations (5.82a) and (5.82b) respectively and $\boldsymbol{\mathcal{L}}^{(0)} = \mathbf{K}$. We now prove that using such a strategy leads to similar convergence guarantees as the SDE method. To do so, we define

$$\Delta_{\boldsymbol{\mathcal{L}}}^{(k)} = \max_{\boldsymbol{z}\in\boldsymbol{\mathcal{L}}^{(k)}} \Psi_k(\boldsymbol{z}), \tag{5.119}$$

and the merit function

$$f_{\mathcal{L}}^{(k)}(\boldsymbol{y}) \coloneqq \max_{\boldsymbol{z}\in\mathcal{L}^{(k)}} \left\{ \max_{\boldsymbol{v}\in\boldsymbol{\Xi}(\boldsymbol{z})} \{\langle \boldsymbol{v}\,,\, \boldsymbol{y}-\boldsymbol{z}\rangle\} + \frac{1}{2}\mu \left\|\boldsymbol{y}-\boldsymbol{z}\right\|^2 \right\}. \tag{5.120}$$

The difference of $\Delta_{\mathcal{L}}^{(k)}$ with $\Delta_{\mathbf{K}}^{(k)}$, and $f_{\mathcal{L}}^{(k)}$ with $f_{\mathbf{K}}^{(k)}$ is that the optimisation problem is on the localisation set $\mathcal{L}^{(k)}$ instead of $\mathbf{K}$.

LEMMA 5.8. *It holds that*

$$f_{\mathcal{L}}^{(k)}\left(\bar{\boldsymbol{y}}^{(k)}\right) \leq \frac{1}{S^{(k)}}\Delta_{\mathcal{L}}^{(k)}. \tag{5.121}$$

PROOF. The proof follows the same steps as for Lemmas 5.4 and 5.5. □

THEOREM 5.10. *If* $\alpha^{(k+1)} \leq \frac{\mu}{\beta}S^{(k)}$ *then*

$$\Delta_{\mathcal{L}}^{(k+1)} \leq \Delta_{\mathcal{L}}^{(k)} + \frac{1}{2}\alpha^{(k+1)}\left[\frac{1}{\mu+\beta}\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|^2 - \beta\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2\right]. \tag{5.122}$$

PROOF. The same steps as for the proof of Theorems 5.5 and 5.7 can be followed. We first observe that $\Psi_k\left(\boldsymbol{x}^{(k)}\right) = \Delta_{\mathcal{L}}^{(k)}$ based on (5.118b) and the definition of $\Delta_{\mathcal{L}}^{(k)}$, and that $\Psi_{k+1}(\boldsymbol{z}) = \Psi_k(\boldsymbol{z}) + \alpha^{(k+1)}\psi^{\mu}_{\boldsymbol{y}^{(k+1)},\boldsymbol{u}^{(k+1)}}(\boldsymbol{z})$. We can consequently bound $\Delta_{\mathcal{L}}^{(k+1)}$ as follows:

$$\begin{aligned}
\Delta_{\mathcal{L}}^{(k+1)} &\overset{(5.119)}{=} \max_{\boldsymbol{z}\in\mathcal{L}^{(k+1)}} \left\{\Psi_k(\boldsymbol{z}) + \alpha^{(k+1)}\psi^{\mu}_{\boldsymbol{y}^{(k+1)},\boldsymbol{u}^{(k+1)}}(\boldsymbol{z})\right\}, && \text{(5.123a)}\\
&\leq \max_{\boldsymbol{z}\in\mathcal{L}^{(k)}} \left\{\Psi_k(\boldsymbol{z}) + \alpha^{(k+1)}\psi^{\mu}_{\boldsymbol{y}^{(k+1)},\boldsymbol{u}^{(k+1)}}(\boldsymbol{z})\right\}, && \text{(5.123b)}\\
&\overset{(5.64)}{\leq} \max_{\boldsymbol{z}\in\mathcal{L}^{(k)}} \left\{\Delta_{\mathcal{L}}^{(k)} - \frac{1}{2}\mu S^{(k)}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2 + \alpha^{(k+1)}\psi^{\mu}_{\boldsymbol{y}^{(k+1)},\boldsymbol{u}^{(k+1)}}(\boldsymbol{z})\right\}, && \text{(5.123c)}\\
&\overset{(5.82a)}{=} \Delta_{\mathcal{L}}^{(k)} + \max_{\boldsymbol{z}\in\mathcal{L}^{(k)}} \Bigg\{ -\frac{1}{2}\mu S^{(k)}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2 \\
&\qquad\qquad +\alpha^{(k+1)}\left(\left\langle \boldsymbol{u}^{(k+1)}\,,\, \boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2\right)\Bigg\}, && \text{(5.123d)}
\end{aligned}$$

where the first inequality holds since $\mathcal{L}^{(k+1)} \subseteq \mathcal{L}^{(k)}$, and the second equality arises from the strong concavity of Ψ_k with parameter $\mu S^{(k)}$ and the fact that $\boldsymbol{x}^{(k)}$ maximises Ψ_k over $\mathcal{L}^{(k)}$. Then, as $\boldsymbol{y}^{(k+1)}$ maximises $\psi^{\beta}_{\boldsymbol{x}^{(k)},\boldsymbol{v}^{(k)}}$ on $\mathcal{L}^{(k)}$, we have that

$$\left\langle -\boldsymbol{v}^{(k)} - \beta\left(\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right)\,,\, \boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\rangle \leq 0, \quad \forall \boldsymbol{z}\in\mathcal{L}^{(k)}. \tag{5.124}$$

The remainder of the proof follows the same steps as for Theorem 5.5. □

COROLLARY 5.4. *Assume that* $\boldsymbol{\Xi}$ *is* μ*-strongly monotone and* δ*-far-sighted Lipschitz continuous with constant* L_δ *on* $\mathbf{K}$ *for a given* $\delta > 0$*. With the choice*

$$\beta = L_\delta, \quad \alpha^{(k+1)} = \frac{1}{\gamma_\delta}S^{(k)}, \tag{5.125}$$

where $\gamma_\delta = \frac{L_\delta}{\mu}$ *is the condition number of* $\boldsymbol{\Xi}$ *related to* δ*, we obtain that* $\Delta_{\mathcal{L}}^{(k+1)} \leq \Delta_{\mathcal{L}}^{(k)}$ *while* $\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\| \geq \delta$.

PROOF. The proof is identical to that of Corollary 5.2. □

THEOREM 5.11. *Let $\boldsymbol{\Xi}$ be a μ-strongly monotone and δ-far-sighted Lipschitz continuous with parameter L_δ on $\mathbf{K}$. Let $\left\{\bar{\boldsymbol{y}}^{(k)}\right\}_{k\in[\![0,\bar{k}]\!]}$ be the sequence generated by (5.118), where $\bar{k}$ is the first iteration number such that $\left\|\boldsymbol{y}^{(\bar{k}+1)}-\boldsymbol{x}^{(\bar{k})}\right\| < \delta$ holds true. If $\left\|\boldsymbol{y}^{(0)}-\boldsymbol{y}^\star\right\| > \delta$, then for any $k \in [\![0,\bar{k}]\!]$:*

$$\begin{aligned}\frac{1}{2}\mu\left\|\bar{\boldsymbol{y}}^{(k)}-\boldsymbol{y}^\star\right\|^2 \leq f_{\mathcal{L}}^{(k)}\left(\bar{\boldsymbol{y}}^{(k)}\right) &\leq \left[f_{\mathbf{K}}\left(\boldsymbol{y}^{(0)}\right)+\frac{\mu(\gamma_\delta^2-1)}{2}\left\|\boldsymbol{y}^{(0)}-\boldsymbol{y}^\star\right\|^2\right]\exp\left(-\frac{k}{\gamma_\delta+1}\right)\\ &\leq f_{\mathbf{K}}\left(\boldsymbol{y}^{(0)}\right)\cdot\gamma_\delta^2\cdot\exp\left(-\frac{k}{\gamma_\delta+1}\right).\end{aligned} \tag{5.126}$$

PROOF. For any $k \in [\![0,\bar{k}]\!]$, the solution $\boldsymbol{y}^\star$ of GVI$(\mathbf{K},\boldsymbol{\Xi})$ belongs to the localisation set $\mathcal{L}^{(k)}$, which implies that $\boldsymbol{y}^\star$ is also the solution of GVI$\left(\mathcal{L}^{(k)},\boldsymbol{\Xi}\right)$, and consequently that $f_{\mathcal{L}}^{(k)}(\boldsymbol{y}^\star)=0$. Therefore, by the strong convexity of the function $f_{\mathcal{L}}^{(k)}$ with parameter μ, we have that

$$\frac{1}{2}\mu\left\|\bar{\boldsymbol{y}}^{(k)}-\boldsymbol{y}^\star\right\|^2 \overset{(5.64)}{\leq} f_{\mathcal{L}}^{(k)}\left(\bar{\boldsymbol{y}}^{(k)}\right). \tag{5.127}$$

In view of Corollary 5.4, $\Delta_{\mathcal{L}}^{(k)} \leq \Delta_{\mathcal{L}}^{(k-1)}$ for any $k \in [\![1,\bar{k}]\!]$. Together with Lemma 5.8, this leads to

$$f_{\mathcal{L}}^{(k)}\left(\bar{\boldsymbol{y}}^{(k)}\right) \overset{(5.121)}{\leq} \frac{1}{S^{(k)}}\Delta_{\mathcal{L}}^{(k)} \leq \Delta_{\mathcal{L}}^{(0)}\left(1-\frac{1}{\gamma_\delta+1}\right)^k \leq \Delta_{\mathcal{L}}^{(0)}\exp\left(-\frac{k}{\gamma_\delta+1}\right). \tag{5.128}$$

The rest of the proof follows the one of Theorem 5.6 after noticing that $\Delta_{\mathcal{L}}^{(0)}=\Delta_{\mathbf{K}}^{(0)}$. □

We have just proven that a convergence guarantee is obtained similar to that for the SDE method when, at each iteration $k \in [\![0,\bar{k}]\!]$, we replace the feasible set $\mathbf{K}$ be the localisation set $\mathcal{L}^{(k)}$ when we compute the next iterates $\boldsymbol{x}^{(k)}$ and $\boldsymbol{y}^{(k+1)}$.

5.4.7.2 Partial dual extrapolation method with cutting planes (PDE-CP)

Using cutting planes as suggested in the SDE-CP is the first part to prevent the sequence of iterates to oscillate around $\boldsymbol{y}^\star$ as depicted in the numerical results (see Section 5.4.4). However, in the current formulation of the SDE-CP method, the volume of the localisation set does not necessarily decrease with k, which is required to have a convergent method. Hence, in the same vein as for Algorithm 5.5, we study the impact of including a partial step of randomised length in the SDE or SDE-CP methods. In these methods, the iterate $\boldsymbol{y}^{(k+1)}$ is defined as the maximizer of $\psi^\beta_{\boldsymbol{x}^{(k)},\boldsymbol{v}^{(k)}}$ (see Equations (5.84c) and (5.118d) respectively). We investigate here the convergence results when defining $\boldsymbol{y}^{(k+1)}$ as the point resulting from a partial step of randomised relative length $\kappa \in_R [\kappa_-,\kappa_+]$ with $\frac{1}{2} < \kappa_- \leq \kappa_+ \leq 1$ from $\boldsymbol{x}^{(k)}$ towards the maximizer $\boldsymbol{y}_\star^{(k+1)}$ of $\psi^\beta_{\boldsymbol{x}^{(k)},\boldsymbol{v}^{(k)}}$. The partial step with randomised relative length is illustrated in Figure 5.12.

In the following, we prove the convergence of the SDE method when taking a partial step when combined with cutting planes. We should mention that the proofs provided subsequently also hold when the cutting planes are not added since the arguments remain valid when substituting the localisation sets $\left\{\mathcal{L}^{(k)}\right\}_{k\geq 0}$ with $\mathbf{K}$.

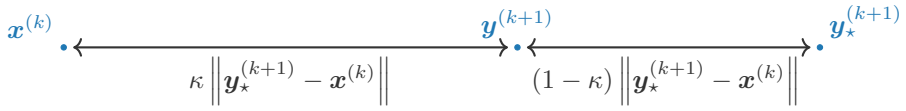


FIGURE 5.12: Illustration of the partial step in the dual extrapolation method.

Given a sequence of previous iterates $\left\{\boldsymbol{y}^{(i)}\right\}_{i=0}^{k}$, the SDE method with a randomised step and cutting planes is mathematically translated as follows:

$$\mathcal{L}^{(k)} := \mathcal{L}^{(k-1)} \cap \left\{ \boldsymbol{z} \,\middle|\, \left\langle \boldsymbol{u}^{(k)}, \boldsymbol{z} - \boldsymbol{y}^{(k)} \right\rangle \leq 0 \right\}, \tag{5.129a}$$

$$\boldsymbol{x}^{(k)} := \arg\max_{\boldsymbol{z} \in \mathcal{L}^{(k)}} \Psi_k(\boldsymbol{z}), \tag{5.129b}$$

$$\boldsymbol{v}^{(k)} \in_R \boldsymbol{\Xi}\left(\boldsymbol{x}^{(k)}\right), \tag{5.129c}$$

$$\boldsymbol{y}_\star^{(k+1)} := \arg\max_{\boldsymbol{z} \in \mathcal{L}^{(k)}} \psi^{\beta}_{\boldsymbol{x}^{(k)}, \boldsymbol{v}^{(k)}}(\boldsymbol{z}), \tag{5.129d}$$

$$\kappa \in_R [\kappa_-, \kappa_+], \tag{5.129e}$$

$$\boldsymbol{y}^{(k+1)} := \kappa \boldsymbol{y}_\star^{(k+1)} + (1-\kappa)\, \boldsymbol{x}^{(k)}, \tag{5.129f}$$

$$\boldsymbol{u}^{(k+1)} \in_R \boldsymbol{\Xi}\left(\boldsymbol{y}^{(k+1)}\right), \tag{5.129g}$$

where $\psi^{\beta}_{\boldsymbol{x},\boldsymbol{u}}$ and Ψ_k are defined as in Equations (5.82a) and (5.82b) respectively and $\mathcal{L}^{(0)} = \mathbf{K}$. For $\kappa_- = \kappa_+ = 1$, we recover the iterations defined in the SDE method with cutting planes (SDE-CP) defined in (5.118). We show now that a similar convergence guarantee towards a neighbourhood of $\boldsymbol{y}^\star$ exists when taking a partial step with randomised relative length. First, it is clear that Lemma 5.8 continues to hold. We now show a similar result as Theorem 5.10:

THEOREM 5.12. *If* $\alpha^{(k+1)} \leq \frac{\mu}{\beta} S^{(k)}$ *then*

$$\Delta_{\mathcal{L}}^{(k+1)} \leq \Delta_{\mathcal{L}}^{(k)} + \frac{1}{2}\alpha^{(k+1)} \left[\frac{1}{\mu+\beta} \left\| \boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)} \right\|^2 + \frac{2\beta}{\mu+\beta} \frac{(1-\kappa)}{\kappa} \left\| \boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)} \right\| \left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\| \right.$$
$$\left. + \frac{\beta^2}{\mu+\beta} \frac{(1-\kappa)^2}{\kappa^2} \left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|^2 - \beta \left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|^2 \right], \tag{5.130}$$

where $\Delta_{\mathcal{L}}^{(k)} = \max_{\boldsymbol{z} \in \mathcal{L}^{(k)}} \Psi_k(\boldsymbol{z})$.

PROOF. The proof is provided in Appendix C.2 at page 190. □

COROLLARY 5.5. *Assume that* $\boldsymbol{\Xi}$ *is* μ*-strongly monotone and* δ*-far-sighted Lipschitz continuous with constant* L_δ *on* $\mathbf{K}$ *for a given* $\delta > 0$. *If* $0 < \kappa_- < \kappa_+ \leq 1$, *then the choice*

$$\beta = \frac{\kappa_-}{2\kappa_- - 1} L_\delta, \quad \alpha^{(k+1)} = \frac{2\kappa_- - 1}{\kappa_- \gamma_\delta} S^{(k)}, \tag{5.131}$$

implies that $\Delta_{\mathcal{L}}^{(k+1)} \leq \Delta_{\mathcal{L}}^{(k)}$ *while* $\left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\| \geq \delta$.

PROOF. First

$$\alpha^{(k+1)} = \frac{2\kappa_- - 1}{\kappa_- \gamma_\delta} S^{(k)} = \frac{(2\kappa_- - 1)\,\mu}{\kappa_- L_\delta} S^{(k)} = \frac{\mu}{\beta} S^{(k)}, \tag{5.132}$$

which implies that Theorem 5.12 holds. By assumption $\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right\| \geq \delta$, we can consequently use the δ-far-sighted Lipschitz continuity in (5.130), which gives us

$$\Delta_{\mathcal{L}}^{(k+1)} \leq \Delta_{\mathcal{L}}^{(k)} + \frac{1}{2}\alpha^{(k+1)}\left[\frac{1}{\mu+\beta}\left(L_\delta^2 + 2\beta L_\delta \frac{(1-\kappa)}{\kappa} + \beta^2\frac{(1-\kappa)^2}{\kappa^2}\right) - \beta\right]\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right\|^2 \tag{5.133a}$$

$$\leq \Delta_{\mathcal{L}}^{(k)} + \frac{1}{2}\alpha^{(k+1)}\left[\frac{1}{\beta}\left(L_\delta^2 + 2\beta L_\delta \frac{(1-\kappa)}{\kappa} + \beta^2\frac{(1-\kappa)^2}{\kappa^2}\right) - \beta\right]\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right\|^2, \tag{5.133b}$$

where the second inequality holds since $\mu + \beta > \beta$. If we assume now that $\beta = cL_\delta$ with $c > 0$, (5.133b) becomes

$$\Delta_{\mathcal{L}}^{(k+1)} \leq \Delta_{\mathcal{L}}^{(k)} + \frac{L_\delta}{2c}\alpha^{(k+1)}\left(1 + c\left(\frac{2(1-\kappa)}{\kappa}\right) + c^2\left(\frac{1-2\kappa}{\kappa^2}\right)\right)\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right\|^2. \tag{5.134}$$

For (5.134) to hold, we must find the range of c such that the polynomial $p(c) \coloneqq 1 + c\left(\frac{2(1-\kappa)}{\kappa}\right) + c^2\left(\frac{1-2\kappa}{\kappa^2}\right)$ is less than or equal to 0. If $\kappa = \frac{1}{2}$, then for all $c > 0$, we have that $p(c) > 0$. If $\kappa \neq \frac{1}{2}$, the second order polynomial p has two roots located at $c_1 = -\kappa < 0$ and $c_2 = \frac{\kappa}{2\kappa-1}$. If $\kappa \in \left(0, \frac{1}{2}\right)$ then $c_2 < c_1 < 0$ and $p \leq 0$ for $c \in [c_2, c_1]$, which are not acceptable values since $c > 0$. If $\kappa \in \left(\frac{1}{2}, 1\right]$ then $c_2 > 0$ and $p \leq 0$ for any value $c \geq c_2$. We must thus set a value β such that $\beta \geq c_2 L_\delta = \frac{\kappa}{2\kappa-1}L_\delta$ for any possible value κ. Hence, since $\frac{\kappa}{2\kappa-1}$ is decreasing for with κ on $\left(\frac{1}{2}, 1\right]$, we must have that $\beta \geq \frac{\kappa_-}{2\kappa_- - 1}L_\delta$. This completes the proof. □

Based on Corollary 5.5, we designed the *partial dual extrapolation method with cutting planes* (PDE-CP) to solve strongly monotone generalised variational inequalities (see Algorithm 5.6).

ALGORITHM 5.6 (PARTIAL DUAL EXTRAPOLATION METHOD WITH CUTTING PLANES (PDE-CP)).

Input: *Choose* $\boldsymbol{y}^{(0)} \in \mathbf{K}$, $0 < \kappa_- < \kappa_+ \leq 1$ *and* $\beta > 0$.

Initialisation: *Set* $\mathcal{L}^{(-1)} = \mathbf{K}$, $\alpha^{(0)} = 1$, *and* $\gamma = \frac{\beta}{\mu}$.

Iteration ($k \geq 0$)**:**

$$\begin{aligned}
&S^{(k)} = \textstyle\sum_{i=0}^{k} \alpha^{(i)} \text{ and } \alpha^{(k+1)} = \frac{1}{\gamma}S^{(k)},\\
&\boldsymbol{u}^{(k)} \in_R \boldsymbol{\Xi}\left(\boldsymbol{y}^{(k)}\right),\\
&\mathcal{L}^{(k)} = \mathcal{L}^{(k-1)} \cap \left\{\boldsymbol{z} \,\middle|\, \left\langle \boldsymbol{u}^{(k)}, \boldsymbol{z} - \boldsymbol{y}^{(k)}\right\rangle \leq 0\right\},\\
&\boldsymbol{y}_\star^{(k+1)} = \arg\max_{\boldsymbol{z}\in\mathcal{L}^{(k)}} \psi^{\beta}_{\boldsymbol{x}^{(k)},\boldsymbol{v}^{(k)}}(\boldsymbol{z}),\\
&\kappa \in_R [\kappa_-, \kappa_+],\\
&\boldsymbol{y}^{(k+1)} = (1-\kappa)\,\boldsymbol{x}^{(k)} + \kappa \boldsymbol{y}_\star^{(k+1)}.
\end{aligned}$$

Output: $\bar{\boldsymbol{y}}^{(k)} = \frac{1}{S^{(k)}}\sum_{i=0}^{k}\alpha^{(i)}\boldsymbol{y}^{(i)}$.

This method generalises the dual extrapolation method of Nesterov and Scrimali (2011) for variational inequalities. Indeed, it accepts the SDE method as special case for the choice $\kappa_- = \kappa_+ = 1$ and when no cutting planes are added. More specifically, Theorem 5.7 and Corollary 5.3 follow from Theorem 5.12 and Corollary 5.5 when substituting κ and κ by 1 and $\mathcal{L}$ by $\mathbf{K}$. As shown subsequently in Theorem 5.13, the PDE-CP method benefits from similar convergence guarantees as the SDE method.

THEOREM 5.13. *Let $\boldsymbol{\Xi}$ be a μ-strongly monotone and δ-far-sighted Lipschitz continuous with parameter L_δ on $\mathbf{K}$. Let $\left\{\bar{\boldsymbol{y}}^{(k)}\right\}_{k\in[\![0,\bar{k}]\!]}$ be the sequence generated by Algorithm 5.6, where $\bar{k}$ is the first iteration number such that $\left\|\boldsymbol{y}^{(\bar{k}+1)}-\boldsymbol{x}^{(\bar{k})}\right\|<\delta$ holds true. If $\left\|\boldsymbol{y}^{(0)}-\boldsymbol{y}^\star\right\|>\delta$, then for any $k\in[\![0,\bar{k}]\!]$:*

$$\begin{aligned}\frac{1}{2}\mu\left\|\bar{\boldsymbol{y}}^{(k)}-\boldsymbol{y}^\star\right\|^2 \leq f_{\mathcal{L}}^{(k)}\left(\bar{\boldsymbol{y}}^{(k)}\right) &\leq \left[f_{\mathbf{K}}\left(\boldsymbol{y}^{(0)}\right)+\frac{\mu(\gamma_\delta^2-1)}{2}\left\|\boldsymbol{y}^{(0)}-\boldsymbol{y}^\star\right\|^2\right]\exp\left(-\frac{k}{\gamma_\delta+1}\right)\\ &\leq f_{\mathbf{K}}\left(\boldsymbol{y}^{(0)}\right)\cdot\gamma_\delta^2\cdot\exp\left(-\frac{k}{\gamma_\delta+1}\right).\end{aligned} \tag{5.135}$$

PROOF. The proof can be shown using the same steps as for Theorem 5.11. □

5.4.8 Numerical results (Part II)

We present now the numerical results related to the PPS-CP, SDE-CP, PDE and PDE-CP methods when applied to the GVI problem formulated in Section 5.4.4. For all the methods, we use the following parameter values: $\alpha=\mu\beta^{-2}=0.056$, $\beta=2.618$, $\kappa_-=0.6$ and $\kappa_+=0.9$. In Table 5.2, are reported the number of iterations required by each of these methods to reach different levels of precisions. In Figures 5.13 and 5.14 are illustrated the numerical results related to the PPS-CP and PDE-CP methods respectively.

By comparing this table with Table 5.1, we first observe that for the situation (a) the results generated the PDE method are nearly identical as for the SDE method. The SDE-CP and PDE-CP methods appear to have a slightly faster convergence, which can be justified by the gain provided by the addition of cutting planes. The PDE-CP method preserves consequently the good behaviour on well-conditioned problems such as in situation (a).

For situation (b), both the SDE-CP and PDE methods fail to provide an accurate estimate of $\boldsymbol{y}_{(b)}^\star$, whereas the PDE-CP method is able to do so. This observation highlights that it is the combination of these two effects, namely the use of location sets on the one hand, and taking partial steps on the other, that leads to a convergent method. The PPS-CS is also able to converge with high precision. We note that the number of iterations required by the PPS-CP and PDE-CP methods to reach accurate estimates of the solution $\boldsymbol{y}_{(b)}^\star$ is well inferior compared to the approach aforementioned in Section 5.4.4, which consists of either decreasing the step size α for the PS method or increasing β for the SDE method (see Figures 5.9 and 5.10 and Table 5.2).

By comparing the PPS-CP and PDE-CP methods in Table 5.2, we observe that the PDE-CP method requires fewer iterations compared to the PPS-CP method. However, an iteration of the PDE-CP method is twice as expensive as one of the PPS-CP method, which could decrease the appeal for using the PDE-CP. We should however keep in mind how the convergence guarantee depends on how well the problem is conditioned. Indeed, with Theorem 5.9, we know that the convergence of the PPS-CP method is dominated by the exponential term in which enters γ^2, whereas only γ appears in the exponential term in the convergence guarantee of the PDE-CP method as shown in Theorem 5.13. It is consequently interesting to compare both

methods on less well conditioned problems. To this end, we use the GVI problem defined in the situation (b) of Section 5.4.4, with the difference that we multiply one term by a constant ψ, *i.e.* $\mathbf{K} = [-3,3] \times [-3,3] \subset \mathbb{R}^2$ and

$$\boldsymbol{\Xi}(\boldsymbol{y}) = \begin{cases} [y_1 + y_2 + s_1,\ y_1 + \psi(2y_2) + s_2]^{\mathrm{T}}, & \text{if } y_1 y_2 > 0, \\ [y_1 - y_2 + s_1,\ -y_1 + \psi(2y_2) + s_2]^{\mathrm{T}}, & \text{if } y_1 y_2 < 0, \\ \mathbf{conv}\left\{ \begin{array}{l} [y_1 - y_2 + s_1,\ -y_1 + \psi(2y_2) + s_2]^{\mathrm{T}}, \\ [y_1 - y_2 + s_1,\ -y_1 + \psi(2y_2) + s_2]^{\mathrm{T}} \end{array} \right\}, & \text{if } y_1 y_2 = 0, \end{cases}$$

with $\boldsymbol{s} = [2,0]^{\mathrm{T}}$. The GVI problem defined as such is strongly monotone for values of ψ greater than or equal to 1 and has as solution $\boldsymbol{y}^{\star}_{(b)} = [-2,0]^{\mathrm{T}}$. We should note that for $\psi = 1$, we recover the situation (b) of Section 5.4.4.

In Table 5.3, we report the average number of iterations required by the PPS-CP and PDE-CP methods to reach different levels of precision for different values of ψ, *i.e.* $\psi \in \{1,2,4,8\}$. We also report the upper-bound γ^{con} on the largest condition number over any subset $\mathcal{W}$ of $\mathbf{K}$ on which $\boldsymbol{\Xi}$ is Lipschitz continuous. It is clear that the more ill conditioned the problem is, *i.e.* the larger γ^{con}, the more advantageous in terms of number of iterations it is to use the PDE-CP method rather than the PPS-CP method.

TABLE 5.2: Average number of iterations required by the different methods to reach different level of precision for both test cases. The entry "-" means that the level of precision has not been reached.

(a) Case (a)

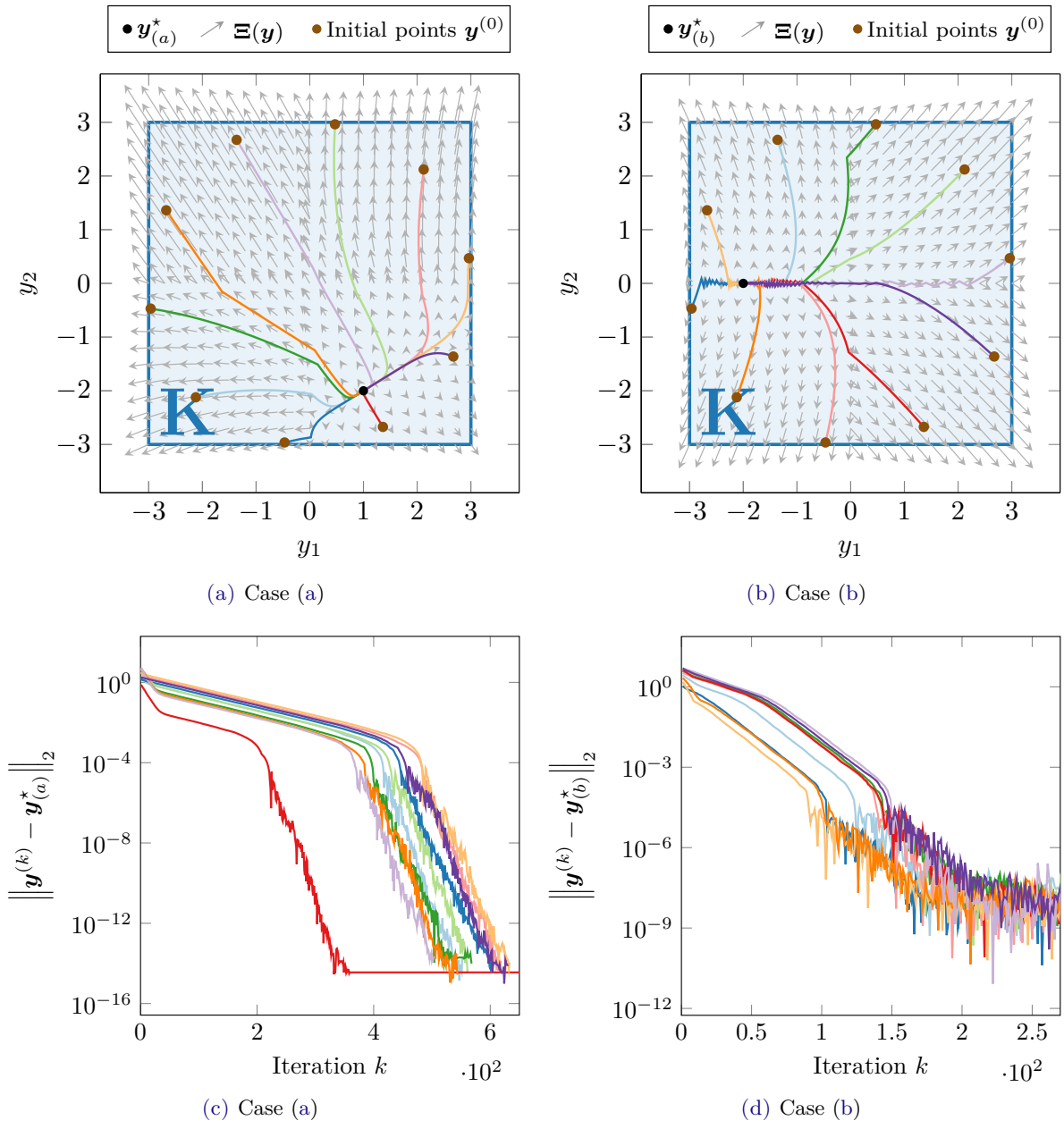
Error	PPS-CP	SDE-CP	PDE	PDE-CP
1.0e0	23.8	9.4	8.1	8.3
1.0e-1	134.3	33.1	26.4	29.5
1.0e-2	269.6	50.4	55.3	49.8
1.0e-4	406.9	84.5	91.1	85.1
1.0e-6	424.4	118.4	126.6	119.0
1.0e-8	457.6	152.6	161.4	152.9

(b) Case (b)

Error	PPS-CP	SDE-CP	PDE	PDE-CP
1.0e0	22.9	14.5	9.9	9.1
1.0e-1	58.0	-	-	25.6
1.0e-2	83.4	-	-	43.4
1.0e-4	126.7	-	-	77.2
1.0e-6	147.2	-	-	111.0
1.0e-8	185.7	-	-	147.0

5.5 Discussion and future research avenues

Firstly, we notice in Table 5.3 that the PDE-CP method does not converge to a ball of radius 10^{-8} around $\boldsymbol{y}^{\star}$. This reflects not a limitation of the method but one the solver employed. Indeed, the PDE-CP (and PPS-CP) requires at each iteration k two orthogonal projections on to the localisation set over the $\mathcal{L}^{(k)}$. However, as k grows, $\mathcal{L}^{(k)}$ becomes more complex (or degenerate) and the resolution of Equations (5.129b) and (5.129f) is susceptible to suffer from numerical issues. In the numerical results section, we used the Gurobi Optimization (2022)



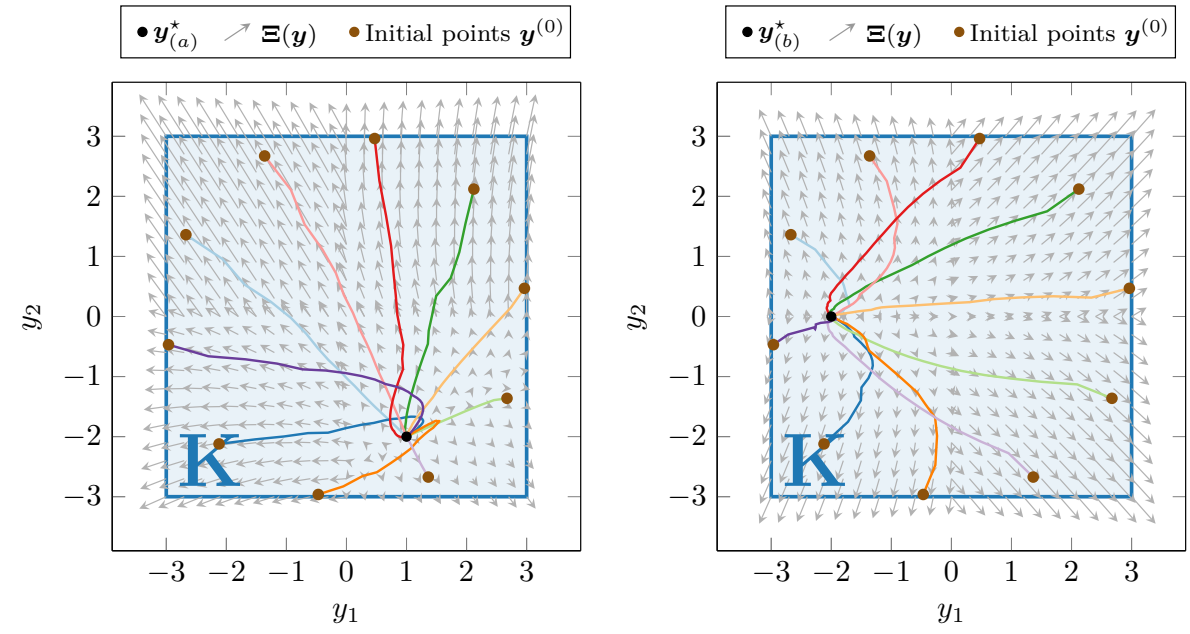
(a) Case (a)

(b) Case (b)

(c) Case (a)

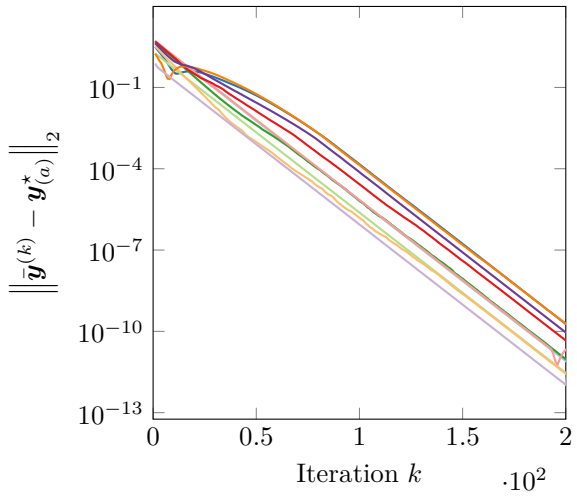
(d) Case (b)

FIGURE 5.13: **PPS-CP method.** Figures 5.13a and 5.13b: Trajectories of the iterates $\left(\boldsymbol{y}^{(k)}\right)_{k\geq 0}$ for different initial points $\boldsymbol{y}^{(0)}$. Figures 5.13a and 5.13b: Convergence of $\left(\boldsymbol{y}^{(k)}\right)_{k\geq 0}$ towards the solution $\boldsymbol{y}^\star$ of the GVI.

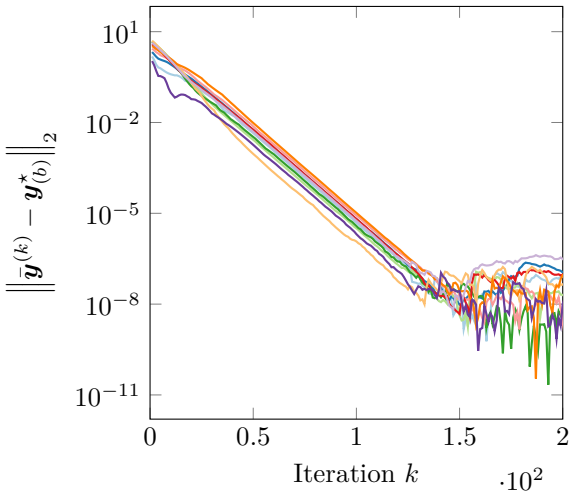


(a) Case (a)

(b) Case (b)



(c) Case (a)



(d) Case (b)

FIGURE 5.14: **PDE-CP method.** Figures 5.14a and 5.14b: Trajectories of the iterates $\left(\bar{\boldsymbol{y}}^{(k)}\right)_{k\geq 0}$ for different initial points $\boldsymbol{y}^{(0)}$. Figures 5.14c and 5.14d: Convergence of $\left(\bar{\boldsymbol{y}}^{(k)}\right)_{k\geq 0}$ towards the solution $\boldsymbol{y}^\star$ of the GVI.

TABLE 5.3: Average number of iterations required by the PPS-CP and PDE-CP methods to reach different levels of precision for different values of ψ. The entry "-" means that the level of precision has not been reached.

Error	$\psi = 1, \gamma^{con} \simeq 6.854$		$\psi = 2, \gamma^{con} \simeq 6.171$		$\psi = 4, \gamma^{con} \simeq 9.466$		$\psi = 8, \gamma^{con} \simeq 17.209$	
	PPS-CP	PDE-CP	PPS-CP	PDE-CP	PPS-CP	PDE-CP	PPS-CP	PDE-CP
1.0e0	22.9	9.1	29.3	8.5	77.4	12.4	266.2	20.8
1.0e-1	58.0	25.1	80.0	23.9	218.8	35.2	766.8	59.6
1.0e-2	83.6	43.5	118.8	41.0	334.3	61.8	1150.0	106.5
1.0e-4	126.7	77.2	179.9	72.5	487.2	111.0	1615.6	197.7
1.0e-6	145.1	111.2	197.0	103.0	502.3	157.2	1634.6	279.1
1.0e-8	181.8	164.1	247.6	136.2	537.7	203.3	1674.4	-

solver with the parameters `FeasibilityTol` and `OptimalityTol` both fixed to `1e-9`, which is the minimum value allowed by the solver. If for instance we reduce both of these parameters to `1e-4`, both the PPS-CP and PDE-CP methods struggle to find a solution with accuracy `1e-4`. Hence, we should not conclude that these methods cannot theoretically return solutions of any desired precision and keep in mind that their precision also depends on the underlying solvers.

Secondly, the test cases presented in this work are performed for a feasible set in $\mathbb{R}^2$. We should mention that it is known that the cutting plane scheme scales relatively poorly as the dimension of the feasible set increases (Nesterov, 2018), which could impact the final convergence of the PDE-CP (and PPS-CP) method and mitigate its relatively good convergence results.

Finally, in the PDE-CP method presented in Algorithm 5.6, we add a cutting plane at each iteration. As the number of iterations grows, some of these cutting planes could become redundant. Hence, it could be interesting in a future research to couple this method with a pruning scheme that eliminates redundant cuts in order to facilitate the resolution of the orthogonal projections.

5.6 Conclusion

The contribution of this work is two-fold and focused on the resolution of a generalised variational inequality (GVI) involving an upper hemicontinuous, compact- and convex-valued and strongly monotone operator over a compact and convex set. On the one hand, we show that the problem can be reformulated as finding an equilibrium point of a differential inclusion for which the operator is maximal monotone, and prove the uniqueness and exponential global stability of such equilibrium point. On the other hand, we devise a versatile method, namely the partial dual extrapolation method with cutting planes (PDE-CP), that is able to solve such GVI problem. We demonstrated its good numerical performance based on two test cases. The method combines the structure of a dual extrapolation method, which guarantees a fast convergence towards the neighbourhood of the solution of the GVI, together with a cutting plane scheme to iteratively reduce the size of the localisation set and guarantee an exact convergence. By doing so, the

method can adapt its behaviour on whether there exists a neighbourhood of the solution of the GVI problem on which the operator is single valued.

Conclusions

In this thesis we considered the problem of electricity prosumers who hedge against risk by trading future contracts on a market with limited liquidity. In Chapter 2, we introduced a new model for the optimal execution problem which takes into account the uncertainty related to the volume to be traded. We showed that the model leads to strategies which are able to adapt to the uncertain nature of the volume target via predetermined recourse actions. In Chapter 3, we showed that this problem could also be solved via a dynamic programming approach that is computationally economical. Then, in Chapter 4, we generalised the single-player model derived in Chapter 2 to the multi-player case. Lastly, Chapter 5 introduced a new numerical method for calculating solutions to strongly monotone generalised variational inequalities.

To conclude this thesis, we outline some directions for future work, some of which are the subject of my ongoing research. In Chapter 2, we considered linear predetermined recourse actions parametrised by $\boldsymbol{\beta}$ and mentioned that numerical evidence showed limited gain in optimising $\boldsymbol{\beta}$. However, we believe that a more detailed analysis of the recourse rule would be interesting because even small savings in the trading cost make a large difference over time, due to the cumulative effect of sequential trades.

In Chapter 3, we assumed the absence of permanent impact of the trades. This restrictive assumption was convenient in order to prove the convexity of the value function and show that the dynamic program can be solved efficiently. Investigating whether this assumption can be lifted would be of great interest, as this would lead to a more widely applicable framework.

In Chapter 4, we formulated Conjecture 4.1. Proving this result would be interesting as this would provide new insights into not only the particular problem of optimal execution considered in this thesis, but any Nash equilibrium problem where players are risk-averse and estimate risk with the $\mathbb{E}$-CVaR$_{\alpha,\lambda}$ measure. Another avenue for future research concerns the method employed to derive the equilibria in the numerical section of Chapter 4. We employed a method similar to the one derived in Chapter 5, but in a real game, it is realistic to assume that traders would adapt their strategy by taking the best response given the current market situation, by using either a Jacobi or a Gauss-Seidel approach (see details in Fischer et al. (2014)). This could potentially lead to a situation where there the theoretical equilibrium is not reached and traders oscillate around it. Finally, it would be interesting to consider a leader-follower model, where the players with greater market power play first.

In Chapter 5, we introduce the PDE-CP method and proved its convergence towards a

neighbour of the solution of the GVI. As mentioned in the chapter, it would be interesting to prove the exact convergence of the method, *i.e.* the convergence not to a neighborhood of the solution but to the solution itself. This could be achieved, for example, by proving that the size of the localisation sets strictly decreases with a certain rate.

Finally, it would be interesting to study whether the trading strategy approach under uncertainty presented in this thesis can be combined with other works (*e.g.* de Maere d'Aertrycke and Smeers (2013); Gérard et al. (2018)) in order to derive electricity prices endogenously. According to us, the main challenge to this end would be to enforce the market clearing constraints, which we have neglected using the argument that it could be modelled via the temporary impact of traders. Indeed, we assumed that a power generator is always ready to produce electricity provided that a wholesaler offers a sufficiently large price for a desired volume. In our model, this is reflected in the temporary impact of the trades on the transaction price. We believe that a promising approach to impose the clearing constraint while ensuring convergence to a Nash equilibrium would be to study the effects of a tax, which would provide an incentive for prosumers to alter their strategies (*e.g.* Ruddell et al. (2017)).

Appendix A

Support materials related to Chapter 3

A.1 Auxiliary results

PROPOSITION A.1. *Let A and B be independent real-valued random variables taking values in finite discrete spaces Ω_A and Ω_B with associated laws $\mathbb{P}_A$ and $\mathbb{P}_B$ such that:*

$$\mathbb{E}^{\mathbb{P}_A}[A] = 0. \tag{A.1}$$

Let ϱ be the CVaR_α risk measure for a risk threshold $\alpha \in [0, 1)$. Then,

$$\forall c \geq 0, \quad \forall \nu > 0 : \qquad \varrho[cA + B] \leq \varrho[(c + \nu) A + B]\,. \tag{A.2}$$

PROOF. W.l.o.g. we assume that $\Omega_A = \{\omega_{A,1}, \ldots, \omega_{A,m}\}$ such that $A(\omega_{A,i}) > A(\omega_{A,j})$ for $i < j$. Let Ω be the sample space defined as $\Omega := \Omega_A \times \Omega_B$, let $\mathcal{P}_\Omega$ be the set of probability measures on Ω, and let $\mathcal{D}$ be the risk-set associated to CVaR_α, which is defined as follows:

$$\mathcal{D} := \left\{ \mu \in \mathcal{P}_\Omega \,\middle|\, \forall \omega = (\omega_A, \omega_B) \in \Omega : 0 \leq \mu(\omega) \leq \min\left(\frac{\pi_A(\omega_A)\, \pi_B(\omega_B)}{\alpha}, 1 \right) \right\}, \tag{A.3}$$

where $\pi_A(\omega_A)$ should be understood as $\mathbb{P}_A[\omega_A]$ and similarly for B. Then, with Shapiro et al. (2009, Theorems 6.4 and 6.6) we have that

$$\varrho[cA + B] = \max_{\mu \in \mathcal{D}} \mathbb{E}^\mu[cA + B] \tag{A.4}$$

Let $\mu^* \in \mathcal{D}$ be the *worst case probability measure* solving (A.4). Then

$$\varrho[cA + B] = \mathbb{E}^{\mu^*}[cA + B] = c\, \mathbb{E}^{\mu^*}[A] + \mathbb{E}^{\mu^*}[B]\,. \tag{A.5}$$

We must have that $\mathbb{E}^{\mu^*}[A] \geq 0$. Indeed, by contradiction, if $\mathbb{E}^{\mu^*}[A] < 0$, there exists a $\bar{\omega}_B \in \Omega_B^*$ with

$$\Omega_B^* = \{\omega_B \in \Omega_B \,|\, \exists \omega_A \in \Omega_A : \mu^*(\omega_A, \omega_B) > 0\} \tag{A.6}$$

such that

$$\mathbb{E}^{\mu^*}[A \mid \omega_B = \bar{\omega}_B] = \frac{1}{\mu^*(\bar{\omega}_B)} \sum_{\omega_A \in \Omega_A} A(\omega_A)\, \mu^*(\omega_A, \bar{\omega}_B) < 0, \tag{A.7}$$

where $\mu^*(\bar{\omega}_B) \coloneqq \sum_{\omega_A \in \Omega_A} \mu^*(\omega_A, \bar{\omega}_B)$. If, we define the probability measure $\mu^\dagger$ as:

$$\forall \omega_B \notin \Omega_B^* : \mu^\dagger(\omega_A, \omega_B) = \mu^*(\omega_A, \omega_B)\,, \tag{A.8a}$$

$$\forall i \in [\![1, m]\!] : \mu^\dagger(\omega_{A,i}, \bar{\omega}_B) = \min\left(\frac{\pi_A(\omega_{A,i})\, \pi_B(\bar{\omega}_B)}{\alpha}, \max\left(\mu^*(\bar{\omega}_B) - \sum_{k=1}^{i-1} \mu^\dagger(\omega_{A,k}, \bar{\omega}_B)\,, 0 \right) \right). \tag{A.8b}$$

Then, $\mu^\dagger \in \mathcal{D}$ and

$$\mathbb{E}^{\mu^*}[A \mid \omega_B = \bar{\omega}_B] < \mathbb{E}^{\mu^\dagger}[A \mid \omega_B = \bar{\omega}_B]\,, \tag{A.9}$$

where we used the fact that $A(\omega_{A,i}) > A(\omega_{A,j})$ for $i < j$. Also, $\mathbb{E}^{\mu^\dagger}[B] = \mathbb{E}^{\mu^*}[B]$. Consequently,

$$\begin{aligned}
\mathbb{E}^{\mu^*}[cA + B] &= c\,\mathbb{E}^{\mu^*}[A \mid \omega_B = \bar{\omega}_B] + c\,\mathbb{E}^{\mu^*}[A \mid \omega_B \neq \bar{\omega}_B] + \mathbb{E}^{\mu^*}[B] && \text{(A.10)}\\
&= c\,\mathbb{E}^{\mu^*}[A \mid \omega_B = \bar{\omega}_B] + c\,\mathbb{E}^{\mu^\dagger}[A \mid \omega_B \neq \bar{\omega}_B] + \mathbb{E}^{\mu^\dagger}[B] && \text{(A.11)}\\
&< c\,\mathbb{E}^{\mu^\dagger}[A \mid \omega_B = \bar{\omega}_B] + c\,\mathbb{E}^{\mu^\dagger}[A \mid \omega_B \neq \bar{\omega}_B] + \mathbb{E}^{\mu^\dagger}[B] && \text{(A.12)}\\
&= \mathbb{E}^{\mu^\dagger}[cA + B]\,, && \text{(A.13)}
\end{aligned}$$

which violates the fact that μ^* is the worst case probability measure. Hence, as $\mathbb{E}^{\mu^*}[A] \geq 0$ we have that

$$\begin{aligned}
\varrho[(c+\nu)\, A + B] &\geq \mathbb{E}^{\mu^*}[(c+\nu)\, A + B]\\
&= (c+\nu)\, \mathbb{E}^{\mu^*}[A] + \mathbb{E}^{\mu^*}[B] \geq c\,\mathbb{E}^{\mu^*}[A] + \mathbb{E}^{\mu^*}[B] = \varrho[cA + B]\,, \quad \text{(A.14)}
\end{aligned}$$

where the first inequality comes from the fact that μ^* is not necessarily the worst case probability measure for $(c+\nu)\, A + B$. This terminates the proof. □

Corollary A.1. *Let A and B be independent real-valued random variables taking values in finite discrete spaces Ω_A and Ω_B with associated laws $\mathbb{P}_A$ and $\mathbb{P}_B$ such that:*

$$\mathbb{E}^{\mathbb{P}_A}[A] = 0. \tag{A.15}$$

Let ϱ be the $\mathbb{E}$-CVaR$_{\alpha,\lambda}$ risk measure for a risk threshold $\alpha \in [0, 1)$ and $\lambda \in [0, 1]$. Then,

$$\forall c \geq 0, \quad \forall \nu > 0 : \qquad \varrho[cA + B] \leq \varrho[(c+\nu)\, A + B]\,. \tag{A.16}$$

Proof. The proof follows from Proposition A.1 and the fact that

$$\forall c \geq 0, \forall \nu > 0 : \quad \mathbb{E}[cA + B] = c\,\mathbb{E}[A] + \mathbb{E}[B] = (c+\nu)\, \mathbb{E}[A] + \mathbb{E}[B] = \mathbb{E}[(c+\nu)\, A + B]\,. \tag{A.17}$$

□

A.2 Proofs

PROPOSITION A.2 (PROOF OF PROPOSITION 3.7). *Given Assumptions 3.1 to 3.9, the following assertions hold for all* $k \in [\![0, m-1]\!]$*:*

(i) For any $x_k \in \bar{q}\,\mathbb{Z} : q^*_{k+1}(x_k + \bar{q}) \in \left\{q^*_{k+1}(x_k)\,, q^*_{k+1}(x_k) + \bar{q}\right\}$*, where* $q^*_{k+1}(x_k)$ *is the solution of Optimisation Problem (3.49) for* $x = x_k$*.*

(ii) $\{f_k(j\bar{q})\}_{j\in\mathbb{Z}}$ *is a convex sequence.*

PROOF. We proceed by backward induction to show Proposition A.2. For the last stage $k = m-1$, $q^*_m(x) = x$ and $f_{m-1}(x) = \epsilon_m |x| + \frac{\eta_m x^2}{\tau_m}$, hence it is straightforward that both assertions hold. We now assume that assertions A.2*i* and A.2*ii* hold for stage $k+1$ and show that it implies that they also hold for stage k. First, since $\{f_{k+1}(j\bar{q})\}_{j\in\mathbb{Z}}$ is a convex sequence, it implies, for all $\omega \in \Omega_{\xi_{k+1}\times\delta_{k+1}}$, that the sequence $r_{k+1}[\omega] \coloneqq \{\xi_{k+1}[\omega](j\bar{q}) + f_{k+1}(j\bar{q} + \delta_{k+1}[\omega])\}_{j\in\mathbb{Z}}$ is convex, which in turn implies that the following sequence is convex:

$$r_{k+1} \coloneqq \{\varrho_k[\xi_{k+1} j\bar{q} + f_{k+1}(j\bar{q} + \delta_{k+1})]\}_{j\in\mathbb{Z}} = \{l_k(j\bar{q})\}_{j\in\mathbb{Z}}\,. \tag{A.18}$$

This can be shown by first taking the linear interpolation $\bar{f}_{k+1}[\omega]$ of the sequence $r_{k+1}[\omega]$ for any $\omega \in \Omega_{\xi_{k+1}\times\delta_{k+1}}$, and then use the argument that, since each function $\bar{f}_{k+1}[\omega](\cdot)$ is convex, $\varrho_k\left[\bar{f}_{k+1}(\cdot)\right]$ is also a convex function, which finally implies the convexity of the sequence r_{k+1}. As a consequence of Proposition 3.6, the sequence r_{k+1} has a minimum that is attained. W.l.o.g. we assume that this minimum is unique and is attained for index j^*_{k+1}, and thus for $x^*_k \coloneqq j^*_{k+1}\bar{q}$. As a consequence, for any state $x_k \coloneqq j_k\bar{q} \in \bar{q}\,\mathbb{Z}$, the objective of a risk-averse trader is to get closer from x^*_k while factoring into the balance their trading cost. This can been seen as finding the minimum of the sequence:

$$\left\{\epsilon_{k+1}\,|j\bar{q}| + \frac{\eta_{k+1}}{\tau_{k+1}}(j\bar{q})^2 + l_k((j - j_k)\,\bar{q})\right\}_{j\in\mathbb{Z}}. \tag{A.19}$$

This sequence is convex and has a unique minimum that is attained. Based on the convexity of the sequence r_{k+1}, it is straightforward to show the following assertions:

$$\begin{cases} x_k < x^*_k \Rightarrow 0 \geq q^*_{k+1}(x_k) \geq x_k - x^*_k, \\ x_k = x^*_k \Rightarrow q^*_{k+1}(x_k) = 0, \\ x_k > x^*_k \Rightarrow 0 \leq q^*_{k+1}(x_k) \leq x_k - x^*_k. \end{cases} \tag{A.20}$$

For readability purpose, we use for the rest of this proof q, q^- and q^+ as shorthand for $q^*_{k+1}(x_k)$, $q^*_{k+1}(x_k - \bar{q})$ and $q^*_{k+1}(x_k + \bar{q})$ respectively.

- To prove that Proposition A.2i holds, we first show that if $x_k \geq x^*_k$, then the following inequalities hold: $q^+ \leq q + \bar{q}$ and $q^+ \geq q$, which imply that for all $x_k \geq x^*_k$ that $q^+ \in \{q, q + \bar{q}\}$. Since we assume that $x_k \geq x^*_k$, we must have with Equation (A.20) that $q \geq 0$ and $q^+ \geq 0$.

- We now show that $q^+ \leq q + \bar{q}$. By contradiction, we assume that $q^+ = q + (\mu + 1)\,\bar{q}$ with $\mu \in \mathbb{N} \setminus \{0\}$. Then, since q^+ is optimal, we must have that

$$\begin{aligned}
&\epsilon_{k+1} q^+ + \frac{\eta_{k+1}}{\tau_{k+1}} \Big(q^+\Big)^2 + l_k\Big(x_k + \bar{q} - q^+\Big) \\
&\leq \epsilon_{k+1}(q + \bar{q}) + \frac{\eta_{k+1}}{\tau_{k+1}}(q + \bar{q})^2 + l_k(x_k + \bar{q} - (q + \bar{q})) \\
&\leq \epsilon_{k+1}(q + \bar{q}) + \frac{\eta_{k+1}}{\tau_{k+1}}(q + \bar{q})^2 + l_k(x_k - q)\,.
\end{aligned} \tag{A.21}$$

  By the optimality of q, we also have that

$$\begin{aligned}
&\epsilon_{k+1} q + \frac{\eta_{k+1}}{\tau_{k+1}} q^2 + l_k(x_k - q) \\
&\leq \epsilon_{k+1}\Big(q^+ - \bar{q}\Big) + \frac{\eta_{k+1}}{\tau_{k+1}}\Big(q^+ - \bar{q}\Big)^2 + l_k\Big(x_k - \Big(q^+ - \bar{q}\Big)\Big) \\
&\leq \epsilon_{k+1}\Big(q^+ - \bar{q}\Big) + \frac{\eta_{k+1}}{\tau_{k+1}}\Big(q^+ - \bar{q}\Big)^2 + l_k\Big(x_k + \bar{q} - q^+\Big)\,.
\end{aligned} \tag{A.22}$$

  Therefore, by combining Inequalities (A.21) and (A.22), we obtain

$$\frac{\eta_{k+1}}{\tau_{k+1}}\Big(q^+\Big)^2 + \frac{\eta_{k+1}}{\tau_{k+1}} q^2 \leq \frac{\eta_{k+1}}{\tau_{k+1}}(q + \bar{q})^2 + \frac{\eta_{k+1}}{\tau_{k+1}}\Big(q^+ - \bar{q}\Big)^2\,,$$

  which further simplifies to $2\eta\mu\bar{q}^2 \leq 0$ and leads to the desired contradiction since $\mu > 0$.

- We now show that $q^+ \geq q$. By contradiction, we assume that $0 \leq q^+ < q$. Since q is the optimal trading quantity for $x = x_k$, we must have for all $\tilde{q} \in [0, q - \bar{q}] \cap \bar{q}\,\mathbb{Z}$ that

$$\begin{aligned}
l_k(x_k - \tilde{q}) - l_k(x_k - (\tilde{q} + \bar{q})) &> \epsilon_{k+1}(\tilde{q} + \bar{q}) + \frac{\eta_{k+1}}{\tau_{k+1}}(\tilde{q} + \bar{q})^2 - \Big(\epsilon_{k+1}\tilde{q} + \frac{\eta_{k+1}}{\tau_{k+1}}\tilde{q}^2\Big) \\
&> \epsilon_{k+1}\bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}}\Big(\bar{q}^2 + 2\bar{q}\tilde{q}\Big)\,.
\end{aligned} \tag{A.23}$$

  In words, Equation (A.23) means that, for any $0 \leq \tilde{q} \in \bar{q}\,\mathbb{Z} < \bar{q}$, if ones trades $\bar{q}$ more units, then the gain in risk reduction (LHS), which is given by the function l_k, is greater than the increase in the deterministic cost (RHS). Now, if $q^+ < q$, then it means that trading $\bar{q}$ more units compared to q^+, would lead to an increase in deterministic cost that is greater than the resulting risk reduction:

$$\begin{aligned}
&l_k\Big(x_k + \bar{q} - q^+\Big) - l_k\Big(x_k + \bar{q} - \Big(q^+ + \bar{q}\Big)\Big) \\
&< \epsilon_{k+1}\Big(q^+ + \bar{q}\Big) + \frac{\eta_{k+1}}{\tau_{k+1}}\Big(q^+ + \bar{q}\Big)^2 - \Big(\epsilon_{k+1} q^+ + \frac{\eta_{k+1}}{\tau_{k+1}} q^+\Big) \\
&< \epsilon_{k+1}\bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}}\Big(\bar{q}^2 + 2\bar{q}q^+\Big)\,.
\end{aligned} \tag{A.24}$$

  Since the sequence r_{k+1} is convex, we must have that $r_{k+1,j} - r_{k+1,j-1} \leq r_{k+1,j+1} - r_{k+1,j}$. We have therefore that

$$l_k\Big(x_k - q^+\Big) - l_k\Big(x_k - \Big(q^+ + \bar{q}\Big)\Big) \leq l_k\Big(x_k - \Big(q^+ - \bar{q}\Big)\Big) - l_k\Big(x_k - q^+\Big)\,. \tag{A.25}$$

By combining Equations (A.23) to (A.25) and using the fact that $q^+ \in [0, q-\bar{q}] \cap \bar{q}\,\mathbb{Z}$, we obtain that

$$\begin{aligned} \epsilon_{k+1}\bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}}\left(\bar{q}^2 + 2\bar{q}q^+\right) &\overset{(A.23)}{<} l_k\left(x_k - q^+\right) - l_k\left(x_k - \left(q^+ + \bar{q}\right)\right) \\ &\overset{(A.25)}{\leq} l_k\left(x_k - \left(q^+ - \bar{q}\right)\right) - l_k\left(x_k - q^+\right) \\ &\overset{(A.24)}{<} \epsilon_{k+1}\bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}}\left(\bar{q}^2 + 2\bar{q}q^+\right), \end{aligned} \tag{A.26}$$

which leads to the desired contradiction.

Similarly, it can be shown that if $x_k \leq x_k^*$ then the following inequalities hold: $q^- \geq q - \bar{q}$ and $q^- \leq q$, which imply that for all $x_k \leq x_k^*$ that $q^- \in \{q, q-\bar{q}\}$, and thus that $q^+ \in \{q, q+\bar{q}\}$ for all $x_k < x_k^*$. As a consequence, Proposition A.2i is proven.

- We show now that Proposition A.2ii holds too. To show the convexity of the sequence $\{f_k(j\bar{q})\}_{j\in\mathbb{Z}}$, we must show, for all $x_k \in \bar{q}\,\mathbb{Z}$, that

$$f_k(x_k) - f_k(x_k - \bar{q}) \leq f_k(x_k + \bar{q}) - f_k(x_k)\,. \tag{A.27}$$

For a given x_k, we denote by q^-, q and q^+ the optimal quantity to be traded for $x = x_k - \bar{q}$, x_k and $x_k + \bar{q}$ respectively. With Proposition A.2i, there might be only 4 possibilities:

  - $q^- = q = q^+$: In this case Equation (A.27) becomes:

$$l_k(x_k) - l_k(x_k - \bar{q}) \leq l_k(x_k + \bar{q}) - l_k(x_k)\,,$$

    which holds true since r_{k+1} is a convex sequence.

  - $q - \bar{q} = q^- < q < q^+ = q + \bar{q}$: Equation (A.27) then becomes:

$$\begin{aligned} \epsilon_{k+1}q + \frac{\eta_{k+1}}{\tau_{k+1}}q^2 - \left(\epsilon_{k+1}(q-\bar{q}) + \frac{\eta_{k+1}}{\tau_{k+1}}(q-\bar{q})^2\right) \\ \leq \epsilon_{k+1}(q+\bar{q}) + \frac{\eta_{k+1}}{\tau_{k+1}}(q+\bar{q})^2 - \left(\epsilon_{k+1}q + \frac{\eta_{k+1}}{\tau_{k+1}}q^2\right), \end{aligned} \tag{A.28}$$

    which holds true by the convexity of the function $y \to \epsilon_{k+1}y + \frac{\eta_{k+1}}{\tau_{k+1}}y^2$.

  - $q^- = q < q^+ = q + \bar{q}$: The optimality of q for $x = x_k$ implies that

$$\epsilon_{k+1}q + \frac{\eta_{k+1}}{\tau_{k+1}}q^2 + l_k(x_k - q) < \epsilon_{k+1}(q+\bar{q}) + \frac{\eta_{k+1}}{\tau_{k+1}}(q+\bar{q})^2 + l_k(x_k - q - \bar{q})\,, \tag{A.29}$$

    which in turn implies that

$$l_k(x_k - q) - l_k(x_k - q - \bar{q}) < \epsilon_{k+1}\bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}}\left(\bar{q}^2 + 2q\bar{q}\right). \tag{A.30}$$

    Since $q^- = q$, we have that $f_k(x_k) - f_k(x_k - \bar{q}) = l_k(x_k - q) - l_k(x_k - \bar{q} - q)$. Besides, since $q^+ = q + \bar{q}$, we have that

$$f_k(x_k + \bar{q}) - f_k(x_k) = \epsilon_{k+1}(q+\bar{q}) + \frac{\eta_{k+1}}{\tau_{k+1}}(q+\bar{q})^2 - \left(\epsilon_{k+1}q + \frac{\eta_{k+1}}{\tau_{k+1}}q^2\right) \tag{A.31}$$

$$= \epsilon_{k+1}\bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}}\left(\bar{q}^2 + 2q\bar{q}\right). \tag{A.32}$$

This implies with (A.30) that (A.27) holds.

- $q - \bar{q} = q^- < q = q^+$: The optimality of q for $x = x_k$ implies that

$$\epsilon_{k+1}q + \frac{\eta_{k+1}}{\tau_{k+1}}q^2 + l_k(x_k - q) < \epsilon_{k+1}(q - \bar{q}) + \frac{\eta_{k+1}}{\tau_{k+1}}(q - \bar{q})^2 + l_k(x_k - q + \bar{q}), \tag{A.33}$$

which in turn implies that

$$l_k(x_k - q) - l_k(x_k - q + \bar{q}) < -\epsilon_{k+1}\bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}}\left(\bar{q}^2 - 2q\bar{q}\right). \tag{A.34}$$

Since $q^+ = q$, we have that

$$f_k(x_k + \bar{q}) - f_k(x_k) = l_k(x_k + \bar{q} - q) - l_k(x_k - q). \tag{A.35}$$

Besides this, since $q^- = q - \bar{q}$, we have that

$$f_k(x_k) - f_k(x_k - \bar{q}) = \epsilon_{k+1}q + \frac{\eta_{k+1}}{\tau_{k+1}}q^2 - \left(\epsilon_{k+1}(q - \bar{q}) + \frac{\eta_{k+1}}{\tau_{k+1}}(q - \bar{q})^2\right) \tag{A.36}$$

$$= \epsilon_{k+1}\bar{q} + \frac{\eta_{k+1}}{\tau_{k+1}}\left(2q\bar{q} - \bar{q}^2\right). \tag{A.37}$$

This implies with (A.34) that (A.27) holds.

This proves the desired result. □

Appendix B

Support materials related to Chapter 4

B.1 Auxiliary results

PROPOSITION B.1. *For any matrices* $\mathbf{M} \in \mathbb{R}^{m\times m}$ *and* $\mathbf{N} \in \mathbb{R}^{km\times km}$ *defined as follows for* $k \geq 2$*:*

$$\mathbf{N} \coloneqq \begin{bmatrix} 2\,\mathbf{M} & \mathbf{M} & \cdots & \cdots & \mathbf{M} \\ \mathbf{M} & 2\,\mathbf{M} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mathbf{M} \\ \mathbf{M} & \cdots & \cdots & \mathbf{M} & 2\,\mathbf{M} \end{bmatrix}, \tag{B.1}$$

we have that the following assertions holds:

(i) If λ *is an eigenvalue of* $\mathbf{M}$*, then* λ *and* $(k+1)\,\lambda$ *are eigenvalues of* $\mathbf{N}$ *with geometric multiplicity* $k-1$ *and 1 respectively.*

(ii) If λ *is an eigenvalue of* $\mathbf{N}$*, then either* λ *or* $\frac{\lambda}{k+1}$ *is an eigenvalue of* $\mathbf{M}$*.*

PROOF.

- Proposition B.1i: Let $\lambda \in \mathbb{C}$ and $\boldsymbol{u} \in \mathbb{R}^m$ be such that $\mathbf{M}\,\boldsymbol{u} = \lambda\boldsymbol{u}$. If we define the vector $\boldsymbol{v} \coloneqq [\boldsymbol{u}, \ldots, \boldsymbol{u}]^{\mathrm{T}} \in \mathbb{R}^{mk}$, then the following shows that $(k+1)\,\lambda$ is an eigenvalue of $\mathbf{N}$:

$$\mathbf{N}\,\boldsymbol{v} = [(k+1)\,\mathbf{M}\,\boldsymbol{u}, \ldots, (k+1)\,\mathbf{M}\,\boldsymbol{u}]^{\mathrm{T}} = (k+1)\,\lambda[\boldsymbol{u}, \ldots, \boldsymbol{u}]^{\mathrm{T}} = (k+1)\,\lambda\boldsymbol{v}. \tag{B.2}$$

  If we define the vector $\boldsymbol{w} \coloneqq [\boldsymbol{u}, -\boldsymbol{u}, \boldsymbol{0}, \ldots, \boldsymbol{0}]^{\mathrm{T}} \in \mathbb{R}^{mk}$, then the following shows that λ is an eigenvalue of $\mathbf{N}$:

$$\mathbf{N}\,\boldsymbol{w} = [\mathbf{M}\,\boldsymbol{u}, -\mathbf{M}\,\boldsymbol{u}, \boldsymbol{0}, \ldots, \boldsymbol{0}]^{\mathrm{T}} = \lambda[\boldsymbol{u}, -\boldsymbol{u}, \boldsymbol{0}, \ldots, \boldsymbol{0}]^{\mathrm{T}} = \lambda\boldsymbol{w}. \tag{B.3}$$

- Proposition B.1ii: Let $\lambda \in \mathbb{C}$ and $\boldsymbol{v} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k]^{\mathrm{T}} \in \mathbb{R}^{mk}$ such that $\mathbf{N}\,\boldsymbol{v} = \lambda \boldsymbol{v}$. Then, it implies the following equality holds that for any block row $i \in [\![1, k]\!]$:

$$\mathbf{M}\,\boldsymbol{u}_i + \mathbf{M}\left(\sum_{r=1}^{k} \boldsymbol{u}_r\right) = \lambda \boldsymbol{u}_i. \tag{B.4}$$

  If $\exists \boldsymbol{u} \in \mathbb{R}^m$ such that $\boldsymbol{u}_i = \boldsymbol{u}$ for all $i \in [\![1, m]\!]$, then Equation (B.4) implies that

$$(k+1)\,\mathbf{M}\,\boldsymbol{u} = \lambda \boldsymbol{u}, \tag{B.5}$$

  which in turn implies that $\frac{\lambda}{k+1}$ is an eigenvalue of $\mathbf{M}$. If such $\boldsymbol{u}$ does not exist, there must exist two indices i and j where $\boldsymbol{u}_i \neq \boldsymbol{u}_j$ and where at least one of them is different from the null vector since $\boldsymbol{v} \neq 0$ is an eigenvector of $\mathbf{N}$. Hence, we have with Equation (B.4) that

$$\mathbf{M}(\boldsymbol{u}_i - \boldsymbol{u}_j) = \lambda(\boldsymbol{u}_i - \boldsymbol{u}_j)\,, \tag{B.6}$$

  which implies that λ is an eigenvalue of $\mathbf{M}$ with associated eigenvector $\boldsymbol{u}_i - \boldsymbol{u}_j$.

This completes the proof. □

Corollary B.1. *The matrix $\mathbf{M} \in \mathbb{R}^{m \times m}$ is a positive (semi-)definite if and only if the matrix $\mathbf{N} \in \mathbb{R}^{km \times km}$ with $k \in \mathbb{N} \geq 1$ is positive (semi-)definite, where*

$$\mathbf{N} = \begin{bmatrix} 2\,\mathbf{M} & \mathbf{M} & \cdots & \cdots & \mathbf{M} \\ \mathbf{M} & 2\,\mathbf{M} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mathbf{M} \\ \mathbf{M} & \cdots & \cdots & \mathbf{M} & 2\,\mathbf{M} \end{bmatrix}. \tag{B.7}$$

Moreover,

$$\lambda_{\min}(\mathbf{M}) = \lambda_{\min}(\mathbf{N})\,. \tag{B.8}$$

Proof. The proof derives directly from Proposition B.1. □

B.2 Proofs

Proposition B.2 (Proof of Proposition 4.5). *If $\bar{\mathbf{M}} \succ 0$, then, for any realisation $\omega \in \Omega$, the truncated pseudo-gradient map $\boldsymbol{\nabla}^{\mathrm{tr}(\omega)} \widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}$ defined in (4.51) is strongly monotone on $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$ with constant $\frac{1}{2}\lambda_{\min}\left(\bar{\mathbf{M}}\right)$, where $\lambda_{\min}\left(\bar{\mathbf{M}}\right)$ denotes the smallest eigenvalue of $\bar{\mathbf{M}}$.*

Proof. For a point $\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)} \in \widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$, we have defined in Equation (4.50) the truncated gradient as follows

$$\nabla_{\nu}^{\mathrm{tr}(\omega)} \widehat{\mathcal{C}}_{\nu}^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) := \left[\frac{\partial \widehat{\mathcal{C}}_{\nu}^{\mathrm{tr}(\omega)}}{\partial q_{\nu,\omega,1}^{+}}, \dots, \frac{\partial \widehat{\mathcal{C}}_{\nu}^{\mathrm{tr}(\omega)}}{\partial q_{\nu,\omega,m-1}^{+}}, \frac{\partial \widehat{\mathcal{C}}_{\nu}^{\mathrm{tr}(\omega)}}{\partial q_{\nu,\omega,1}^{-}}, \dots, \frac{\partial \widehat{\mathcal{C}}_{\nu}^{\mathrm{tr}(\omega)}}{\partial q_{\nu,\omega,m}^{-}}\right]^{\mathrm{T}}, \tag{B.9}$$

where for all $k \in [\![1, m-1]\!]$

$$
\begin{aligned}
&\frac{\partial \widehat{\mathcal{C}}_{\nu}^{\mathrm{tr}(\omega)}}{\partial q_{\nu,\omega,k}^{+}}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) \\
&\quad = \gamma_k \left(D_{\nu,m}(\omega) - \sum_{i=1}^{k} \left(q_{\nu,\omega,i}^{+} - q_{\nu,\omega,i}^{-} \right) \right) - \sum_{i=k}^{m-1} \xi_{\nu,i}(\omega) \\
&\quad - \sum_{i=k}^{m-1} \gamma_i \left(\sum_{\tilde{\nu} \in [\![1,n]\!]} \left(q_{\tilde{\nu},\omega,i}^{+} - q_{\tilde{\nu},\omega,i}^{-} \right) \right) + \epsilon_k + \frac{\eta_k}{\tau_k} \left(2 q_{\nu,\omega,k}^{+} + \sum_{\tilde{\nu} \neq \nu} q_{\tilde{\nu},\omega,k}^{+} \right) \\
&\quad - \epsilon_m - \frac{2\eta_m}{\tau_m} \left(D_{\nu,m}(\omega) + q_{\nu,\omega,m}^{-} - \sum_{k=1}^{m-1} \left(q_{\nu,\omega,k}^{+} - q_{\nu,\omega,k}^{-} \right) \right) \\
&\quad - \frac{\eta_m}{\tau_m} \left(\sum_{\tilde{\nu} \neq \nu} \left(D_{\tilde{\nu},m}(\omega) + q_{\tilde{\nu},\omega,m}^{-} - \sum_{k=1}^{m-1} \left(q_{\tilde{\nu},\omega,k}^{+} - q_{\tilde{\nu},\omega,k}^{-} \right) \right) \right),
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial \widehat{\mathcal{C}}_{\nu}^{\mathrm{tr}(\omega)}}{\partial q_{\nu,\omega,k}^{-}}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) \\
&\quad = -\gamma_k \left(D_{\nu,m}(\omega) - \sum_{i=1}^{k} \left(q_{\nu,\omega,i}^{+} - q_{\nu,\omega,i}^{-} \right) \right) + \sum_{i=k}^{m-1} \xi_{\nu,i}(\omega) \\
&\quad + \sum_{i=k}^{m-1} \gamma_i \left(\sum_{\tilde{\nu} \in [\![1,n]\!]} \left(q_{\tilde{\nu},\omega,i}^{+} - q_{\tilde{\nu},\omega,i}^{-} \right) \right) + \epsilon_k + \frac{\eta_k}{\tau_k} \left(2 q_{\nu,\omega,k}^{-} + \sum_{\tilde{\nu} \neq \nu} q_{\tilde{\nu},\omega,k}^{-} \right) \\
&\quad + \epsilon_m + \frac{2\eta_m}{\tau_m} \left(D_{\nu,m}(\omega) + q_{\nu,\omega,m}^{-} - \sum_{k=1}^{m-1} \left(q_{\nu,\omega,k}^{+} - q_{\nu,\omega,k}^{-} \right) \right) \\
&\quad + \frac{\eta_m}{\tau_m} \left(\sum_{\tilde{\nu} \neq \nu} \left(D_{\tilde{\nu},m}(\omega) + q_{\tilde{\nu},\omega,m}^{-} - \sum_{k=1}^{m-1} \left(q_{\tilde{\nu},\omega,k}^{+} - q_{\tilde{\nu},\omega,k}^{-} \right) \right) \right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\partial \widehat{\mathcal{C}}_{\nu}^{\mathrm{tr}(\omega)}}{\partial q_{\nu,\omega,m}^{-}}\left(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}\right) \\
&\quad = \epsilon_m + \frac{\eta_m}{\tau_m} \left(2 q_{\nu,\omega,m}^{-} + \sum_{\tilde{\nu} \neq \nu} q_{\tilde{\nu},\omega,m}^{-} \right) \\
&\quad + \epsilon_m + \frac{2\eta_m}{\tau_m} \left(D_{\nu,m}(\omega) + q_{\nu,\omega,m}^{-} - \sum_{k=1}^{m-1} \left(q_{\nu,\omega,k}^{+} - q_{\nu,\omega,k}^{-} \right) \right) \\
&\quad + \frac{\eta_m}{\tau_m} \left(\sum_{\tilde{\nu} \neq \nu} \left(D_{\tilde{\nu},m}(\omega) + q_{\tilde{\nu},\omega,m}^{-} - \sum_{k=1}^{m-1} \left(q_{\tilde{\nu},\omega,k}^{+} - q_{\tilde{\nu},\omega,k}^{-} \right) \right) \right).
\end{aligned}
$$

Since the cost function $\widehat{\mathcal{C}}_{\nu}^{\mathrm{tr}(\omega)}$ is a quadratic function, the Jacobian $\mathbf{J}^{\mathrm{tr}}$ of the truncated pseudo-gradient $\boldsymbol{\nabla}^{\mathrm{tr}(\omega)} \widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}$ defined in Equation (4.51) is independent of the realisation ω and of $\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}$,

and is given by

$$\mathbf{J}^{\mathrm{tr}} = \begin{bmatrix} \mathbf{J}^{\mathrm{tr}}_{=} & \mathbf{J}^{\mathrm{tr}}_{\neq} & \cdots & \cdots & \mathbf{J}^{\mathrm{tr}}_{\neq} \\ \mathbf{J}^{\mathrm{tr}}_{\neq} & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mathbf{J}^{\mathrm{tr}}_{\neq} \\ \mathbf{J}^{\mathrm{tr}}_{\neq} & \cdots & \cdots & \mathbf{J}^{\mathrm{tr}}_{\neq} & \mathbf{J}^{\mathrm{tr}}_{=} \end{bmatrix}, \tag{B.10}$$

where the sub-matrices $\mathbf{J}^{\mathrm{tr}}_{=}$ and $\mathbf{J}^{\mathrm{tr}}_{\neq}$ are defined as follows

$$\mathbf{J}^{\mathrm{tr}}_{=} = \begin{bmatrix} 2\,\mathbf{E} + 2\,\mathbf{E}_m - \mathbf{\Gamma} & -2\,\mathbf{E}_m + \mathbf{\Gamma} & -\frac{2\eta_m}{\tau_m}\,\mathbf{1}_{m-1} \\ -2\,\mathbf{E}_m + \mathbf{\Gamma} & 2\,\mathbf{E} + 2\,\mathbf{E}_m - \mathbf{\Gamma} & \frac{2\eta_m}{\tau_m}\,\mathbf{1}_{m-1} \\ -\frac{2\eta_m}{\tau_m}\,\mathbf{1}^{\mathrm{T}}_{m-1} & \frac{2\eta_m}{\tau_m}\,\mathbf{1}^{\mathrm{T}}_{m-1} & \frac{4\eta_m}{\tau_m} \end{bmatrix}, \tag{B.11}$$

and

$$\mathbf{J}^{\mathrm{tr}}_{\neq} = \begin{bmatrix} \mathbf{E} + \mathbf{E}_m - \mathbf{\Lambda} & -\mathbf{E}_m + \mathbf{\Lambda} & -\frac{\eta_m}{\tau_m}\,\mathbf{1}_{m-1} \\ -\mathbf{E}_m + \mathbf{\Lambda} & \mathbf{E} + \mathbf{E}_m - \mathbf{\Lambda} & \frac{\eta_m}{\tau_m}\,\mathbf{1}_{m-1} \\ -\frac{\eta_m}{\tau_m}\,\mathbf{1}^{\mathrm{T}}_{m-1} & \frac{\eta_m}{\tau_m}\,\mathbf{1}^{\mathrm{T}}_{m-1} & \frac{2\eta_m}{\tau_m} \end{bmatrix}, \tag{B.12}$$

where $\mathbf{E}$, $\mathbf{E}_m$, $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ are defined as in Definition 4.10. If we define $\mathbf{\Delta}^{\mathrm{tr}} = \frac{1}{2}\Big(\mathbf{J}^{\mathrm{tr}} + \big(\mathbf{J}^{\mathrm{tr}}\big)^{\mathrm{T}}\Big)$, then

$$\mathbf{\Delta}^{\mathrm{tr}} = \frac{1}{2}\begin{bmatrix} 2\bar{\mathbf{M}} & \bar{\mathbf{M}} & \cdots & \cdots & \bar{\mathbf{M}} \\ \bar{\mathbf{M}} & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \bar{\mathbf{M}} \\ \bar{\mathbf{M}} & \cdots & \cdots & \bar{\mathbf{M}} & 2\bar{\mathbf{M}} \end{bmatrix}, \tag{B.13}$$

with $\bar{\mathbf{M}}$ defined as in Definition 4.10. As by hypothesis $\bar{\mathbf{M}}$ is positive definite, $\mathbf{\Delta}^{\mathrm{tr}}$ is also positive definite with Proposition B.1 and has a Cholesky decomposition of the form $\mathbf{\Delta}^{\mathrm{tr}} = \big(\mathbf{C}^{\mathrm{tr}}\big)^{\mathrm{T}}\,\mathbf{C}^{\mathrm{tr}}$. We now show that $\mathbf{\Delta}^{\mathrm{tr}} \succeq 0$ implies that truncated pseudo-gradient map $\boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}$ is strongly monotone on $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$, *i.e.* that there exists a $\mu > 0$ such that for any distinct points $\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(1)} \neq \hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(2)}$ in $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$, we have that the following inequality holds:

$$\left\langle \boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}\Big(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(2)}\Big) - \boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}\Big(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(1)}\Big)\,,\, \hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(2)} - \hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(1)} \right\rangle \geq \mu \left\| \hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(2)} - \hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(1)} \right\|^2. \tag{B.14}$$

To do so, we define $\hat{\boldsymbol{y}}(\theta) = \theta\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(2)} + (1-\theta)\,\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(1)}$. Since $\widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$ is convex, $\hat{\boldsymbol{y}}(\theta) \in \widehat{\mathcal{Y}}^{\mathrm{tr}(\omega)}$ for any $\theta \in [0,1]$. Using the fundamental theorem of calculus, we have that

$$\frac{\mathrm{d}\Big(\boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}(\hat{\boldsymbol{y}}(\theta))\Big)}{\mathrm{d}\theta} = \mathbf{J}^{\mathrm{tr}}\,\frac{\mathrm{d}\hat{\boldsymbol{y}}(\theta)}{\mathrm{d}\theta} = \mathbf{J}^{\mathrm{tr}}\Big(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(2)} - \hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(1)}\Big). \tag{B.15}$$

This implies that

$$\boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}\Big(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(2)}\Big) - \boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}\Big(\hat{\boldsymbol{y}}^{\mathrm{tr}(\omega)}_{(1)}\Big) = \int_0^1 \frac{\mathrm{d}\Big(\boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}(\hat{\boldsymbol{y}}(\theta))\Big)}{\mathrm{d}\theta}\,\mathrm{d}\theta$$

$$= \mathbf{J}^{\mathrm{tr}}\left(\hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)}\right), \tag{B.16}$$

which in turn implies when taking the dot product with $\hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)}$ that

$$\begin{aligned}
&\left\langle \boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)}\right) - \boldsymbol{\nabla}^{\mathrm{tr}(\omega)}\widehat{\mathcal{C}}^{\mathrm{tr}(\omega)}\left(\hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)}\right),\ \hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)} \right\rangle \\
&= \left\langle \mathbf{J}^{\mathrm{tr}}\left(\hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)}\right),\ \hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)} \right\rangle \\
&= \left\langle \boldsymbol{\Delta}^{\mathrm{tr}}\left(\hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)}\right),\ \hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)} \right\rangle \\
&= \left\| \mathbf{C}^{\mathrm{tr}}\left(\hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)}\right) \right\|_2^2 \\
&\geq \left(\sigma_{\min}\left(\mathbf{C}^{\mathrm{tr}}\right)\right)^2 \left\| \hat{\boldsymbol{y}}_{(2)}^{\mathrm{tr}(\omega)} - \hat{\boldsymbol{y}}_{(1)}^{\mathrm{tr}(\omega)} \right\|^2 \geq 0,
\end{aligned} \tag{B.17}$$

where $\sigma_{\min}\left(\mathbf{C}^{\mathrm{tr}}\right)$ is the smallest eigenvalue value of $\mathbf{C}^{\mathrm{tr}}$. Finally, we have that

$$\left(\sigma_{\min}\left(\mathbf{C}^{\mathrm{tr}}\right)\right)^2 = \sigma_{\min}\left(\boldsymbol{\Delta}^{\mathrm{tr}}\right) = \lambda_{\min}\left(\boldsymbol{\Delta}^{\mathrm{tr}}\right) = \frac{\lambda_{\min}\left(\bar{\mathbf{M}}\right)}{2}, \tag{B.18}$$

where the last equality holds with Corollary B.1 and the definition of $\boldsymbol{\Delta}^{\mathrm{tr}}$. This completes the proof. □

Appendix C

Support materials related to Chapter 5

C.1 Auxiliary results

PROPOSITION C.1. *If* $\boldsymbol{y} \in \mathbf{K}$ *and* $\boldsymbol{y} \notin \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$, $\boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y})$ *and* $\alpha > 0$, *then*

$$\|\boldsymbol{y} - \boldsymbol{y}^\star\| \leq \left(\frac{L_{\boldsymbol{y}^\star,\delta} + \frac{1}{\alpha}}{\mu}\right) \left\|\boldsymbol{\Xi}^{\mathbf{nat}}_{\mathbf{K}}(\boldsymbol{y}, \alpha\boldsymbol{u})\right\|. \tag{C.1}$$

PROOF. The proof is similar to Facchinei and Pang (2003a, Theorem 2.3.3(c)). To lighten the notation, we define $\boldsymbol{r} \coloneqq \boldsymbol{\Xi}^{\mathbf{nat}}_{\mathbf{K}}(\boldsymbol{y}, \alpha\boldsymbol{u})$. Hence by the definition of the natural map, we have with (5.22) that

$$\boldsymbol{y} - \boldsymbol{r} = \Pi_{\mathbf{K}}(\boldsymbol{y} - \alpha\boldsymbol{u})\,. \tag{C.2}$$

Also, with the variational formulation of the orthogonal projection defined in (5.7), we obtain that

$$\forall \boldsymbol{z} \in \mathcal{Y}: \quad \langle \boldsymbol{z} - (\boldsymbol{y} - \boldsymbol{r})\,, (\boldsymbol{y} - \boldsymbol{r}) - (\boldsymbol{y} - \alpha\boldsymbol{u})\rangle \geq 0 \quad \Rightarrow \quad \langle \boldsymbol{z} - \boldsymbol{y} + \boldsymbol{r}\,,\, \alpha\boldsymbol{u} - \boldsymbol{r}\rangle \geq 0, \tag{C.3}$$

This holds for $\boldsymbol{z} = \boldsymbol{y}^\star$, which means that

$$\langle \boldsymbol{y}^\star - \boldsymbol{y} + \boldsymbol{r}\,,\, \alpha\boldsymbol{u} - \boldsymbol{r}\rangle \geq 0 \Leftrightarrow \alpha\,\langle \boldsymbol{y}^\star - \boldsymbol{y} + \boldsymbol{r}\,,\, \boldsymbol{u}\rangle - \langle \boldsymbol{y}^\star - \boldsymbol{y} + \boldsymbol{r}\,,\, \boldsymbol{r}\rangle \geq 0, \tag{C.4}$$

which in turn implies that

$$\alpha\,\langle \boldsymbol{y}^\star - \boldsymbol{y} + \boldsymbol{r}\,,\, \boldsymbol{u}\rangle - \langle \boldsymbol{y}^\star - \boldsymbol{y}\,,\, \boldsymbol{r}\rangle \geq 0. \tag{C.5}$$

Moreover, since $\boldsymbol{y}^\star \in \mathrm{SOL}(\mathrm{GVI}(\mathbf{K}, \boldsymbol{\Xi}))$, there exists a $\boldsymbol{u}^\star \in \boldsymbol{\Xi}(\boldsymbol{y}^\star)$ such that the following inequality holds:

$$\langle \boldsymbol{u}^\star\,,\, \boldsymbol{y} - \boldsymbol{r} - \boldsymbol{y}^\star\rangle \geq 0 \Leftrightarrow \alpha\,\langle \boldsymbol{u}^\star\,,\, \boldsymbol{y} - \boldsymbol{r} - \boldsymbol{y}^\star\rangle \geq 0, \tag{C.6}$$

where we used (5.1) since $\boldsymbol{y} - \boldsymbol{r} \in \mathbf{K}$. When combining Equations (C.5) and (C.6), we obtain

$$\langle \boldsymbol{u} - \boldsymbol{u}^\star\,,\, \boldsymbol{y} - \boldsymbol{y}^\star\rangle \leq \langle \boldsymbol{u} - \boldsymbol{u}^\star\,,\, \boldsymbol{r}\rangle + \frac{1}{\alpha}\,\langle \boldsymbol{y} - \boldsymbol{y}^\star\,,\, \boldsymbol{r}\rangle \tag{C.7}$$

Since $\boldsymbol{\Xi}$ is strongly monotone on $\mathbf{K}$ with parameter μ, the left-hand side of (C.7) can be lower bounded by $\mu \|\boldsymbol{y} - \boldsymbol{y}^\star\|^2$. If $\boldsymbol{y} \notin \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$ then the right-hand side of (C.7) can be upper bounded by

$$\left(L_{\boldsymbol{y}^\star,\delta} + \frac{1}{\alpha}\right) \|r\| \, \|\boldsymbol{y} - \boldsymbol{y}^\star\|$$

using the δ-far-sighted Lipschitz continuity at the point $\boldsymbol{y}^\star$. This implies (C.1) and completes the proof. □

Corollary C.1. *If $\boldsymbol{y} \in \mathbf{K}$ and $\boldsymbol{y} \notin \mathcal{B}(\boldsymbol{y}^\star\,;\,\delta)$ and $\boldsymbol{u} \in \boldsymbol{\Xi}(\boldsymbol{y})$, then*

$$\|\boldsymbol{y} - \boldsymbol{y}^\star\| \leq \alpha \frac{L_{\boldsymbol{y}^\star,\delta} + \frac{1}{\alpha}}{\mu} \|\boldsymbol{u}\| \,. \tag{C.8}$$

Proof. This holds with Proposition C.1 and the fact that the norm of the natural map $\boldsymbol{\Xi}^{\mathbf{nat}}_{\mathbf{K}}(\boldsymbol{y}, \alpha\boldsymbol{u})$ is bounded by $\alpha \|\boldsymbol{u}\|$ as shown as follows:

$$\left\|\boldsymbol{\Xi}^{\mathbf{nat}}_{\mathbf{K}}(\boldsymbol{y}, \alpha\boldsymbol{u})\right\| = \|\boldsymbol{y} - \Pi_{\mathbf{K}}(\boldsymbol{y} - \alpha\boldsymbol{u})\| = \|\Pi_{\mathbf{K}}(\boldsymbol{y}) - \Pi_{\mathbf{K}}(\boldsymbol{y} - \alpha\boldsymbol{u})\| \overset{(5.8)}{\leq} \|\boldsymbol{y} - (\boldsymbol{y} - \alpha\boldsymbol{u})\| = \alpha \|\boldsymbol{u}\| \,. \tag{C.9}$$

□

Proposition C.2. *Let $\mathbf{K} \subseteq \mathbb{R}^d$ be a closed convex set, $\alpha > 0$, $\boldsymbol{y} \in \mathbf{K}$, and $\boldsymbol{u} \in \mathbb{R}^d$. Let $\boldsymbol{y}^\star_{\mathbb{R}^d}$ be the maximizer of the following unconstrained optimisation problem*

$$\underset{\boldsymbol{z} \,\in\, \mathbb{R}^d}{\text{maximise}} \quad \langle \boldsymbol{u} \,,\, \boldsymbol{y} - \boldsymbol{z} \rangle - \frac{1}{2\alpha} \|\boldsymbol{y} - \boldsymbol{z}\|_2^2 \,, \tag{C.10}$$

i.e. $\boldsymbol{y}^\star_{\mathbb{R}^d} = \boldsymbol{y} - \alpha\boldsymbol{u}$. *If $\boldsymbol{y}^\star_{\mathbf{K}}$ denotes the maximizer of the following optimisation problem*

$$\underset{\boldsymbol{z} \,\in\, \mathbf{K}}{\text{maximise}} \quad \langle \boldsymbol{u} \,,\, \boldsymbol{y} - \boldsymbol{z} \rangle - \frac{1}{2\alpha} \|\boldsymbol{y} - \boldsymbol{z}\|_2^2 \,, \tag{C.11}$$

then

$$\boldsymbol{y}^\star_{\mathbf{K}} = \Pi_{\mathbf{K}}\left(\boldsymbol{y}^\star_{\mathbb{R}^d}\right) = \Pi_{\mathbf{K}}(\boldsymbol{y} - \alpha\boldsymbol{u}) \,. \tag{C.12}$$

Proof. Since $\boldsymbol{y}^\star_{\mathbf{K}}$ is the optimal solution of (C.11), the following assertion holds

$$\forall \boldsymbol{z} \in \mathbf{K} : \quad \left\langle -\boldsymbol{u} + \frac{1}{\alpha}(\boldsymbol{y} - \boldsymbol{y}^\star_{\mathbf{K}}) \;,\; \boldsymbol{z} - \boldsymbol{y}^\star_{\mathbf{K}} \right\rangle \leq 0. \tag{C.13}$$

Then, by the variational inequality formulation of the orthogonal projection, we have with (5.7) that $\bar{\boldsymbol{y}}$ is the orthogonal projection of $\boldsymbol{y}^\star_{\mathbb{R}^d}$ on $\mathbf{K}$ if and only if it satisfies

$$\forall \boldsymbol{z} \in \mathbf{K} : \quad \left\langle \boldsymbol{z} - \bar{\boldsymbol{y}} \,,\, \bar{\boldsymbol{y}} - \boldsymbol{y}^\star_{\mathbb{R}^d} \right\rangle \geq 0. \tag{C.14}$$

It can be shown that the solution of the unconstrained Optimisation Problem (C.10) has the closed form formula $\boldsymbol{y}^\star_{\mathbb{R}^d} = \boldsymbol{y} - \alpha\boldsymbol{u}$, which gives with (C.14) that $\bar{\boldsymbol{y}} = \Pi_{\mathbf{K}}\left(\boldsymbol{y}^\star_{\mathbb{R}^d}\right)$ if and only if it satisfies

$$\forall \boldsymbol{z} \in \mathbf{K} : \quad \langle \boldsymbol{z} - \bar{\boldsymbol{y}} \,,\, \bar{\boldsymbol{y}} - (\boldsymbol{y} - \alpha\boldsymbol{u}) \rangle \geq 0 \Leftrightarrow \left\langle \boldsymbol{z} - \bar{\boldsymbol{y}} \,,\, \frac{1}{\alpha}(\bar{\boldsymbol{y}} - \boldsymbol{y}) + \boldsymbol{u} \right\rangle \geq 0. \tag{C.15}$$

This completes the proof since taking $\bar{\boldsymbol{y}} = \boldsymbol{y}^\star_{\mathbf{K}}$ satisfies (C.15) given (C.13). □

LEMMA C.1. *Let $\boldsymbol{\Xi} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be a μ-strongly monotone and δ-far-sighted Lipschitz continuous set-valued map on $\mathbf{K}$. Let $\boldsymbol{w} \in \mathbb{R}^d$. Then the set-valued map $\hat{\boldsymbol{\Xi}}_{\boldsymbol{w}} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d; \boldsymbol{x} \mapsto \boldsymbol{\Xi}(\boldsymbol{x}) - \boldsymbol{w}$ is also a μ-strongly monotone and δ-far-sighted Lipschitz continuous set-valued map on $\mathbf{K}$.*

PROOF. The proof is straightforward and left to the reader. □

C.2 Proofs

PROPOSITION C.3. *Let $r \geq 1$, $\delta > 0$ and $\mathbf{K}$ be a compact convex set. If the set-valued map $\boldsymbol{\Xi}$ is μ-strongly monotone on $\mathbf{K}^{[\delta]}$ and δ-far-sighted Lipschitz continuous with constant L on $\mathbf{K}^{[\delta]}$ at any point $\boldsymbol{x} \in \mathbf{K}$. Then $\boldsymbol{\Xi}$ is $\frac{\delta}{r}$-far-sighted Lipschitz continuous with constant rL on $\mathbf{K}^{[\delta]}$ at any point $\boldsymbol{x} \in \mathbf{K}$.*

PROOF. We must show for any $\boldsymbol{x} \in \mathbf{K}$ and $\boldsymbol{v} \in \boldsymbol{\Xi}(\boldsymbol{x})$ that the following inequality holds true for all $\boldsymbol{z} \in \mathbf{K}^{[\delta]} \setminus \mathcal{B}\left(\boldsymbol{x}\,;\,\frac{\delta}{r}\right)$ and $\boldsymbol{w} \in \boldsymbol{\Xi}(\boldsymbol{z})$:

$$\|\boldsymbol{w} - \boldsymbol{v}\| \leq rL \|\boldsymbol{z} - \boldsymbol{x}\| . \tag{C.16}$$

If $\|\boldsymbol{z} - \boldsymbol{x}\| \geq \delta$, then (C.16) follows from the δ-far-sighted Lipschitz continuity of $\boldsymbol{\Xi}$ on $\mathbf{K}^{[\delta]}$ at $\boldsymbol{x}$, *i.e.*

$$\|\boldsymbol{w} - \boldsymbol{v}\| \leq L \|\boldsymbol{z} - \boldsymbol{x}\| < rL \|\boldsymbol{z} - \boldsymbol{x}\| . \tag{C.17}$$

It remains to show that (C.16) holds when $\boldsymbol{z} \in \left(\mathbf{K}^{[\delta]} \cap \mathcal{B}(\boldsymbol{x}\,;\,\delta)\right) \setminus \mathcal{B}\left(\boldsymbol{x}\,;\,\frac{\delta}{r}\right)$. First, let $\boldsymbol{d}$ denote the direction of $\boldsymbol{w} - \boldsymbol{v}$, *i.e.* $\boldsymbol{d} = \frac{\boldsymbol{w}-\boldsymbol{v}}{\|\boldsymbol{w}-\boldsymbol{v}\|}$; we assume w.l.o.g. that $\boldsymbol{w} \neq \boldsymbol{v}$. Since $\boldsymbol{x} \in \mathbf{K}$, $\mathcal{B}[\boldsymbol{x}\,;\,\delta] \subseteq \mathbf{K}^{[\delta]}$. Hence, as $\boldsymbol{z} \in \mathcal{B}(\boldsymbol{x}\,;\,\delta)$, there exist two points $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(2)}$ in $\mathbf{K}^{[\delta]}$ on the boundary of $\mathcal{B}[\boldsymbol{x}\,;\,\delta]$ (see Figure C.1), such that

$$\boldsymbol{y}^{(2)} - \boldsymbol{y}^{(1)} = \kappa \boldsymbol{d} \text{ with } \kappa \geq 0 , \quad \textit{and} \quad \boldsymbol{z} = \varsigma \boldsymbol{y}^{(1)} + (1 - \varsigma)\, \boldsymbol{y}^{(2)} \text{ with } \varsigma \in [0, 1]. \tag{C.18}$$

Let $\hat{\boldsymbol{\Xi}}_{\boldsymbol{v}}$ be the set-valued map defined as $\hat{\boldsymbol{\Xi}}_{\boldsymbol{v}} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d; \boldsymbol{x} \mapsto \boldsymbol{\Xi}(\boldsymbol{x}) - \boldsymbol{v}$. With Lemma C.1, $\hat{\boldsymbol{\Xi}}_{\boldsymbol{v}}$ is μ-strongly monotone on $\mathbf{K}^{[\delta]}$, which implies with (5.24) that, for any $\boldsymbol{u}^{(1)} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(1)}\right)$ and $\boldsymbol{u}^{(2)} \in \boldsymbol{\Xi}\left(\boldsymbol{y}^{(2)}\right)$, there exists a value $\lambda \in (0,\, 1)$ such that

$$\boldsymbol{w} - \boldsymbol{v} = \lambda\, \Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(1)} - \boldsymbol{v}\right) + (1 - \lambda)\, \Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(2)} - \boldsymbol{v}\right) . \tag{C.19}$$

Hence

$$\begin{aligned}
\|\boldsymbol{w} - \boldsymbol{v}\| &= \left\|\lambda\, \Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(1)} - \boldsymbol{v}\right) + (1 - \lambda)\, \Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(2)} - \boldsymbol{v}\right)\right\| && \text{(C.20a)}\\
&= \lambda \left\|\Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(1)} - \boldsymbol{v}\right)\right\| + (1 - \lambda) \left\|\Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(2)} - \boldsymbol{v}\right)\right\| && \text{(C.20b)}\\
&\overset{(5.24)}{<} \max\left\{\left\|\Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(1)} - \boldsymbol{v}\right)\right\|, \left\|\Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(2)} - \boldsymbol{v}\right)\right\|\right\} . && \text{(C.20c)}
\end{aligned}$$

We assume w.l.o.g. that $\left\|\Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(1)} - \boldsymbol{v}\right)\right\| > \left\|\Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(2)} - \boldsymbol{v}\right)\right\|$. Then, (C.20c) implies with the δ-far-sighted Lipschitz continuity of $\boldsymbol{\Xi}$ on $\mathbf{K}^{[\delta]}$ at $\boldsymbol{x}$ that:

$$\|\boldsymbol{w} - \boldsymbol{v}\| < \left\|\Pi_{\boldsymbol{d}}\left(\boldsymbol{u}^{(1)} - \boldsymbol{v}\right)\right\| \leq \left\|\boldsymbol{u}^{(1)} - \boldsymbol{v}\right\| \leq L \left\|\boldsymbol{y}^{(1)} - \boldsymbol{x}\right\| = L\delta \leq rL \|\boldsymbol{x} - \boldsymbol{z}\| , \tag{C.21}$$

where the last inequality holds since by hypothesis $\|\boldsymbol{x} - \boldsymbol{z}\| \geq \frac{\delta}{r}$. This concludes the proof. □

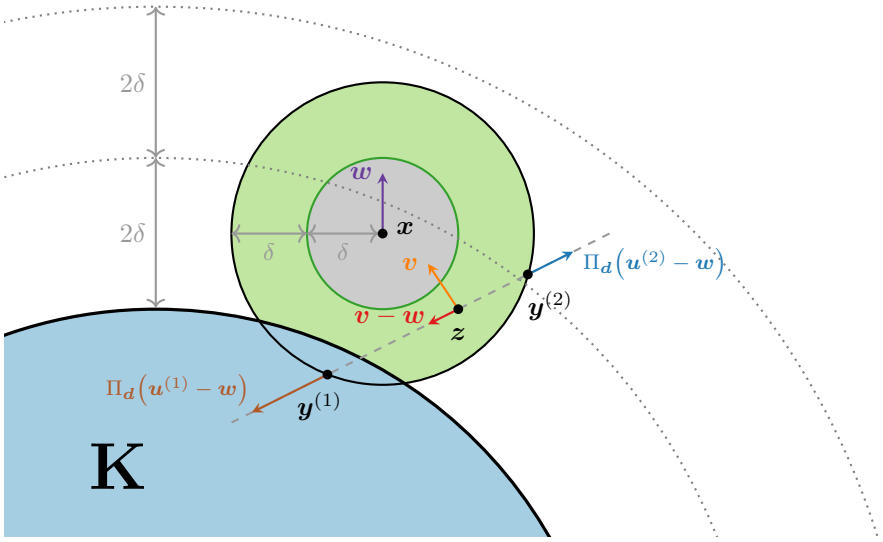


FIGURE C.1: Illustration of Proposition 5.11.

THEOREM C.1. *If* $\alpha^{(k+1)} \leq \frac{\mu}{\beta} S^{(k)}$ *then*

$$\Delta_{\mathcal{L}}^{(k+1)} \leq \Delta_{\mathcal{L}}^{(k)} + \frac{1}{2}\alpha^{(k+1)}\left[\frac{1}{\mu+\beta}\left\|\boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)}\right\|^2 + \frac{2\beta}{\mu+\beta}\frac{(1-\kappa)}{\kappa}\left\|\boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right\| + \frac{\beta^2}{\mu+\beta}\frac{(1-\kappa)^2}{\kappa^2}\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right\|^2 - \beta\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)}\right\|^2\right], \quad \text{(C.22)}$$

where $\Delta_{\mathcal{L}}^{(k)} = \max_{\boldsymbol{z}\in\mathcal{L}^{(k)}} \Psi_k(\boldsymbol{z})$.

PROOF. By proceeding in a similar way as for Theorem 5.10, we obtain the following inequality

$$\Delta_{\mathcal{L}}^{(k+1)} \leq \Delta_{\mathcal{L}}^{(k)} + \max_{\boldsymbol{z}\in\mathcal{L}^{(k)}}\left\{-\frac{1}{2}\mu S^{(k)}\left\|\boldsymbol{z} - \boldsymbol{x}^{(k)}\right\|^2 + \alpha^{(k+1)}\left(\left\langle \boldsymbol{u}^{(k+1)}, \boldsymbol{y}^{(k+1)} - \boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z} - \boldsymbol{y}^{(k+1)}\right\|^2\right)\right\}. \quad \text{(C.23)}$$

Similarly as done in the proof of Theorem 5.5, we focus in a first step on finding an upper-bound for

$$\begin{aligned}
&\left\langle \boldsymbol{u}^{(k+1)}, \boldsymbol{y}^{(k+1)} - \boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z} - \boldsymbol{y}^{(k+1)}\right\|^2 \\
&= \left\langle \boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)}, \boldsymbol{y}^{(k+1)} - \boldsymbol{z}\right\rangle + \left\langle \boldsymbol{v}^{(k)}, \boldsymbol{y}^{(k+1)} - \boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z} - \boldsymbol{y}^{(k+1)}\right\|^2 \\
&\leq \left\|\boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{z}\right\| + \left\langle \boldsymbol{v}^{(k)}, \boldsymbol{y}^{(k+1)} - \boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z} - \boldsymbol{y}^{(k+1)}\right\|^2 \\
&\leq \left\|\boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)} - \boldsymbol{z}\right\| + \left\langle \boldsymbol{v}^{(k)}, \boldsymbol{y}_\star^{(k+1)} - \boldsymbol{z}\right\rangle - \frac{1}{2}\mu\left\|\boldsymbol{z} - \boldsymbol{y}^{(k+1)}\right\|^2 + \left\langle \boldsymbol{v}^{(k)}, \boldsymbol{y}^{(k+1)} - \boldsymbol{y}_\star^{(k+1)}\right\rangle. \quad \text{(C.24)}
\end{aligned}$$

First, we derive an upper-bound for the first three terms of Equation (C.24). By definition, $\boldsymbol{y}_\star^{(k+1)}$ maximises $\psi^{\beta}_{\boldsymbol{x}^{(k)},\boldsymbol{v}^{(k)}}$ on $\mathcal{L}^{(k)}$, hence the following inequality holds:

$$\left\langle -\boldsymbol{v}^{(k)} - \beta\left(\boldsymbol{y}_\star^{(k+1)} - \boldsymbol{x}^{(k)}\right), \boldsymbol{z} - \boldsymbol{y}_\star^{(k+1)}\right\rangle \leq 0, \quad \forall \boldsymbol{z} \in \mathcal{L}^{(k)}. \quad \text{(C.25)}$$

This inequality and (5.70) lead to the following upper-bound:

$$\begin{aligned}
&\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|+\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{z}\right\rangle-\frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2\\
&\overset{(C.25)}{\leq}\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|\\
&\qquad\qquad+\beta\left\langle \boldsymbol{y}_\star^{(k+1)}-\boldsymbol{x}^{(k)}\,,\,\boldsymbol{z}-\boldsymbol{y}_\star^{(k+1)}\right\rangle-\frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2\\
&\overset{(5.70)}{=}\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|+\frac{\beta}{2}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2\\
&\qquad\qquad-\frac{\beta}{2}\left\|\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2-\frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2-\frac{\beta}{2}\left\|\boldsymbol{z}-\boldsymbol{y}_\star^{(k+1)}\right\|^2.
\end{aligned}\tag{C.26}$$

With the triangular inequality, we have that

$$\left\|\boldsymbol{z}-\boldsymbol{y}_\star^{(k+1)}\right\|=\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}+\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|\geq\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|-\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|,$$

which in turn implies that

$$\begin{aligned}
\left\|\boldsymbol{z}-\boldsymbol{y}_\star^{(k+1)}\right\|^2\geq\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2+\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2\\
-2\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|.
\end{aligned}\tag{C.27}$$

Hence, (C.26) can be rewritten as

$$\begin{aligned}
&\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|+\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{z}\right\rangle-\frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2\\
&\overset{(C.27)}{\leq}\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|+\frac{\beta}{2}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2-\frac{\beta}{2}\left\|\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2\\
&\quad-\frac{(\mu+\beta)}{2}\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2-\frac{\beta}{2}\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2+\beta\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|\\
&=\left(\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|+\beta\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|\right)\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|-\frac{(\mu+\beta)}{2}\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2\\
&\qquad+\frac{\beta}{2}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2-\frac{\beta}{2}\left\|\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2-\frac{\beta}{2}\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2.
\end{aligned}\tag{C.28}$$

The first two terms of (C.28) can be seen as the polynomial $c_1r-\frac{1}{2}c_2r^2$ with $c_1=\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|+\beta\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|$, $c_2=\mu+\beta$, and $r=\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|$. This polynomial reaches its maximum at $r^\star=c_1/c_2$ and can be upper-bounded by $(c_1^2/2c_2)$. Hence

$$\begin{aligned}
&\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|+\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{z}\right\rangle-\frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2\\
&\leq\frac{1}{2(\mu+\beta)}\Big(\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|^2+2\beta\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|\\
&+\beta^2\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2\Big)+\frac{\beta}{2}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2-\frac{\beta}{2}\left\|\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2-\frac{\beta}{2}\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2.
\end{aligned}\tag{C.29}$$

Using the equalities

$$\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{x}^{(k)}=\frac{1}{\kappa}\Big(\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\Big)\tag{C.30}$$

$$\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}=\frac{(1-\kappa)}{\kappa}\Big(\boldsymbol{x}^{(k)}-\boldsymbol{y}^{(k+1)}\Big),\tag{C.31}$$

we can finally reformulate (C.29) in terms of $\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|$ as follows

$$\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{z}\right\|+\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{z}\right\rangle-\frac{1}{2}\mu\left\|\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\|^2$$

$$\leq \frac{1}{2(\mu+\beta)}\Big(\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|^2 + \frac{2\beta(1-\kappa)}{\kappa}\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{v}^{(k)}\right\|\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\| + \frac{\beta^2(1-\kappa)^2}{\kappa^2}\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2\Big) + \frac{\beta}{2}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2 - \frac{\beta}{2\kappa^2}\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2 - \frac{\beta(1-\kappa)^2}{2\kappa^2}\left\|\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|^2. \tag{C.32}$$

With (C.32), we have found an upper-bound for the first three terms of (C.24). We now focus on finding an upper-bound for the last term of (C.24), *i.e.* $\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle$. To do so, given a $c>0$, we consider the following optimisation problem:

$$\underset{t\,\in\,\mathbb{R}}{\text{maximise}} \quad \left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle t - \left(\frac{\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle}{\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2}\right)\frac{t^2}{2c}\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2. \tag{C.33}$$

Based on the definition $\boldsymbol{y}_\star^{(k+1)}$, we have that $\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle \geq 0$, which implies that Optimisation Problem (C.33) reaches its maximum for $t_\star(c)=c$. We now define the set

$$\hat{\mathcal{L}}^{(k)} \coloneqq \left\{ \boldsymbol{z}\in\mathcal{L}^{(k)} \;\middle|\; \exists t\in\mathbb{R}:\ \boldsymbol{x}^{(k)}-\boldsymbol{z}=t\Big(\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\Big)\right\}, \tag{C.34}$$

and the optimisation problem

$$\underset{\boldsymbol{z}\,\in\,\hat{\mathcal{L}}^{(k)}}{\text{maximise}} \quad \left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{z}\right\rangle - \left(\frac{\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle}{\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2}\right)\frac{1}{2c}\left\|\boldsymbol{z}-\boldsymbol{x}^{(k)}\right\|^2. \tag{C.35}$$

By comparing Optimisation Problems (C.33) and (C.35), it is straightforward to show that the maximum of Optimisation Problem (C.35) is attained for $\boldsymbol{z}_\star(c)=\boldsymbol{x}^{(k)}-c\Big(\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\Big)$. In particular, $\boldsymbol{z}_\star(\kappa)=\boldsymbol{x}^{(k)}-\kappa\Big(\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\Big)=\boldsymbol{y}^{(k+1)}$ when $c=\kappa$. The variational equivalent formulation of the optimality of $\boldsymbol{y}^{(k+1)}$ for $c=\kappa$ for Optimisation Problems (C.35), which corresponds to the first order necessary condition, can be expressed as follows

$$\left\langle -\boldsymbol{v}^{(k)} - \frac{1}{\kappa}\left(\frac{\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle}{\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2}\right)\Big(\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\Big)\,,\,\boldsymbol{z}-\boldsymbol{y}^{(k+1)}\right\rangle \leq 0, \quad \forall \boldsymbol{z}\in\hat{\mathcal{L}}^{(k)}. \tag{C.36}$$

In particular, (C.36) holds for $\boldsymbol{z}=\boldsymbol{y}_\star^{(k+1)}$, which gives

$$\left\langle -\boldsymbol{v}^{(k)} - \frac{1}{\kappa}\left(\frac{\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle}{\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2}\right)\Big(\boldsymbol{y}^{(k+1)}-\boldsymbol{x}^{(k)}\Big)\,,\,\boldsymbol{y}_\star^{(k+1)}-\boldsymbol{y}^{(k+1)}\right\rangle \leq 0. \tag{C.37}$$

Hence, the last term (C.24) can be bounded as follows:

$$\begin{aligned}\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle &\overset{(C.37)}{\leq} \frac{1}{\kappa}\left(\frac{\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle}{\left\|\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\|^2}\right)\left\langle \boldsymbol{x}^{(k)}-\boldsymbol{y}^{(k+1)}\,,\,\boldsymbol{y}^{(k+1)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle \\ &= \frac{(1-\kappa)}{\kappa}\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}^{(k+1)}\right\rangle = (1-\kappa)\left\langle \boldsymbol{v}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle \\ &\overset{(C.25)}{\leq} -\beta(1-\kappa)\left\langle \boldsymbol{y}_\star^{(k+1)}-\boldsymbol{x}^{(k)}\,,\,\boldsymbol{x}^{(k)}-\boldsymbol{y}_\star^{(k+1)}\right\rangle\end{aligned}$$

$$= \beta \frac{(1-\kappa)}{\kappa^2} \left\| \boldsymbol{x}^{(k)} - \boldsymbol{y}^{(k+1)} \right\|^2. \tag{C.38}$$

Hence, by combining (C.32) and (C.38), we can bound (C.24) as follows

$$\begin{aligned} \left\langle \boldsymbol{u}^{(k+1)}, \boldsymbol{y}^{(k+1)} - \boldsymbol{z} \right\rangle - \frac{1}{2}\mu \left\| \boldsymbol{z} - \boldsymbol{y}^{(k+1)} \right\|^2 \leq \frac{\beta}{2} \left\| \boldsymbol{z} - \boldsymbol{x}^{(k)} \right\|^2 - \frac{\beta}{2} \left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|^2 \\ + \frac{1}{2(\mu+\beta)} \Bigg(\left\| \boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)} \right\|^2 + \frac{2\beta(1-\kappa)}{\kappa} \left\| \boldsymbol{u}^{(k+1)} - \boldsymbol{v}^{(k)} \right\| \left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\| \\ + \frac{\beta^2 (1-\kappa)^2}{\kappa^2} \left\| \boldsymbol{y}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|^2 \Bigg). \end{aligned} \tag{C.39}$$

This completes the proof since Equation (C.22) is obtained by substituting (C.39) into (C.23) and using the assumption that $\alpha^{(k+1)} \leq \frac{\mu}{\beta} S^{(k)}$. □

Bibliography

R. Almgren. Execution Costs. In *Encycl. Quant. Financ.*, pages 51–66. John Wiley & Sons, Ltd, 2010. doi:10.1002/9780470061602.eqf18002.

R. Almgren. Optimal trading with stochastic liquidity and volatility. *SIAM J. Financ. Math.*, 3(1): 163–181, 2012. doi:10.1137/090763470.

R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *J. Risk*, 3(2):5–39, 2001. doi:10.21314/jor.2001.041.

R. Almgren and J. Lorenz. Adaptive arrival price. 2007. URL https://www.semanticscholar.org/paper/Adaptive-Arrival-Price-Almgren-Lorenz/a0db429b08bc56ae9ae50789bee05c071a1d51bc.

R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct Estimation of Equity Market Impact. 2005. URL https://www.semanticscholar.org/paper/Direct-Estimation-of-Equity-Market-Impact-Almgren-Thum/00777edc168f26633de9f5b9ff4c4f74bd9790e3.

R. F. Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Appl. Math. Financ.*, 10(1):1–18, 2003. doi:10.1080/135048602100056.

R. W. Anderson and J.-P. Danthine. Hedging and Joint Production: Theory and Illustrations. *J. Finance*, 35(2):487, 1980. doi:10.2307/2327409.

K. J. Arrow and G. Debreu. Existence of an Equilibrium for a Competitive Economy. *Econometrica*, 22 (3):265, 1954. doi:10.2307/1907353.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent Measures of Risk. *Math. Financ.*, 9(3): 203–228, 1999. doi:10.1111/1467-9965.00068.

J.-P. Aubin and A. Cellina. *Differential Inclusions*, volume 264. Springer Berlin Heidelberg, 1984. doi:10.1007/978-3-642-69512-4.

J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Birkhäuser Boston, 2009. doi:10.1007/978-0-8176-4848-0.

M. T. Barlow. A diffusion model for electricity prices. *Math. Financ.*, 12(4):287–298, 2002. doi:10.1111/j.1467-9965.2002.tb00125.x.

H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer New York, 2011. doi:10.1007/978-1-4419-9467-7.

A. Bensoussan. Points de Nash dans le cas de fonctionnelles quadratiques et jeux differentiels linéaires à N Personnes. *SIAM J. Control*, 12(3):460–499, 1974. doi:10.1137/0312037.

T. G. Benton, A. Froggatt, L. Wellesley, O. Grafham, R. King, N. Morisetti, J. Nixey, and P. Schröder. The Ukraine war and threats to food and energy security. Technical report, 2022. URL https://www.chathamhouse.org/sites/default/files/2022-04/2022-04-12-ukraine-war-threats-food-energy-security-benton-et-al.pdf.

H. Bessembinder and M. L. Lemmon. Equilibrium pricing and optimal hedging in electricity forward markets. *J. Finance*, 57(3):1347–1382, 2002. doi:10.1111/1540-6261.00463.

M. Bianchi and A. W. Lo. Optimal control of execution costs. *J. Financ. Mark.*, 1(1):1–50, 1998. doi:10.1016/S1386-4181(97)00012-8.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. doi:10.1017/CBO9780511804441.

W. Bühler and J. Müller-Merbach. Valuation of Electricity Futures: Reduced-Form vs. Dynamic Equilibrium Models. *SSRN Electron. J.*, (July):3–5, 2011. doi:10.2139/ssrn.983862.

B. Bulthuis, J. Concha, T. Leung, and B. Ward. Optimal execution of limit and market orders with trade director, speed limiter, and fill uncertainty. *Int. J. Financ. Eng.*, 4(2-3):1750020, 2017. doi:10.1142/S2424786317500207.

R. Burachik and J. Dutta. Inexact Proximal Point Methods for Variational Inequality Problems. *SIAM J. Optim.*, 20(5):2653–2678, 2010. doi:10.1137/080733437.

S. Butenko, A. Golodnikov, and S. Uryasev. Optimal Security Liquidation Algorithms. *Comput. Optim. Appl.*, 32:9–27, 2005. doi:10.1007/s10589-005-2052-9.

B. I. Carlin, M. S. Lobo, and S. Viswanathan. Episodic liquidity crises: Cooperative and predatory trading. *J. Finance*, 62(5):2235–2274, 2007. doi:10.1111/j.1540-6261.2007.01274.x.

R. Carmona, M. Coulon, and D. Schwarz. Electricity price modeling and asset valuation: A multi-fuel structural approach. *Math. Financ. Econ.*, 7(2):167–202, 2013. doi:10.1007/s11579-012-0091-4.

M. Carrion, A. B. Philpott, A. J. Conejo, and J. M. Arroyo. A Stochastic Programming Approach to Electric Energy Procurement for Large Consumers. *IEEE Trans. Power Syst.*, 22(2):744–754, 2007. doi:10.1109/TPWRS.2007.895164.

Á. Cartea and M. G. Figueroa. Pricing in Electricity Markets: A Mean Reverting Jump Diffusion Model with Seasonality. *Appl. Math. Financ.*, 12(4):313–335, 2005. doi:10.1080/13504860500117503.

Á. Cartea and S. Jaimungal. Incorporating order-flow into optimal execution. *Math. Financ. Econ.*, 10 (3):339–364, 2016. doi:10.1007/s11579-016-0162-z.

Á. Cartea, S. Jaimungal, and J. Penalva. *Algorithmic and high-frequency trading*. Cambridge University Press, 2015. URL www.cambridge.org/9781107091146.

E. Cavazzuti, M. Pappalardo, and M. Passacantando. Nash Equilibria, Variational Inequalities, and Dynamical Systems. *J. Optim. Theory Appl.*, 114(3):491–506, 2002. doi:10.1023/A:1016056327692.

Y. Censor, A. Gibali, and S. Reich. The Subgradient Extragradient Method for Solving Variational Inequalities in Hilbert Space. *J. Optim. Theory Appl.*, 148(2):318–335, 2011. doi:10.1007/s10957-010-9757-3.

D. Chan and J. S. Pang. The Generalized Quasi-Variational Inequality Problem. *Math. Oper. Res.*, 7(2): 211–222, 1982. doi:10.1287/moor.7.2.211.

K. Chan, Y. P. Chung, and H. Johnson. The Intraday Behavior of Bid-Ask Spreads for NYSE Stocks and CBOE Options. *J. Financ. Quant. Anal.*, 30(3):329, 1995. doi:10.2307/2331344.

X. Cheng, M. Di Giacinto, and T.-H. Wang. Optimal execution with uncertain order fills in Almgren–Chriss framework. *Quant. Financ.*, 17(1):55–69, 2017. doi:10.1080/14697688.2016.1185531.

X. Cheng, M. Di Giacinto, and T.-H. Wang. Optimal execution with dynamic risk adjustment. *J. Oper. Res. Soc.*, 70(10):1662–1677, 2019. doi:10.1080/01605682.2019.1644143.

P. Cheridito, F. Delbaen, and M. Kupper. Dynamic Monetary Risk Measures for Bounded Discrete-Time Processes. *Electron. J. Probab.*, 11(none):57–106, 2006. doi:10.1214/EJP.v11-302.

J. Cortes. Discontinuous dynamical systems. *IEEE Control Syst.*, 28(3):36–73, 2008. doi:10.1109/MCS.2008.919306.

M. Coulon and S. Howison. Stochastic behavior of the electricity bid stack: from fundamental drivers to power prices. *J. Energy Mark.*, 2(1):29–69, 2009. doi:10.21314/jem.2009.032.

G. Curato, J. Gatheral, and F. Lillo. Optimal execution with non-linear transient market impact. *Quant. Financ.*, 17(1):41–54, 2017. doi:10.1080/14697688.2016.1181274.

G. de Maere d'Aertrycke and Y. Smeers. Liquidity risks on power exchange. 2010.

G. de Maere d'Aertrycke and Y. Smeers. Liquidity risks on power exchanges: a generalized Nash equilibrium model. *Math. Program.*, 140(2):381–414, 2013. doi:10.1007/s10107-013-0694-4.

G. Debreu. A Social Equilibrium Existence Theorem. *Proc. Natl. Acad. Sci.*, 38(10):886–893, 1952. doi:10.1073/pnas.38.10.886.

C. Dutang. Existence theorems for generalized Nash equilibrium problems. 4(2):115–126, 2013. URL `https://hal.archives-ouvertes.fr/hal-00828948`.

J. Eckstein and D. P. Bertsekas. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55(1-3):293–318, 1992. doi:10.1007/BF01581204.

F. Espen Benth and T. Meyer-Brandis. The information premium for non-storable commodities. *J. Energy Mark.*, 2(3):111–140, 2009. doi:10.21314/jem.2009.021.

F. Facchinei and C. Kanzow. Generalized Nash Equilibrium Problems. *Ann. Oper. Res.*, 175(1):177–211, 2010. doi:10.1007/s10479-009-0653-x.

F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I.* Springer New York, 2003a. doi:10.1007/b97543.

F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume II*. Springer-Verlag, New York, 2003b. doi:10.1007/b97544.

F. Facchinei and J.-S. Pang. Nash equilibria: the variational approach. In *Convex Optim. Signal Process. Commun.*, pages 443–493. Cambridge University Press, 2009. doi:10.1017/CBO9780511804458.013.

Y. Feng, F. Rubio, and D. P. Palomar. Optimal order execution for algorithmic trading: A CVaR approach. In *2012 IEEE 13th Int. Work. Signal Process. Adv. Wirel. Commun.*, pages 480–484. IEEE, 2012. doi:10.1109/SPAWC.2012.6292954.

M. Ferris and A. Philpott. Dynamic Risked Equilibrium. *Oper. Res.*, 70(3):1933–1952, 2022. doi:10.1287/opre.2019.1958.

A. Fischer, M. Herrich, and K. Schönefeld. Generalized Nash equilibrium problems - Recent advances and challenges. *Pesqui. Operacional*, 34(3):521–558, 2014. doi:10.1590/0101-7438.2014.034.03.0521.

H. Föllmer and A. Schied. *Stochastic Finance*. De Gruyter, 2016. doi:10.1515/9783110463453.

P. A. Forsyth, J. S. Kennedy, S. T. Tse, and H. Windcliff. Optimal trade execution: A mean quadratic variation approach. *J. Econ. Dyn. Control*, 36(12):1971–1991, 2012. doi:10.1016/j.jedc.2012.05.007.

M. Fukushima. Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Math. Program.*, 53(1-3):99–110, 1992. doi:10.1007/BF01585696.

J. Gatheral and A. Schied. Optimal trade execution under geometric brownian motion in the Almgren and Chriss framework. *Int. J. Theor. Appl. Financ.*, 14(3):353–368, 2011. doi:10.1142/S0219024911006577.

J. Gatheral and A. Schied. Dynamical Models of Market Impact and Algorithms for Order Execution. *SSRN Electron. J.*, pages 579–599, 2012. doi:10.2139/ssrn.2034178.

H. Gérard, V. Leclère, and A. Philpott. On risk averse competitive equilibrium. *Oper. Res. Lett.*, 46(1): 19–26, 2018. doi:10.1016/j.orl.2017.10.011.

R. Green. Electricity liberalisation in Europe-how competitive will it be? *Energy Policy*, 34(16):2532–2541, 2006a. doi:10.1016/j.enpol.2004.08.016.

R. Green. Market power mitigation in the UK power market. *Util. Policy*, 14(2):76–89, 2006b. doi:10.1016/j.jup.2005.09.001.

D. A. Gregory. Upper Semi-Continuity of Subdifferential Mappings. *Can. Math. Bull.*, 23(1):11–19, 1980. doi:10.4153/CMB-1980-002-9.

O. Gueant. *The Financial Mathematics of Market Liquidity*. Chapman and Hall/CRC, 2016. doi:10.1201/b21350.

L. Gurobi Optimization. Gurobi Optimizer Reference Manual, 2022. URL `https://www.gurobi.com`.

B. He. Inexact implicit methods for monotone general variational inequalities. *Math. Program.*, 86(1): 199–217, 1999. doi:10.1007/s101070050086.

G. Huberman and W. Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4):1247–1275, 2004. doi:10.1111/j.1468-0262.2004.00531.x.

G. Huberman and W. Stanzl. Optimal liquidity trading. *Rev. Financ.*, 9(2):165–200, 2005. doi:10.1007/s10679-005-7591-5.

V. S. Kirilyuk. The Class of Polyhedral Coherent Risk Measures. *Cybern. Syst. Anal.*, 40(4):599–609, 2004. doi:10.1023/B:CASA.0000047881.82280.e2.

R. Kissell and R. Malamut. Algorithmic Decision-Making Framework. *J. Trading*, 1(1):12–21, 2005. doi:10.3905/jot.2006.609171.

G. M. Korpelevič. An extragradient method for finding saddle points and for other problems. *Èkonomika i Mat. Metod.*, 12(4):747–756, 1976.

G. M. Korpelevič. Extrapolation gradient methods and their relation to modified Lagrange functions. *Èkonomika i Mat. Metod.*, 19(4):694–703, 1983.

G. Kotsalis, G. Lan, and T. Li. Simple and optimal methods for stochastic variational inequalities, I: operator extrapolation. pages 1–25, 2020. URL http://arxiv.org/abs/2011.02987.

D. Kraljic, M. Troha, and B. Sobocan. Extracting physical power plant parameters from historical behaviour. In *2019 16th Int. Conf. Eur. Energy Mark.*, pages 1–5. IEEE, 2019. doi:10.1109/EEM.2019.8916516.

P. Krokhmal and S. Uryasev. A sample-path approach to optimal position liquidation. *Ann. Oper. Res.*, 152:193–225, 2007. doi:10.1007/s10479-006-0143-3.

Q. Lin, X. Chen, and J. Peña. A trade execution model under a composite dynamic coherent risk measure. *Oper. Res. Lett.*, 43(1):52–58, 2015. doi:10.1016/j.orl.2014.11.005.

P. L. Lions and B. Mercier. Splitting Algorithms for the Sum of Two Nonlinear Operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979. doi:10.1137/0716071.

C. Lorenz and A. Schied. Drift dependence of optimal trade execution strategies under transient price impact. *Financ. Stochastics*, 17(4):743–770, 2013. doi:10.1007/s00780-013-0211-x.

J. Lorenz and R. Almgren. Mean-Variance Optimal Adaptive Execution. *Appl. Math. Financ.*, 18(5): 395–422, 2011. doi:10.1080/1350486X.2011.560707.

J. J. Lucia and E. S. Schwartz. Electricity prices and power derivatives: Evidence from the Nordic Power Exchange. *Rev. Deriv. Res.*, 5(1):5–50, 2002. doi:10.1023/A:1013846631785.

X. Luo and A. Schied. Nash equilibrium for risk-averse investors in a market impact game with transient price impact. pages 1–21, 2018. URL http://arxiv.org/abs/1807.03813.

F. Mariani and L. Fatone. Optimal solution of the liquidation problem under execution and price impact risks. *Quant. Financ.*, 22(6):1037–1049, 2022. doi:10.1080/14697688.2022.2047219.

H. Markowitz. Portfolio Selection. *J. Finance*, 7(1):77–91, 1952. doi:10.2307/2975974.

B. Martinet. Brève communication. Régularisation d'inéquations variationnelles par approximations successives. *Rev. française d'informatique Rech. opérationnelle. Série rouge*, 4(R3):154–158, 1970. doi:10.1051/m2an/197004R301541.

G. J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.*, 29(3), 1962. doi:10.1215/S0012-7094-62-02933-2.

G. J. Minty. On the generalization of a direct method of the calculus of variations. *Bull. Am. Math. Soc.*, 73(3):315–322, 1967. doi:10.1090/S0002-9904-1967-11732-4.

J. Nash. Equilibrium Points in n-Person Games. *Proc. Natl. Acad. Sci.*, (36):48–49, 1950. doi:10.2307/j.ctv173f1fh.6.

J. Nash. Non-Cooperative Games. *Ann. Math.*, 54(2):286, 1951. doi:10.2307/1969529.

Y. Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Math. Program.*, 109(2-3):319–344, 2007. doi:10.1007/s10107-006-0034-z.

Y. Nesterov. *Lectures on Convex Optimization*, volume 137. Springer International Publishing, 2018. doi:10.1007/978-3-319-91578-4.

Y. Nesterov and L. Scrimali. Solving strongly monotone variational and quasi-variational inequalities. *Discret. Contin. Dyn. Syst. - A*, 31(4):1383–1396, 2011. doi:10.3934/dcds.2011.31.1383.

A. A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *J. Financ. Mark.*, 16(1):1–32, 2013. doi:10.1016/j.finmar.2012.09.001.

M. Ohnishi and M. Shimoshimizu. Optimal and equilibrium execution strategies with generalized price impact. *Quant. Financ.*, 20(10):1625–1644, 2020. doi:10.1080/14697688.2020.1749294.

G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.*, 72(2):383–390, 1979. doi:10.1016/0022-247X(79)90234-8.

A. Philpott, M. Ferris, and R. Wets. Equilibrium, uncertainty and risk in hydro-thermal electricity systems. *Math. Program.*, 157(2):483–513, 2016. doi:10.1007/s10107-015-0972-4.

W. B. Powell. *Approximate Dynamic Programming*. John Wiley & Sons, Inc., 2011. doi:10.1002/9781118029176.

D. Ralph and Y. Smeers. Pricing Risk Under Risk Measures: An Introduction to Stochastic-Endogenous Equilibria. *SSRN Electron. J.*, (August):1–45, 2011. doi:10.2139/ssrn.1903897.

D. Ralph and Y. Smeers. Risk Trading and Endogenous Probabilities in Investment Equilibria. *SIAM J. Optim.*, 25(4):2589–2611, 2015. doi:10.1137/110851778.

F. Riedel. Dynamic coherent risk measures. *Stoch. Process. their Appl.*, 112(2):185–200, 2004. doi:10.1016/j.spa.2004.03.004.

R. T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, 33(1): 209–216, 1970a. doi:10.2140/pjm.1970.33.209.

R. T. Rockafellar. On the maximality of sums of nonlinear monotone operators. *Trans. Am. Math. Soc.*, 149(1):75–75, 1970b. doi:10.1090/S0002-9947-1970-0282272-5.

R. T. Rockafellar. Monotone Operators and the Proximal Point Algorithm. *SIAM J. Control Optim.*, 14 (5):877–898, 1976. doi:10.1137/0314056.

R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *J. Risk*, 2(3):21–41, 2000. doi:10.21314/jor.2000.038.

R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *J. Bank. Financ.*, 26(7):1443–1471, 2002. doi:10.1016/S0378-4266(02)00271-6.

J. B. Rosen. Existence and Uniqueness of Equilibrium Points for Concave N-Person Games. *Econometrica*, 33(3):520–534, 1965. doi:10.2307/1911749.

K. Ruddell, A. Philpott, and A. Downward. Supply function equilibrium with taxed benefits. *Oper. Res.*, 65(1):1–18, 2017. doi:10.1287/opre.2016.1555.

K. Ruddell, A. Downward, and A. Philpott. Market power and forward prices. *Econ. Lett.*, 166:6–9, 2018. doi:10.1016/j.econlet.2018.02.016.

W. Rudin. *Functional analysis*. McGraw-Hill, Inc., New York, 1991.

B. Rudloff, A. Street, and D. M. Valladão. Time consistency and risk averse dynamic decision models: Definition, interpretation and practical consequences. *Eur. J. Oper. Res.*, 234(3):743–750, 2014. doi:10.1016/j.ejor.2013.11.037.

A. Ruszczyński. Risk-averse dynamic programming for Markov decision processes. *Math. Program.*, 125 (2):235–261, 2010. doi:10.1007/s10107-010-0393-3.

E. K. Ryu and S. Boyd. A Primer on Monotone Operator Methods (Survey). *Appl. Comput. Math*, (1): 3–43, 2016. URL `http://acmij.az/view.php?lang=az{&}menu=journal{&}id=405`.

A. Schied and T. Schöneborn. Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. *Financ. Stochastics*, 13:181–204, 2009. doi:10.1007/s00780-008-0082-8.

A. Schied and T. Zhang. A market impact game under transient price impact. *Math. Oper. Res.*, 44(1): 102–121, 2019. doi:10.1287/moor.2017.0916.

A. Schied, T. Schöneborn, and M. Tehranchi. Optimal Basket Liquidation for CARA Investors is Deterministic. *Appl. Math. Financ.*, 17(6):471–489, 2010. doi:10.1080/13504860903565050.

T. Schöneborn. *Trade execution in illiquid markets*. PhD thesis, 2008.

T. Schöneborn. Adaptive basket liquidation. *Financ. Stochastics*, 20:455–493, 2016. doi:10.1007/s00780-016-0290-6.

P. M. Schwarz. Electricity restructuring and deregulation. In *Energy Econ.*, pages 294–321. Routledge, 2022. doi:10.4324/9781003163190-17.

A. Shapiro. On a time consistency concept in risk averse multistage stochastic programming. *Oper. Res. Lett.*, 37(3):143–147, 2009. doi:10.1016/j.orl.2009.02.005.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, 2009. doi:10.1137/1.9780898718751.

A. Smith. The Wealth of Nations. In *Adam Smith*, number 1776, pages 168–185. Routledge, 2014. doi:10.4324/9780203092736-14.

M. Troha. *Fundamental model for dynamics of electricity prices*. PhD thesis, 2015. URL https://ora.ox.ac.uk/objects/uuid:60e77641-389d-4cab-a80f-b6b532f014ea.

M. Troha and R. Hauser. The impact of startup costs and the grid operator on the power price equilibrium. 2014a. URL http://arxiv.org/abs/1412.0148.

M. Troha and R. Hauser. Calculation of a power price equilibrium. 2014b. URL http://arxiv.org/abs/1409.6645.

M. Troha and R. Hauser. The existence and uniqueness of a power price equilibrium. 2014c. URL http://arxiv.org/abs/1408.2464.

J. Vaes and R. Hauser. Solving strongly monotone generalised variational inequalities: theoretical and numerical approaches. *In preparation*, a.

J. Vaes and R. Hauser. Volume-Uncertain Trade Execution via Dynamic Coherent Risk Measures. *In preparation*, b.

J. Vaes and R. Hauser. Uncertain Volume Execution in Financial Markets: A Game Theory Perspective. *In preparation*, c.

J. Vaes and R. Hauser. Optimal trade execution with uncertain volume target. *J. Comput. Financ.*, 26 (1), 2022. doi:10.21314/JCF.2022.018.

H. R. Varian. The Arbitrage Principle in Financial Economics. *J. Econ. Perspect.*, 1(2):55–72, 1987. doi:10.1257/jep.1.2.55.

M. Voisei. Characterizations and continuity properties for maximal monotone operators with non-empty domain interior. *J. Math. Anal. Appl.*, 391(1):119–138, 2012. doi:10.1016/j.jmaa.2012.02.025.

A. von Heusinger and C. Kanzow. Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions. *Comput. Optim. Appl.*, 43(3):353–377, 2009. doi:10.1007/s10589-007-9145-6.

Y. Yao, M. Postolache, and Y. C. Liou. Variant extragradient-type method for monotone variational inequalities. *Fixed Point Theory Appl.*, 2013:1–15, 2013. doi:10.1186/1687-1812-2013-185.