Instituto Nacional de Matemática Pura e Aplicada

Decomposition and Approximation Methods for Variational Inequalities, with Applications to Deterministic and Stochastic Energy Markets

Doctoral Thesis

by

Juan Pablo Luna

Advisor:

Mikhail Solodov

Co-Advisor:

Claudia Sagastizábal

March 7, 2013

A *la marlon...*

Agradecimentos

Com a redação desta tese, chega ao final mais uma etapa da minha vida. Etapa de estudante de matemática que começou há muito tempo quando ainda era adolescente. Ao longo deste período tenho conhecido muitas pessoas que tem me influenciado nas mais variadas formas. A elas gostaria de expressar o meu profundo agradecimento. Embora não possa mencionar todas, não posso deixar de fazê-lo com aquelas que aparecem a seguir. Primeiro, gostaria de agradecer aos meus pais Haydee e Uldárico, pela educação recebida, pelo exemplo e porque nunca deixaram faltar um livro em casa.

Agradeço aos amigos de Acecom, com quem passei a maior parte do tempo da graduação. A Orestes Bueno, Rodolfo Canto, Wilfredo Evangelista, Víctor Peña, Daniel Quino, Oswaldo Velásquez, Milagros Zorrilla. Incluo também no grupo os amigos da faculdade: Roberto Colonia, Alba Málaga, Arturo Rabanal, Kenny Venegas e Abraham Zamudio. A John Cotrina e família, por toda a ajuda e carinho brindado a longo do período da faculdade. Também a Janeth Valverde, grande amiga, quase irmã, por todos os momentos inesquecíveis que compartilhamos durante o tempo em que passei em Lima.

Agradeço aos meus professores da faculdade. A Júlio Alcántara, Félix Escalante, Manuel Toribio. Também a Eladio Ocaña, uns dos culpados para eu me iniciar na área de otimização matemática.

A José Cruz e família. Todo o apoio e compreensão que me brindaram foram determinantes para poder formar-me na faculdade.

À senhora Maura Capristano. Nunca serei capaz de expressar o apreço e estima que tenho por ela. A seus filhos Lourdes, César e Javier. Também ao seu sobrinho Daniel Ayala; *la iglesia de las Nazarenas* e as *chocotejas borrachas* estão ligadas às lembranças das nossas aulas de física.

A Cássio Silva, meu primeiro amigo no Brasil, com quem aprendi português. A Tássia Dias e ao Raulzinho. Com muita saudade.

A Liev Maribondo, grande amigo do IMPA, pelo apoio brindado, pela companhia e pela amizade. Aos amigos do IMPA, que diariamente compartilharam as horas de estudo. A Jefferson Chan, Guilherme Freitas, Renato Giroldo, Bruno Gois, Bruno Holanda, Karen Guevara, Ruben Lizarbe, Welington Oliveira, Miguel Orrillo, Lucas Pahl, Marcus Raymundo, Pedro Sabino e Guilherme Salomé. Também a Carlos Eduardo Guedes Belchior, alguém cujo fato de ser flamenguista não impede de ser uma das melhores pessoas que conheci.

Aos amigos que conheci em Lima e ainda continuam comigo no Rio de

Janeiro. A Enrique Arias, Calos Olivares, Carlos León, Cynthia Contreras e Luis Durand.

A Jean Pierre e Raquel Brito, novos amigos que em pouco tempo me levaram por rumos que nunca suspeitei. Também a Úrsula Jeremias, com muito apreço e admiração.

Na vida sempre acontecem fatos afortunados. Com certeza um deles foi conhecer os meus orientadores Claudia Sagastizábal e Mikhail Solodov. A eles agradeço os quatro anos de aprendizado e todos os momentos de trabalho e lazer compartilhados. Todas essas lembranças me acompanharão ao longo da vida.

Finalmente, agradeço aos membros da banca na defesa desta tese. Aos professores Alejandro Jofré, Paulo Silva, Susana Scheimberg e Alfredo Iusem. Também ao IMPA, CNPq e FAPERJ por terem oferecido todo o apoio que me permitiu alcançar os objetivos traçados.

Contents

Chapter 1

Introduction

A common task in optimization and real-life applications is to minimize some quantity subject to some constraints. The complexity of this problem depends on the nature of the functions involved in the problem and on the size of the data. In this work we shall deal with these two aspects in a number of different contexts.

First, instead of considering only one agent trying to minimize a given quantity, we consider a group of agents each one minimizing its quantity, and all the agents facing conflicting interests. The natural way of modeling this situation is the *Generalized Nash Equilibrium Problem* (GNEP). However, the Nash Equilibrium concept usually leads to multiple solutions, many of them without any special significance or even with economically inconsistent meanings. Since the very beginning of research in this area, economists have been defining refinements for the concept of equilibrium. For the models considered in this work we use the so-called *Variational Equilibrium* (VE), which is a refinement that privileges points that lead to equal marginal values for all the players in the game. We believe this property is useful/meaningful because of its fairness for all the players. Another advantage of the VE has to do with the task of actually finding it with computational tools. Whereas computing an arbitrary Nash Equilibrium is equivalent to solving a *Quasi Variational Inequality* (QVI), computing a VE requires to solve a *Variational Inequality* (VI). Numerical methods for QVI cannot be currently considered as well developed. The VI problem, on the other hand, is comparatively well studied, there is considerable theory and some efficient software designed for it. Of course, the efficiency/applicability still depends on the structure of the problem and its size, which is the other issue of this work.

Concerning the modeling aspect, we consider the deterministic and stochastic cases. In the deterministic case, where the models are based on GNEP, our main contribution consists in reducing some extended market models based on *Mixed Complementarity Problem* (MCP) to the GNEP approach. To be more specific, whereas in GNEP all the variables present in each player's problem belong to this player, there exist models where players' objective functions in-

volve *exogenous* variables that do not belong to any player. That is, variables that are exogenous to all the players in the game. This is the typical case of market prices. The players try to maximize their profit using these market prices, satisfying at the same time some constraints, called *market-clearing conditions*, whose role is "to keep the market working". For solving this problem, the common approach consists in putting together the optimality conditions of all the players' profit maximization problems and the market-clearing conditions, building in this way an MCP where the initial exogenous variables become variables of the whole problem (and are treated the same as all the other variables). The main drawback of this approach is that it involves a large number of variables and does not allow to perceive and use some special decomposable structures that are, in fact, intrinsic to the nature of this type of problems. We show that this MCP model can be reformulated as GNEP whose VE provides the same solution. Moreover, the VI associated to this VE of GNEP has "good" structure, amenable to decomposition techniques.

Some of our main contributions consist precisely in developing new results on decomposition of VIs, and in numerical validation of the benefits of the associated techniques when applied to structured but large-scale problems. Specifically, in Chapters 3 and 4 we present a new class of Dantzig-Wolfe and Benders type decomposition algorithms for VIs, which include the previously available methods as special cases and also improve the previously known convergence results.

In Chapter 5 we explore the relations between MCP and game-theoretical models for finding equilibrium prices in energy markets. This is done both in a deterministic and stochastic setting. For the latter, we consider players whose objective functions and decision variables are affected by some random vector. Like in the deterministic framework, we compare the stochastic game and the complementarity formulations, only that in this case the models are not always equivalent. The equivalence depends on the perception of risk of the players, that now can be risk neutral or risk averse. In the first case, the players are trying to minimize the expected value of their objective functions, and we can define a risk neutral GNEP that is equivalent to a risk neutral MCP. The main issue, however, is when we are dealing with risk-averse players. In this setting, instead of the expected value of the players' objective functions, *risk measures* are involved, such as the so-called *average value at risk* (AV@R), a recent renaming of the *Conditional Value-at-Risk*. We show that, in the presence of risk aversion, the MCP and game models are different in essence.

Another contribution of this work is in the solution of stochastic games, and this is the subject of Chapter 6. The use of risk measures yields nondifferentiable objective functions for the players problems and, consequently, the associated VI mapping is multi-valued. This makes the model hard to solve, as for this class of problems there is currently no established efficient software available. Thus, for solving this multi-valued VI we build a family of approximating VIs whose mappings are single-valued. This construction is by ways of smooth approximations of the risk-measure. We consider the risk-measure most widely used in applications, the AV@R, that is convex but nondifferentiable, due to the

positive-part function that enters its definition. We build a family of smooth approximations to AV@R based on twice differentiable approximations of the positive-part function, and establish the validity and convergence of the resulting approximation scheme for solving stochastic GNEPs.

In addition to the nondifferentiability of the functions (or multi-valued VI mappings), the number of variables involved constitutes yet another challenge for solving a stochastic GNEP. In this sense, the decomposition chapters of this work can be useful to deal with this issue. No matter how good are the solvers available or how powerful are the computers being used, the size of a stochastic problem always might present a challenge.

This work is organized as follows. We start in Chapter 2 with some basic notions and introductory material: the Dantzig-Wolfe and Benders decomposition algorithms for linear programming problems; facts and notions concerning the Variational Inequality and Generalized Nash Equilibrium problems; and a general framework for modeling energy markets using GNEPs. These concepts will be useful for the subsequent material, containing the main contributions of this work, summarized below.

- Chapter 3 describes a Dantzig-Wolfe decomposition approach for VIs published in [49], which is applicable to nonmonotone single-valued continuous operators, or to set-valued maximal monotone operators. The method is useful for VIs having feasible sets given by the intersection of two closed convex sets, with one of them coupling the variables while the other one being simple. This situation is typical when searching for variational equilibria of GNEPs.

- Chapter 4 derives, via duality, Benders decomposition from the Dantzig-Wolfe method in the case of VIs of a certain structure.

- Chapter 5 explores the relations between mixed complementarity, variational inequality, and game-theoretical formulations of energy markets, both in deterministic and stochastic settings.

- Chapter 6 addresses the problem of finding variational equilibria for risk-averse stochastic counterparts of GNEPs, by suitably combining smoothing and approximation techniques.

Chapters 3 to 6 start with an introduction that gives the specific context and places the work with respect to the state-of-the-art literature.

The dissertation ends with concluding remarks and a discussion of possible lines of future research.

Chapter 2

Preliminaries

In this chapter we review some concepts that either will be used in the sequel or are important for putting our developments in context. In the first two sections we recall the Dantzig-Wolfe and Benders decomposition algorithms for linear programming problems (the setting for which those techniques were developed originally, and where they are easy to explain). The third section is devoted to some basic notation, facts and notions concerning the Variational Inequality Problem (VI), the Generalized Nash Equilibrium Problem (GNEP), and some relations between them. The last section describes a general framework for modeling markets using GNEPs.

Our notation is mostly standard. For x, y in any given space, we denote by $\langle x, y\rangle$ the usual (Euclidean) inner product; then $\|\cdot\|$ stands for the associated norm. When convenient, we write $x \perp y$ to say that $\langle x, y\rangle = 0$. If clear from the context, for a vector x the notation $|x|$ refers to its dimension. For any set D, $\operatorname{conv} D$ denotes its convex hull and i_D its indicator function. If D is convex, $P_D(x)$ denotes the projection of a point x onto D. By $\mathcal{N}_D(x)$ we denote the normal cone to the convex set D at x, that is $\mathcal{N}_D(x) = \{w : \langle w, y - x\rangle \leq 0, \text{ for all } y \in D\}$ if $x \in D$ and $\mathcal{N}_D(x) = \emptyset$ otherwise. In the space $\mathbb{R}^{n+1}$, we denote by Ω_n its unitary simplex, that is $\Omega_n = \{\alpha \in \mathbb{R}^{n+1} : \alpha \geq 0, \sum_{i=1}^{n+1} \alpha_i = 1\}$. The positive-part function is $[\cdot]^+ = \max\{\cdot, 0\}$.

Given a convex function $f : \mathbb{R}^n \to \mathbb{R}$, we denote by $\partial f(x) = \{y \in \mathbb{R}^n : f(z) \geq f(x) + \langle y, z - x\rangle \ \forall\, z \in \mathbb{R}^n\}$ its subdifferential at a point $x \in \mathbb{R}^n$. For any differentiable function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we denote by $f'(x, y)$ its gradient at $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$, i.e., the derivative with respect to all the variables; and by $\nabla_x f(x, y)$ its partial derivative at (x, y) with respect to x (this derivatives notation is therefore a bit inconsistent, but there are reasons why we adopt it and it surely should not lead to any confusion).

Given a polyhedral set D and $C = \{x \in \mathbb{R}^n : h(x) \leq 0\}$ where the components of the vector-function $h : \mathbb{R}^n \to \mathbb{R}^q$ are convex, we say that the (generalized) Slater constraint qualification holds for the set $C \cap D$ if there exists $x \in D$ such that $h(x) < 0$.

A mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ (from $\mathbb{R}^n$ to the subsets of $\mathbb{R}^n$) is monotone if

it holds that $\langle u - v, x - y\rangle \geq 0$ for all $x, y \in \operatorname{dom} F = \{z : F(z) \neq \emptyset\}$ and all $u \in F(x)$, $v \in F(y)$. It is further maximal monotone if, in addition, its graph $\{(x, u) : u \in F(x), x \in \operatorname{dom} F\}$ is not contained in the graph of any other monotone operator. The function F is strictly monotone if $\langle u - v, x - y\rangle > 0$ for all $x, y \in \operatorname{dom} F$ and $u \in F(x)$, $v \in F(y)$ with $x \neq y$; and it is c-strongly monotone if for $c > 0$ it holds that $\langle u - v, x - y\rangle \geq c\|x - y\|^2$ for any choices above.

We say that $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^q$ is outer semicontinuous if for any sequences $\{x^k\}, \{y^k\}$ such that $\{x^k\} \to \bar{x}$ and $\{y^k\} \to \bar{y}$ with $y^k \in F(x^k)$, it holds that $\bar{y} \in F(\bar{x})$. We say that a family of set-valued functions $\{F^k\}$ is equicontinuous on compact sets if for every compact set D and every $\varepsilon > 0$ there exists $\delta > 0$ such that for any $x, y \in D$ with $\|x - y\| < \delta$, for every k it holds that $d_H(F^k(x), F^k(y)) < \varepsilon$, where d_H is the Hausdorff distance between the sets defined by

$$d_H(A, B) = \inf\{r > 0 : A \subset B + B(0, r) \text{ and } B \subset A + B(0, r)\}.$$

Note that a maximal monotone mapping is outer semicontinuous. The inverse of a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^q$ is given by $F^{-1}(u) = \{x \in \mathbb{R}^n : u \in F(x)\}$ for $u \in \mathbb{R}^q$.

2.1 Dantzig-Wolfe decomposition for linear programming

As is well documented, the field of (modern) Optimization started with *Linear Programming* (LP). Even though by now there exist very efficient LP solvers, there are still problems whose size makes them hard to solve, or whose structure allows them to be solved considerably faster if special decomposition techniques rather than general-purpose solvers are applied. This was even more critical around the middle of the XXth century, when George Dantzig and Philip Wolfe published their decomposition algorithm [16] taking advantage of a special decomposable structure of some constraints to solve fast certain LPs.

In this section we describe the Dantzig–Wolfe decomposition algorithm for LPs in order to give a general idea of the nature of this technique, and to set the stage for its generalization to the setting of VIs presented in Chapter 3.

Given affine functions $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^p$ and $g : \mathbb{R}^n \to \mathbb{R}^q$ (there is no need for our purposes to define them here explicitly as matrix/vector products), consider the LP

$$\min \quad f(x) \quad \text{s.t.} \quad h(x) \leq 0,\ g(x) \leq 0. \tag{2.1}$$

We accordingly split the feasible set of (2.1) into

$$S_h = \{x : h(x) \leq 0\}, \quad S_g = \{x : g(x) \leq 0\}.$$

Assuming that the structure of the problem data is such that linear optimization over the set S_g can be carried out easily (as compared to minimizing over

$S_h \cap S_g$), the idea of the Dantzig–Wolfe method is to perform the Lagrangian relaxation of the h-constraints, and to apply the cutting-plane algorithm [5, Sec. 9.3.2] for nonsmooth optimization to solve the Lagrangian dual

$$\max \quad \theta(\mu) \quad \text{s.t.} \quad \mu \in \mathbb{R}^q_+,$$

where

$$\theta(\mu) = \inf\{f(x) + \langle \mu, h(x)\rangle : x \in S_g\}. \tag{2.2}$$

(Note that θ is a concave function.)

When set in an iterative framework, Dantzig-Wolfe method means the following. Given the current dual iterate $\mu_M^k \in \mathbb{R}^q_+$, the k-th subproblem computes the value of the dual function $\theta(\mu_M^k)$, which means obtaining a minimizer x_S^{k+1} for the LP in (2.2). By construction, $x_S^{k+1} \in S_g$ and we have a subgradient $-h(x_S^{k+1}) \in \partial(-\theta)(\mu_M^k)$. For $k \geq 1$, suppose we have computed solutions of the previous subproblems $\{x_S^0, \ldots, x_S^k\} \subset S_g$. Then the k-th master problem replaces the set S_g in the original problem (2.1) by the approximation $\operatorname{conv}\{x_S^0, \ldots, x_S^k\}$, and solves the following LP:

$$\left\{ \begin{array}{l} \min f(x) \\ x \in \operatorname{conv}\{x_S^0, \ldots, x_S^k\} \\ x \in S_h \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{l} \min f\left(\sum_{i=0}^k \alpha_i x_S^i\right) = \sum_{i=0}^k \alpha_i f(x_S^i) \\ \alpha \in \Omega_k \\ h\left(\sum_{i=0}^k \alpha_i x_S^i\right) = \sum_{i=0}^k \alpha_i h(x_S^i) \leq 0. \end{array} \right. \tag{2.3}$$

This gives a solution x_M^k and a multiplier $\mu_M^k \in \mathbb{R}^q_+$ associated to the h-constraint in (2.3). Then the new dual function value $\theta(\mu_M^k)$ is computed, as well as an associated x_S^{k+1} in (2.2), and the process is repeated. From the point of view of maximizing the dual function θ, the dual problem of the master problem (2.3) corresponds to an iteration of the cutting-plane method, i.e., to solving the LP formulation of

$$\max \quad \theta^k(\mu) \quad \text{s.t.} \quad \mu \in \mathbb{R}^q_+,$$

where

$$\theta^k(\mu) = \min\{f(x_S^i) + \langle \mu, h(x_S^i)\rangle : i = 0, \ldots, k\}.$$

The algorithm continues iteratively until some stopping criterion is satisfied. The usual criterion for the cutting plane algorithm is to stop when the difference (gap) between the true value of the dual function (that is being maximized) and its cutting plane model at the current dual variable, that is, $\theta(\mu_M^k) - \theta^k(\mu_M^k)$, becomes small enough. Note that, by construction and using the linearity of the functions involved, it holds that

$$\begin{aligned} 0 \geq \theta(\mu_M^k) - \theta^k(\mu_M^k) &= (f(x_S^{k+1}) + \langle \mu_M^k, h(x_S^{k+1})\rangle) - f(x_M^k) \\ &= (f(x_S^{k+1}) + \langle \mu_M^k, h(x_S^{k+1})\rangle) - (f(x_M^k) + \langle \mu_M^k, h(x_M^k)\rangle) \\ &= \langle f'(x_M^k), x_S^{k+1} - x_M^k\rangle + \langle \mu_M^k, h(x_S^{k+1}) - h(x_M^k)\rangle \\ &= \langle f'(x_M^k) + [h'(x_M^k)]^\top \mu_M^k, x_S^{k+1} - x_M^k\rangle. \end{aligned} \tag{2.4}$$

This relation will be useful to define the "optimality" gap and the stopping rule for the generalization of this decomposition method for VIs in Chapter 3.

To finish, we emphasize that Dantzig–Wolfe approach for LPs is particularly effective when h is the coupling constraint without which the minimization in (2.2) decomposes further according to some favorable (block-separable) structure of g. For some large-scale applications the resulting method is faster than solving the original LP (2.1) directly, even by state-of-the-art software; see, e.g., [41, 15]. Furthermore, stabilization techniques of bundle methods [5, Ch. 10] can be used to define stabilized master problems and improve computational performance of the cutting-plane scheme [3, 6].

2.2 Benders decomposition for linear programming

The Dantzig-Wolfe algorithm was designed for the situation when the obstacle for exploiting some specific structure is the presence of "coupling" constraints. Another possible situation, in a sense dual to the above, is when we can distinguish two kinds of variables (say, x and y) that behave in such a way that if we fix one of them (say, y for our development) then the problem becomes much easier to solve.

Suppose we have affine functions $f : \mathbb{R}^n \to \mathbb{R}$, $r : \mathbb{R}^m \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and $h : \mathbb{R}^m \to \mathbb{R}^p$. We assume, for simplicity and without loss of generality, that $f(0) = 0$. The task is to solve the following LP:

$$\min \quad f(x) + r(y) \quad \text{s.t.} \quad g(x) + h(y) \leq 0. \tag{2.5}$$

As already commented, we are assuming that for any fixed value $\hat{y} \in \mathbb{R}^m$ the problem

$$\min \quad f(x) + r(\hat{y}) \quad \text{s.t.} \quad g(x) + h(\hat{y}) \leq 0 \tag{2.6}$$

is easier to solve. This usually occurs when g has block-decomposable structure, that is

$$g'(x) = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{bmatrix}$$

where A_i, $i = 1, \ldots, k$, are matrices of appropriate dimensions. The dual problem of (2.5) is

$$\begin{cases} \max & r(0) + \langle \mu, g(0) + h(0) \rangle \\ \text{s.t.} & f'(0) + [g'(0)]^\top \mu = 0, \\ & r'(0) + [h'(0)]^\top \mu = 0, \\ & \mu \geq 0. \end{cases} \tag{2.7}$$

Note that if $g'(0)$ has decomposable structure, then $[g'(0)]^\top$ also has decomposable structure. Then without the constraint $r'(0) + [h'(0)]^\top \mu = 0$, (2.7) would become an easier problem.

The idea of Benders decomposition to solving (2.5) is to solve its dual (2.7) using the Dantzig-Wolfe decomposition algorithm described in Section 2.1, regarding $r'(0) + [h'(0)]^\top \mu = 0$ as the "difficult constraint" and relaxing it. In this framework, given at iteration k a Lagrange multiplier estimate y_M^k, the subproblem iteration corresponds to solving

$$\begin{cases} \max & r(y_M^k) + \langle \mu, g(0) + h(y_M^k) \rangle, \\ \text{s.t.} & f'(0) + [g'(0)]^\top \mu = 0, \mu \geq 0. \end{cases} \tag{2.8}$$

As commented, with the structure at hand the latter is easier to solve.

Now, we need to solve (2.8) in order to get a subproblem solution μ_S^{k+1} for performing the Master subproblem step of the Dantzig-Wolfe algorithm. But instead of solving directly (2.8), we solve its dual:

$$\begin{cases} \min & f(x) + r(y_M^k) \\ \text{s.t.} & g(x) + h(y_M^k) \leq 0, \end{cases} \tag{2.9}$$

whose solution is denoted by x_S^{k+1}. Note also that the solution of (2.8) needed by the Dantzig-Wolfe algorithm is just the Lagrange multiplier of (2.9). Then, after performing this iteration, we can assume we have already computed the previous subproblems solutions $\mu_S^0, \mu_S^1, \ldots, \mu_S^k$. We can then perform the Dantzig-Wolfe Master subproblem step, that consists in solving

$$\begin{cases} \max & r(0) + \langle g(0) + h(0), \mu \rangle \\ \text{s.t.} & r'(0) + [h'(0)]^\top \mu = 0, \\ & \mu \in \operatorname{conv}\{\mu_S^0, \mu_S^1, \ldots, \mu_S^k\}, \end{cases} \tag{2.10}$$

from which we obtain a Lagrange multiplier y_M^k associated to the constraint $r'(0) + [h'(0)]^\top \mu = 0$, so we can continue with the next Dantzig-Wolfe algorithm subproblem.

Again, instead of solving directly (2.10), we solve its dual

$$\begin{cases} \min & t + r(y) \\ \text{s.t.} & \langle \mu_S^i, g(0) + h(y) \rangle \leq t, \ i = 0, 2, \ldots, k, \end{cases} \tag{2.11}$$

which can be rewritten, using optimality conditions of (2.9), as

$$\begin{cases} \min & t + r(y) \\ \text{s.t.} & f(x_S^i) + \langle [h'(y_M^{i-1})]^\top \mu_S^i, y - y_M^{i-1} \rangle \leq t, \ i = 1, 2, \ldots, k. \end{cases} \tag{2.12}$$

Summarizing, we observe that solving (2.7) by solving iteratively (2.8) and (2.10) following the Dantzig-Wolfe algorithm, is equivalent to solving iteratively (2.9) and (2.11). The latter is precisely the Benders decomposition method for LPs [4].

Here, the Benders algorithm was presented as a direct application of the Dantzig-Wolfe algorithm, and thus much of the analysis for the former can be based on the latter. However, it is worth to note another interpretation of the Benders algorithm. To that end, define the value function v by

$$v(y) = \begin{cases} \min & f(x) \\ \text{s.t.} & g(x) \leq -h(y). \end{cases}$$

Clearly, solving (2.9) is equivalent to computing $v(y_M^k)$. In addition, we have that $[h'(y_M^k)]^\top \mu_S^{k+1} \in \partial v(y_M^k)$. At iteration k, the values $\mu_S^1, \mu_S^2, \ldots, \mu_S^k$ and $x_S^1, x_S^2, \ldots, x_S^k$ are available. Then we can define a cutting plane approximation v^k of v as

$$v^k(y) = \max\{f(x_S^i) + \langle [h'(y_M^{i-1})]^\top \mu_S^i, y - y_M^{i-1} \rangle : i = 1, 2, \ldots, k\}.$$

Next, we observe that solving the problem (2.12) is equivalent to solving

$$\min \quad v^k(y) + r(y). \tag{2.13}$$

Then the Benders algorithm can be interpreted as follows. At the iteration k it builds a cutting plane approximation v^k of the value function v and then computes y_M^k by solving (2.13). After this, using this master solution we improve the approximation v^k and iterate again. This point of view seems in fact more natural, since the initial problem (2.5) is equivalent to solving

$$\min \quad v(y) + r(y).$$

On the other hand, viewing Benders algorithm as an application of the Dantzig-Wolfe approach to the dual problem, has the advantage of possibly using those ideas for other problem classes. This would be indeed the main idea in Chapter 4, where we extend Benders technique to VI problems.

2.3 Variational inequalities and generalized Nash equilibrium problems

The Variational Inequality (VI) [27] is a mathematical framework covering a number of different problems and applications. The definition is easy to understand. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued mapping from $\mathbb{R}^n$ to subsets of $\mathbb{R}^n$, and let C be a convex closed subset of $\mathbb{R}^n$. The VI problem defined by F and C, denoted by VI(F, C), consists in determining a point $\bar{x} \in C$ such that for some $\bar{w} \in F(\bar{x})$ the following inequality holds, for every $y \in C$:

$$\langle \bar{w}, y - \bar{x} \rangle \geq 0. \tag{2.14}$$

Note also that (2.14) means that $-\bar{w} \in \mathcal{N}_C(\bar{x})$, where $\mathcal{N}_C(\bar{x})$ is the normal cone to C at $\bar{x}$. This relation is in turn equivalent to $P_C(\bar{x} - \bar{w}) = \bar{x}$. Using the set-valued *natural residual* function $F^{\text{nat}} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by $F^{\text{nat}}(x) = P_C(x - F(x)) - x$, we have that $\bar{x}$ is a solution of VI(F, C) if and only if $0 \in F^{\text{nat}}(\bar{x})$. In the single-valued case, we can use $\|F^{\text{nat}}(x)\|$ to measure the "error" made by taking x as an approximation to a solution of VI.

As mentioned, many problems can be considered in a VI framework. For example, critical points of some minimization problem over a convex set C are solutions of (2.14), where F is the derivative or the subgradient of the objective function. Another example, more interesting for our development,

is the Generalized Nash Equilibrium Problem (GNEP) [25]. Before dealing with GNEPs, we describe the Nash Equilibrium Problem (NEP), an important particular case. Suppose we have N functions $f^i : \mathbb{R}^n \to \mathbb{R}$ and convex closed sets $X^i \subset \mathbb{R}^{n_i}$, where $n = \sum_{i=1}^N n_i$. The pair (f^i, X^i) describes the ith *player* with decision variable $x^i \in \mathbb{R}^{n_i}$. The function can be regarded as a *loss* function, in the sense that it can represent some production cost, payment, etc., that the player wants to minimize. What makes NEP interesting is that the loss function f^i depends not only on the player's decision, but also on the other players' decisions, denoted by x^{-i}. In this context, the ith player makes a guess $\hat{x}^{-i}$ about the other players' actions and then tries to take the best decision by solving

$$\min \quad f^i(x^i, \hat{x}^{-i}) \quad \text{s.t.} \quad x^i \in X^i. \tag{2.15}$$

A point $\bar{x} \in \prod_{i=1}^N X^i$ is called a *Nash Equilibrium* if for each player i, its entry $\bar{x}^i$ is the best decision given that the other players decided $\bar{x}^{-i}$. In other words, if all players choose the entries of $\bar{x}$, then no one will change its decision unilaterally.

Now, if each function f^i is differentiable with respect to x^i, by the optimality condition of first order in primal form, we have that for all $x^i \in X^i$

$$\langle \nabla_{x^i} f^i(\bar{x}), x^i - \bar{x}^i \rangle \geq 0.$$

By combining all these inequalities, we have that the Nash equilibrium $\bar{x}$ solves $\mathrm{VI}(F, \prod_i X^i)$ with $F(x) = (\nabla_{x^1} f^1(x), \nabla_{x^2} f^2(x), \ldots, \nabla_{x^N} f^N(x))$. Also we can see that if all functions f^i were convex on the corresponding variable x^i, then any solution of the VI would also be a Nash equilibrium.

In fact, in the convex case, even if the functions are nondifferentiable, the two problems are equivalent. In the nonconvex case, a VI solution provides critical points to each players' problem, which can be considered as Nash equilibrium for practical purposes.

In real-life situations when we have a group of agents, not only the loss functions but also the feasible sets depend on the decision of the other players. This is natural, because the model usually has to consider the fact that all players are sharing some limited resource. This is the key fact that defines GNEP [25, 27]. In this work we consider the so called *GNEP with shared, or coupling, constraints*. Suppose we have a convex closed subset $\mathcal{S} \subset \mathbb{R}^n$. Then the ith player, after guessing the others' decision variables $\hat{x}^{-i}$, is only allowed to choose points x^i in X^i that satisfy $(x^i, \hat{x}^{-i}) \in \mathcal{S}$. Accordingly, instead of (2.15) this player's problem is

$$\min \quad f^i(x^i, \hat{x}^{-i}) \quad \text{s.t.} \quad x^i \in X^i, \ (x^i, \hat{x}^{-i}) \in \mathcal{S}. \tag{2.16}$$

The concept of Generalized Nash Equilibrium Point is the same as the equilibrium concept associated to Nash Problems.

Differently from Nash games, where each equilibrium point is a solution of the associated VI, there can be generalized equilibrium points that cannot be obtained by solving a VI. In fact, if we repeat the steps followed to build the VI

associated to a Nash game, we would arrive to a *Quasi-Variational Inequality* (QVI) [27]. Actually, GNEP and QVI are "equivalent" as much as are NEP and VI; and since it is known that solving a QVI is much more complex than solving a VI, it is clear that solving a GNEP is much harder than solving a NEP.

It is known that GNEP solutions are generally not unique, and in fact, *typically not isolated* (thus, there is an infinite number of solutions). However, many of the GNEP solutions are likely not interesting. For example, given (2.16) and a set of positive numbers $\alpha_1, \alpha_2, \ldots, \alpha_N$, we can define another GNEP replacing, for each player, the objective function f^i by $\alpha_i f^i$. From the definition of equilibrium point, it is easy to show that the two problems have exactly the same equilibrium points; however, the presence of the factors α_i has changed completely the nature of the problems from a practical point of view, since, although all optimization problems still have the same solutions, the corresponding dual variables (marginal values) have changed, which leads to different economic interpretations. So, we see not all equilibrium points are useful or interesting. For example, let us consider a two player GNEP:

$$\left\{ \begin{array}{cc} \min_{x^1} & \frac{1}{2}(x^1-1)^2 \\ \text{s.t.} & x^1 + x^2 = 2 \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{cc} \min_{x^2} & \frac{1}{2}(x^2-1)^2 \\ \text{s.t.} & x^1 + x^2 = 2 \end{array} \right. \tag{2.17}$$

Using only feasibility considerations it is easy to show that the set of equilibrium points is precisely the whole (shared) feasible set. In this case, therefore, the objective functions have no influence on determining the equilibrium points; certainly an unreasonable situation from the practical point of view.

For this reason it is desirable to have some criterion in order to choose an equilibrium point that is most convenient in some sense [47]. One distinguished equilibrium point is the so called *variational equilibrium*, that is a solution of $\mathrm{VI}(F, \prod_i X^i \cap \mathcal{S})$ associated to (2.16). It is not difficult to show that any solution of this VI is in fact an equilibrium point for (2.16). This allows us to find an equilibrium solving only a variational inequality, and in addition this equilibrium turns out to be a better choice because it depends on the players' objective functions.

When the sharing set $\mathcal{S}$ is defined by constraints, say $\mathcal{S} = \{x \in \mathbb{R}^n : g(x) \leq 0\}$ for a convex differentiable function g, then at an equilibrium point $\bar{x}$ we have that for each player, the entry $\bar{x}^i$ solves the problem

$$\min \quad f^i(x^i, \bar{x}^{-i}) \quad \text{s.t.} \quad x^i \in X^i,\ g(x^i, \bar{x}^{-i}) \leq 0.$$

By assuming qualifications of constraints [77], each player will have a Lagrange multiplier $\bar{\mu}^i$ associated to the constraint $g(x^i, \bar{x}^{-i}) \leq 0$. In [26], it was observed that the variational equilibria are precisely those for which multipliers $\bar{\mu}^i$ in all the players' problems are the same. This means that at a variational equilibrium, the optimum value of all players will vary at the same rate (given by the multiplier) after small perturbations that could occur on the sharing constraint. This makes good sense in many models, from the economic point of view. The other equilibria always benefit some players more than others, that is, they are not "fair"; the latter being an important characteristic in various applications

modeled as games. In Section 5.3.1 we comment on this issue for a game with affine coupling constraints that was used to model an energy/capacity market in [23] and the European network of natural gas in [32].

Summarizing, when dealing with GNEP, we shall aim mainly at finding variational equilibria, as this is justified both by the advantage of reducing the problem to a better understood and easier to solve VI, and by the desirable properties that such equilibria have from the practical point of view.

Finally, we fix the notation used when describing a GNEP. For simplicity, instead of emphasizing the guess made by a player about the other players' decisions by writing, for example, $\hat{x}^{-i}$ in (2.16), we drop the "*hat*" mark, and simply write

$$\min_{x^i} \quad f^i(x^i, x^{-i}) \quad \text{s.t.} \quad x^i \in X^i,\ (x^i, x^{-i}) \in \mathcal{S},$$

or

$$\min_{x^i} \quad f^i(x^i, x^{-i}) \quad \text{s.t.} \quad x^i \in X^i,\ x \in \mathcal{S}.$$

Sometimes, when convenient or instructive, we shall describe separately the players' objective functions and endogenous feasible sets, and the coupling constraints, by

$$\textbf{Player}\ \left\{ \min_{x^i} \quad f^i(x^i, x^{-i}) \quad \text{s.t.} \quad x^i \in X^i, \right.$$
$$\textbf{Coupling constraints}\ \left\{ x \in \mathcal{S}. \right.$$

This last notation will be useful when describing games with players of different nature sharing exactly the same coupling constraints.

2.4 A game model for gas and electricity

In this section we describe general structure of a GNEP for modeling markets where we make explicit two variables of different nature, that is, investment and production decisions. The purpose of the modeling is to allow the simultaneous consideration of capacity and energy markets, as in [23]; see Chapter 5. Our general framework based on GNEP with shared constraints is also suitable to model market equilibrium problems like the one in [21].

Let $(z^i, q^i) \in \mathbb{R}^{n_i} \times \mathbb{R}^{m_i}$, $i = 1, \ldots, N$, where z^i represents the investment variables and q^i represents the operational variables. We denote by z and q the vectors $(z^1, z^2, \ldots, z^N)$ and $(q^1, q^2, \ldots, q^N)$, respectively. Let X^i be the set of endogenous constraints of the ith player. This set gathers technological and resource constraints. Also, there are coupling constraints that link the players' decisions, represented by the inequality

$$\sum_{i=1}^{N} [h^i(z^i) + g^i(q^i)] \leq 0, \tag{2.18}$$

where the functions h^i and g^i are convex.

Given strategy vectors $\hat{z}$ and $\hat{q}$, the ith player tries to minimize the quantity $I^i(z^i, \hat{z}^{-i}) + c^i(q^i, \hat{q}^{-i})$. This quantity can represent a loss function (i.e., the negative of the profit function or investment), operational cost function. So, the ith player's problem is

$$\begin{cases} \min & I^i(z^i, z^{-i}) + c^i(q^i, q^{-i}) \\ \text{s.t.} & (z^i, q^i) \in X^i, \\ & \sum_{j=1}^{N} [h^j(z^j) + g^j(q^j)] \leq 0. \end{cases} \tag{2.19}$$

With respect to the notation in (2.16), we have the relations

$$x^i \leftrightarrow (z^i, q^i), \quad f^i(x^i) \leftrightarrow I^i(z^i) + c^i(q^i), \quad \text{and} \quad \mathcal{S} \leftrightarrow \sum_{j=1}^{N} [h^j(z^j) + g^j(q^j)] \leq 0.$$

As explained before, for economic applications it is often not enough to find a Nash equilibrium, and we look for a variational equilibrium (see [47, 26]). Since all functions involved in (2.19) are differentiable, we can compute a variational equilibrium by solving

$$\text{VI}(F, \prod_{i=1}^{N} X^i \cap \mathcal{S}), \tag{2.20}$$

where $\mathcal{S}$ is the set of all points that satisfy the shared constraints (2.18) and

$$F = (F^1, F^2, \ldots, F^N)$$

with

$$F^i(\{(z^j, q^j)\}_j) = \nabla_{(z^i, q^i)}[I^i(z) + c^i(q)], \quad i = 1, \ldots, N.$$

By the convexity and smoothness of the players' objective functions, the function F is continuous. Hence, to ensure the existence of solutions of (2.20) it is enough to guarantee compactness of the VI-feasible set $\prod_{i=1}^{N} X^i \cap \mathcal{S}$. The latter is typically the case in applications. Also, it is worth noting that under constraint qualification hypotheses at a solution $(\bar{z}, \bar{q})$, there exists an "optimal" multiplier $\bar{\pi}$ associated to the shared constraint (2.18). Relaxing this constraint reveals the decomposable structure in the feasible set, and for each player i the point $(\bar{z}^i, \bar{q}^i)$ solves the individual problem

$$\text{VI}(\nabla_{(z^i, q^i)}[I^i(z^i, \bar{z}^{-i}) + c^i(q^i, \bar{q}^{-i}) - [h^i(z^i) + g^i(q^i)]^\top \bar{\pi}], X^i).$$

In other words, $(\bar{z}^i, \bar{q}^i)$ solves the following profit maximization–like problem:

$$\begin{cases} \max & [h^i(z^i) + g^i(q^i)]^\top \bar{\pi} - c^i(q^i, \bar{q}^{-i}) - I^i(z^i, \bar{z}^{-i}) \\ \text{s.t.} & (z^i, q^i) \in X^i. \end{cases} \tag{2.21}$$

This game model (for a simple electricity market) will be used below to show the benefits of the Dantzig-Wolfe decomposition introduced in Chapter 3.

The reformulation of the cost-minimization game in the context of revenue maximization (2.21) will be useful in Chapter 5, devoted to finding equilibria of energy markets. In particular, the analysis in Chapter 5 shows that a variational equilibrium together with the Lagrange multipliers of the coupling constraints provides an equilibrium for complementarity formulations based on (2.21). Furthermore, this equilibrium point actually provides always (without any regularity conditions) a variational equilibrium for the GNEP described above. As a result, equilibrium prices can be found by solving a VI with a smaller number of variables than the one corresponding to the complementarity approach.

Finally, as stated above, the feasible set of the VI associated to a GNEP typically has the form $\prod_i X^i \cap \mathcal{S}$, where the set $\prod_i X^i$ has decomposable structure. Moreover, this structure *must* be used for some benefit, especially when dealing with large problems. The presence of the coupling conditions that define $\mathcal{S}$ prevents this, however. In the next chapter we develop a decomposition algorithm to solve more easily VI problems whose feasible sets follow the described pattern.

Chapter 3

Dantzig-Wolfe Decomposition for Variational Inequalities

In this chapter, we extend the decomposition algorithm described in Section 2.1 to a fairly general VI setting. Our presentation follows [49].

Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued mapping from $\mathbb{R}^n$ to the subsets of $\mathbb{R}^n$, and let S_h and S_g be two closed convex sets in $\mathbb{R}^n$. We consider the *variational inequality problem* $\mathrm{VI}(F, S_h \cap S_g)$ [27], which means to find

$$\bar{x} \in S_h \cap S_g \text{ such that } \langle \bar{w}, x - \bar{x} \rangle \geq 0 \text{ for some } \bar{w} \in F(\bar{x}) \text{ and all } x \in S_h \cap S_g. \tag{3.1}$$

In what follows, we assume that $S_h = \{x : h(x) \leq 0\}$ with $h : \mathbb{R}^n \to \mathbb{R}^q$ being convex and differentiable, and S_g is a generic closed convex set which is easier (in some sense) to handle than the set S_h and the intersection $S_h \cap S_g$. The set S_h is defined by inequality constraints only for simplicity; affine equality constraints can be introduced in our developments without any difficulties. We assume that the operator F is either single-valued and continuous (possibly nonmonotone) or it is maximal monotone. We also assume that $\mathrm{VI}(F, S_h \cap S_g)$ has a nonempty solution set, and that $S_g \subset \mathrm{int}(\mathrm{dom}\, F)$. We note that the latter assumption could be more general; we use the stated one for simplicity, as the issue does not seem to be of real importance in a work devoted to a computational algorithm.

The setting just described suggests trying to deal with the constraint sets S_h and S_g separately, i.e., by some type of decomposition of the problem $\mathrm{VI}(F, S_h \cap S_g)$. Many decomposition techniques (for monotone problems) are explicitly derived from the proximal point method [55, 63] for maximal monotone operators, e.g., [80, 19, 81, 83]. Sometimes the relation to the proximal iterates is less direct, e.g., the methods in [9, 20, 82, 35, 58], which were nevertheless more recently generalized and interpreted in [76, 53] within the hybrid

inexact proximal schemes of [78, 57]. As some other decomposition methods, we might mention [44] which employs projection and cutting-plane techniques for certain structured problems, matrix splitting for complementarity problems in [14], and the applications of the latter to stochastic complementarity problems in [72]. The methods cited above typically assume monotonicity and, from the beginning, some rather specific structure in the mapping F and/or in the constraints defining the feasible set. In that sense, our setting $\mathrm{VI}(F, S_h \cap S_g)$ and the subsequent developments are more general, as in the single-valued case F is allowed to be nonmonotone and no specific structural assumptions are being made about F or about the constraints. That said, if separable features are present, they can be exploited at the stage of solving the subproblems.

Let us recall the ideas of the Dantzig-Wolfe approach to linear programs discussed in Section 2.1, and examine what they might mean in the present more general setting. As is well known VI (3.1) is equivalent to the inclusion $0 \in F(\bar{x}) + \mathcal{N}_{S_h \cap S_g}(\bar{x})$, where $\mathcal{N}_D(x)$ is the normal cone to the convex set D at the point x. Let $\bar{x} \in S_h \cap S_g$. Under appropriate constraint qualification conditions for the sets S_h and S_g (see, e.g., [27, Chapter 3.2] and [77]), it holds that

$$\mathcal{N}_{S_h \cap S_g}(\bar{x}) = \mathcal{N}_{S_h}(\bar{x}) + \mathcal{N}_{S_g}(\bar{x}),$$

and

$$\mathcal{N}_{S_h}(\bar{x}) = \{u : u = [h'(\bar{x})]^\top \mu,\ \mu \in \mathbb{R}^q_+,\ \mu \perp h(\bar{x})\}.$$

In particular, for any solution $\bar{x}$ of (3.1), there exists a multiplier $\bar{\mu} \in \mathbb{R}^q_+$ such that

$$0 \in F(\bar{x}) + [h'(\bar{x})]^\top \bar{\mu} + \mathcal{N}_{S_g}(\bar{x}), \quad 0 \le \bar{\mu} \perp h(\bar{x}) \le 0.$$

Hence, solving problem (3.1) is equivalent to finding $(\bar{x}, \bar{\mu})$ such that

$$\begin{cases} (\bar{x}, \bar{\mu}) \in S_h \times \mathbb{R}^q_+, \quad \bar{\mu} \perp h(\bar{x}), \\ \bar{x} \text{ solves } \mathrm{VI}(F(\cdot) + [h'(\cdot)]^\top \bar{\mu}, S_g). \end{cases} \tag{3.2}$$

A natural extension of the ideas of the Dantzig–Wolfe decomposition for linear programming (described in Section 2.1) to this variational setting is then the following. Using the current multiplier estimate $\mu^k_M \in \mathbb{R}^q_+$ (instead of the unknown "optimal" $\bar{\mu}$), the k-th subproblem consist in solving a variational inequality with the structure in (3.2), to obtain a new primal point x^{k+1}_S. In particular, this variational problem is over the simpler set S_g, with the h-constraint dealt with in a manner similar to the Lagrangian relaxation approach. Using solutions of the previous subproblems $\{x^0_S, \ldots, x^k_S\} \subset S_g$, the k-th master problem solves a variational inequality with the structure in (3.1), except that the set S_g therein is approximated by $\mathrm{conv}\,\{x^0_S, \ldots, x^k_S\}$ (recall (2.3)). This gives a solution x^{k+1}_M and a new multiplier estimate $\mu^{k+1}_M \in \mathbb{R}^q_+$ for the h-constraint, and the process is repeated. Thus, we iteratively generate two sequences of (approximate) solutions of the problems (3.1) and (3.2), using at each iteration the solution of one problem to improve the solution of the other. We shall postpone the details and various possible options to be discussed later.

A Dantzig–Wolfe method along these lines had been introduced for (single-valued) variational inequalities in [29, 12]. In [29] some restrictive assumptions are employed. For example, F is required to be either strictly monotone or to be a separable combination of a strictly monotone part with a gradient of a differentiable convex function. The subproblems have the specific form $\mathrm{VI}(F(\cdot)+[h'(x_M^k)]^\top \mu_M^k, S_g)$. Also, the solvability of all the subproblems is an assumption. Some of the restrictive assumptions have been alleviated in [12], where also a useful feature of approximating F in the subproblems is introduced. The latter can be helpful in applications where the subproblem $\mathrm{VI}(F(\cdot)+[h'(x_M^k)]^\top \mu_M^k, S_g)$ is not decomposable, but using instead of F a suitable approximation makes it decomposable and thus easier to solve. One possibility considered in [12] is fixing the value of F at the last master solution, i.e., solving $\mathrm{VI}(F(x_M^k) + [h'(x_M^k)]^\top \mu_M^k, S_g)$. The other possibility uses a Jacobi-like approximation, where only some components of x are fixed at their values at x_M^k. In this work, we shall also consider other approximations, for example of the Josephy–Newton type [42], which approximates $F(\cdot)$ in the smooth single-valued case by $F(x_M^k)+F'(x_M^k)(\cdot - x_M^k)$. We shall also allow combinations of various approximations. In fact, in our numerical results in Section 3.3, we found that the *combination* of the Newtonian and Jacobi approximations works best for large problems of the structure considered there. In addition, and as compared to [29, 12], our framework also allows for approximations to the derivative h' (including the option of taking the fixed value $h'(x_M^k)$ as in [29, 12] but not limited to it); does not assume solvability of the subproblems; allows for inexact solutions of subproblems; gives an option of generating (cheap) additional cuts by projecting a selection of previous iterates using separation ideas [46, 75]; and can handle the general case of F being set-valued.

The rest of the chapter is organized as follows. In Section 3.1 we formally state the algorithm and discuss the approximation options for F and h', inexact solution of subproblems, and other details. Convergence analysis is given in Section 3.2. Numerical results for computing variational equilibria of game-theoretic models of electricity markets are presented in Section 3.3. In particular, we show that some specific implementations of our approach make it possible to solve problem instances which are too large to be handled by the widely used PATH solver [18, 28] applied directly to the full problem without decomposition. This is also a difference with the methods in [29, 12] where the considered examples were solved faster without decomposition than with decomposition.

3.1 The algorithmic framework

In view of (3.2), having a current multiplier estimate μ_M^k for the h-constraint, perhaps the first natural approach would be to solve the subproblem

$$\mathrm{VI}(F(\cdot) + [h'(\cdot)]^\top \mu_M^k, S_g). \tag{3.3}$$

This is a valid option indeed, but it may have drawbacks, at least for some types of problem structures. For example, if (3.3) involves a general nonlinear (and possibly nonmonotone) mapping F, it may prevent us from taking full advantage of some special structure of the set S_g (e.g., S_g may be block-separable). The same comment applies to the nonlinearity of the derivative of h. Another issue is that the set S_g in (3.3) may be unbounded (even if $S_h \cap S_g$ were bounded), in which case (3.3) is not guaranteed to have solutions if F is merely continuous/monotone. For these reasons, we shall consider various approximations to F and h' that include (3.3) itself as an option, possibly regularized by a variable-metric proximal term to induce solvability of subproblems if needed. The algorithm is as follows.

Algorithm 3.1.1. (Dantzig-Wolfe Decomposition)

1. Choose $x_S^0 \in S_g \cap S_h$, such that $h(x_S^0) < 0$ if h is not affine. Set $x_M^0 = x_S^0$. Choose $\mu_M^0 \in \mathbb{R}_+^q$ and $w_M^0 \in F(x_M^0)$. Set $k := 0$.

2. **The Subproblem:** Choose an approximation $F^k : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ of $F(\cdot)$, an approximation $H^k : \mathbb{R}^n \to \mathbb{R}^{q\times n}$ of $h'(\cdot)$, a possible modification of μ_M^k given by $\mu^k : \mathbb{R}^n \to \mathbb{R}_+^q$, and a positive (semi)definite matrix $Q_k \in \mathbb{R}^{n\times n}$. Find x_S^{k+1}, an approximate solution of the problem

$$\mathrm{VI}(\hat{F}^k, S_g), \tag{3.4}$$

$$\hat{F}^k(x) = F^k(x) + [H^k(x)]^\top \mu^k(x) + Q_k(x - x_M^k). \tag{3.5}$$

3. **The Master Problem:** Choose a finite set $X^{k+1} \subset S_g$ containing $\{x_S^0, \ldots, x_S^{k+1}\}$. Find a solution x_M^{k+1} of the problem

$$\mathrm{VI}(F, S_h \cap \operatorname{conv} X^{k+1}), \tag{3.6}$$

with the associated $w_M^{k+1} \in F(x_M^{k+1})$ and a Lagrange multiplier μ_M^{k+1} associated to the h-constraint.

4. Set $k := k + 1$ and go to Step 2.

Some comments are in order.

In Step 1, choosing a feasible starting point is needed to guarantee that the master problems (3.6) are feasible for all k. When h is not affine, the role of the condition $h(x_S^0) < 0$ is to ensure that the Slater constraint qualification holds for the master problems (3.6) for all k, so that there exist Lagrange multipliers associated to the h-constraint in (3.6). If h is affine, then (3.6) is a linearly constrained problem and the existence of Lagrange multipliers is automatic. That said, computing a (strictly) feasible starting point may be nontrivial in some applications. For this reason, Section 3.1.5 below presents a modification of the algorithm in which the h-constraints are relaxed by introducing slack variables, and computing a starting feasible point is required only for the set S_g (recall that this set is assumed to be simple in our context). Master problems (3.6)

are solved introducing simplicial parametrization of the convex hull, similarly to (2.3) in the case of linear programs.

The options for approximations F^k and H^k in the subproblems, as well as an augmented-Lagrangian type modification μ^k of the multiplier estimate μ_M^k, will be discussed in Section 3.1.1 below. As for the regularization matrix Q_k, it should generally be taken as zero if F (and then also F^k, for natural choices) is known to be strongly monotone; if strong monotonicity does not hold then Q_k should be positive definite (e.g., a multiple of the identity; but more sophisticated choices may be useful depending on the structure [57]). The notion of acceptable approximate solutions of subproblems is discussed in Section 3.1.2.

The set X^{k+1} in the master problem contains previous solutions of subproblems, but we could also add additional points. Section 3.1.3 shows that, at least if the projection onto the simpler set S_g is easy, we can compute explicitly (at negligible computational cost) points that are improvements over the previous iterates in the sense that they are closer to the solution set of (3.1).

Some remarks concerning reasonable stopping rules for Algorithm 3.1.1 will be given in Section 3.1.4.

3.1.1 Approximating the data in the subproblems

We next discuss the options for approximating the problem data in the subproblems. Roughly speaking, possible choices range from the simplest ones of taking the fixed values computed at the previous master solution x_M^k, pass through the Newtonian approximation centered at x_M^k, and arrive to taking the functions themselves ("exact approximation"). Furthermore, different options can be combined. For example, in the differentiable case, we can fix some components of the functions at x_M^k and use Newtonian approximations for the other components. In fact, we found such combinations to be the most efficient ones in our numerical results reported in Section 3.3.

To be deemed admissible, approximating objects must satisfy the following four basic conditions:

$$w_M^k \in F^k(x_M^k) \subset F(x_M^k), \tag{3.7a}$$

$$F^k(x) + [H^k(x)]^\top \mu^k(x) \quad \begin{array}{l}\text{is maximal monotone and}\\ \text{its domain contains } \operatorname{dom} F,\end{array} \tag{3.7b}$$

$$H^k(x_M^k) = h'(x_M^k), \tag{3.7c}$$

$$\mu^k(x_M^k) = \mu_M^k. \tag{3.7d}$$

The mapping approximations. As already commented, F^k estimates F near x_M^k. Some examples are:

$$F_{\text{const}}^k(x) = \{w_M^k\}, \tag{3.8a}$$

$$F_{\text{exact}}^k(x) = F(x), \tag{3.8b}$$

$$F_{\text{N}}^k(x) = F(x_M^k) + F'(x_M^k)(x - x_M^k), \tag{3.8c}$$

(the latter in the single-valued smooth case)

where the subscript N above stands for "Newton". Note that all these approximations are (maximal) monotone if F is (maximal) monotone, and that F_{const}^k is maximal monotone regardless of any assumptions.

We would like to emphasize that even if F is nonmonotone, we can always choose a maximal monotone approximation F^k. For example, taking F_{const}^k. Also, for specific applications there may exist other (more sophisticated options) of choosing a monotone approximation F^k for a nonmonotone F. One example will be discussed in the sequel in the context of VI associated to generalized Nash equilibrium problems and in our numerical results in Section 3.3.

Approximations of the derivative of the h-constraint. Similarly, the function H^k estimates the derivative h' near the point x_M^k, while preserving the monotonicity property (of derivatives of convex functions). Some examples are:

$$H_{\text{const}}^k(x) = h'(x_M^k), \tag{3.9a}$$

$$H_{\text{exact}}^k(x) = h'(x), \tag{3.9b}$$

$$H_{\text{N}}^k(x) = h'(x_M^k) + \sum_{i=1}^{q} h_i''(x_M^k)(x - x_M^k). \tag{3.9c}$$

Note that for all the cases in (3.9), because of the convexity of h, the following monotonicity property holds:

$$(H^k(y) - H^k(x))(y - x) \geq 0, \quad \text{for all } x, y \in \mathbb{R}^n. \tag{3.10}$$

Since $\mu_M^k \geq 0$, it then follows that $[H^k(x)]^\top \mu_M^k$ is also monotone. And if F^k is maximal monotone then (3.7b) holds if we take $\mu^k(x) = \mu_M^k$.

Multiplier modifications. Choices of $\mu^k(x)$ different from μ_M^k are possible if there are linear equality constraints in the definition of the set S_h (formally, in our setting this would correspond to taking two inequalities with opposite signs). Suppose that these equality constraints are given by $\tilde{h}(x) = Ax - a$, where A and a are a matrix and a vector of appropriate dimensions, respectively. We could then use the augmented Lagrangian choice for the corresponding multipliers:

$$\tilde{\mu}^k(x) = \tilde{\mu}_M^k + r_k \tilde{h}(x) = \tilde{\mu}_M^k + r_k A(x - x_M^k),$$

where $r_k > 0$ is the penalty parameter, and we took into account that $x_M^k \in S_h$ and so $\tilde{h}(x_M^k) = Ax_M^k - a = 0$. It can be seen that this choice satisfies the conditions in (3.7).

Jacobi-type approximations in the block-separable case. To conclude this section, we consider the important special case where S_g is a product of convex sets. That is, $S_g = \prod_{i=1}^m S_{g_i}$ where $S_{g_i} \subset \mathbb{R}^{n_i}$ are closed convex, $i = 1, \dots, m$, $n = \sum_{i=1}^m n_i$. Having chosen the approximations F^k and H^k, the matrix Q_k, and taking $\mu^k(x) = \mu^k_M$, we can write the function $\hat{F}^k$ for the subproblem by blocks:

$$\hat{F}^k(x) = (\hat{F}^k_1(x), \dots, \hat{F}^k_m(x)), \quad \text{with } \hat{F}^k_i(x) \in \mathbb{R}^{n_i},$$

and for every $i = 1, 2, \dots, m$, define the Jacobi-like approximations $\hat{F}^k_{\mathrm{J}_i} : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i}$,

$$\hat{F}^k_{\mathrm{J}_i}(x_i) = \hat{F}^k_i(x^k_{M_{-i}}, x_i),$$

where $(x^k_{M_{-i}}, x_i) = (x^k_{M_1}, \dots, x^k_{M_{i-1}}, x_i, x^k_{M_{i+1}}, \dots, x^k_{M_m})$ is the vector with all the blocks of variables, except for the i-th, fixed to the master solution. The corresponding estimate $\hat{F}^k_{\mathrm{J}} : \mathbb{R}^n \to \mathbb{R}^n$ for the subproblem is then given by

$$\hat{F}^k_{\mathrm{J}}(x) = (\hat{F}^k_{\mathrm{J}_1}(x_1), \dots, \hat{F}^k_{\mathrm{J}_m}(x_m)).$$

Accordingly, the objects in (3.5) take the form

$$F^k(x) = (F^k_1(x), \dots, F^k_m(x)), \quad \text{with } F^k_i(x) \in \mathbb{R}^{n_i},$$

$$H^k(x) = [H^k_1(x) | \cdots | H^k_m(x)], \quad \text{with } H^k_i(x) \in \mathbb{R}^q \times \mathbb{R}^{n_i},$$

$$Q_k = [Q_{k_{ij}}], \quad \text{with } Q_{k_{ij}} \in \mathbb{R}^{n_i} \times \mathbb{R}^{n_j}.$$

And, for each $i = 1, 2, \dots, m$, we define the Jacobi-like approximations $F^k_{\mathrm{J}_i} : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i}$ and $H^k_{\mathrm{J}_i} : \mathbb{R}^{n_i} \to \mathbb{R}^q \times \mathbb{R}^{n_i}$ by

$$F^k_{\mathrm{J}_i}(x_i) = F^k_i(x^k_{M_{-i}}, x_i), \quad H^k_{\mathrm{J}_i}(x_i) = H^k_i(x^k_{M_{-i}}, x_i).$$

It is easy to see that

$$\hat{F}^k_{\mathrm{J}_i}(x_i) = F^k_{\mathrm{J}_i}(x_i) + [H^k_{\mathrm{J}_i}(x_i)]^\top \mu^k_M + Q_{k_{ii}}(x_i - x^k_{M_i}),$$

and thus

$$\hat{F}^k_{\mathrm{J}}(x) = F^k_{\mathrm{J}}(x) + [H^k_{\mathrm{J}}(x)]^\top \mu^k_M + Q^{\mathrm{J}}_k(x - x^k_M),$$

where

$$F^k_{\mathrm{J}}(x) = (F^k_{\mathrm{J}_1}(x_1), \dots, F^k_{\mathrm{J}_m}(x_m)), \quad H^k_{\mathrm{J}}(x) = [H^k_{\mathrm{J}_1}(x_1) | \cdots | H^k_{\mathrm{J}_m}(x_m)],$$

$$Q^{\mathrm{J}}_k = \mathrm{diag}(Q_{k_{11}}, Q_{k_{22}}, \dots, Q_{k_{mm}}).$$

The functions F^k_{J} and $H^k_{\mathrm{J}}(x)$ satisfy all the properties in (3.7). Moreover, since for every $i = 1, 2, \dots, m$ and $x_i, y_i \in \mathbb{R}^{n_i}$ it holds that

$$\langle \hat{F}^k_{\mathrm{J}_i}(y_i) - \hat{F}^k_{\mathrm{J}_i}(x_i), y_i - x_i \rangle = \langle \hat{F}^k(x^k_{M_{-i}}, y_i) - \hat{F}^k(x^k_{M_{-i}}, x_i), (x^k_{M_{-i}}, y_i) - (x^k_{M_{-i}}, x_i) \rangle,$$

it follows that if $\hat{F}^k$ were monotone, strictly monotone or strongly monotone, then $\hat{F}^k_{\mathrm{J}_i}$ would inherit the same property; and therefore so would $\hat{F}^k_{\mathrm{J}}$. We again comment that in some applications (for example, VI associated to generalized Nash equilibrium problems, see Section 3.3) the full function $F(x) = (F_1(x), \ldots, F_n(x))$ can be nonmonotone but each component F_i is monotone in the variable x_i. So, even if the approximations F^k_{N} and F^k_{exact} could be nonmonotone, the approximations $F^k_{\text{N-J}}$ (where "N-J" stands for "Newton-Jacobi") and $F^k_{\text{exact-Jacobi}}$ are monotone in that case.

As we found $F^k_{\text{N-J}}$ particularly useful in our numerical experiments, we shall next state it formally. By the definition above, we have $F^k_{\mathrm{N}_i}(x) = F_i(x^k_M) + F_i'(x^k_M)(x - x^k_M)$, and so

$$\begin{aligned} F^k_{\text{N-J}_i}(x_i) &:= F^k_{\mathrm{N}_i}(x^k_{M_{-i}}, x_i) \\ &= F_i(x^k_M) + F_i'(x^k_M)((x^k_{M_{-i}}, x_i) - (x^k_{M_{-i}}, x^k_{M_i})) \\ &= F_i(x^k_M) + F_i'(x^k_M)(0, x_i - x^k_{M_i}) \\ &= F_i(x^k_M) + \nabla_{x_i} F_i(x^k_M)(x_i - x^k_{M_i}). \end{aligned}$$

Then if $F_i(x)$ is monotone in the variable x_i, we have that $F^k_{\text{N-J}_i}(x_i)$ is monotone and so is $F^k_{\text{N-J}}(x)$.

The motivation for the Jacobi approach is that we can take advantage of the separable structure of S_g when solving the subproblems even when F is not separable. Specifically, it can be seen that x^{k+1}_S solves the subproblem $\mathrm{VI}(\hat{F}^k_{\mathrm{J}}, S_g)$ if and only if the components $(x^{k+1}_S)_i$ solve $\mathrm{VI}(\hat{F}^k_{\mathrm{J}_i}, S_{g_i})$, $i = 1, \ldots, m$. Thus, the subproblems in Algorithm 3.1.1 decompose according to the structure of S_g. Clearly, such decomposition is also achieved for $F^k_{\text{N-J}}(x)$.

3.1.2 Inexact solution of subproblems

By approximate solution of subproblem (3.4) we mean computing some

$$\begin{aligned} x^{k+1}_S \in S_g \text{ such that} \quad & \langle v^{k+1}_S + e^k, y - x^{k+1}_S \rangle \geq 0 \\ & \text{for some } v^{k+1}_S \in \hat{F}^k(x^{k+1}_S) \text{ and all } y \in S_g, \end{aligned} \tag{3.11}$$

where $e^k \in \mathbb{R}^n$ is the error term accounting for inexactness. This definition of approximate solutions of variational problems was also employed, e.g., in [54, 75]. In our convergence analysis, we shall use the following two approaches to controlling inexactness. One is the "relative-error" type:

$$\langle e^k, x^k_M - x^{k+1}_S \rangle \leq \sigma \langle Q_k(x^k_M - x^{k+1}_S), x^k_M - x^{k+1}_S \rangle, \quad \sigma \in [0, 1), \tag{3.12}$$

or its stronger version

$$\|e^k\| \|x^k_M - x^{k+1}_S\| \leq \sigma \langle Q_k(x^k_M - x^{k+1}_S), x^k_M - x^{k+1}_S \rangle, \quad \sigma \in [0, 1). \tag{3.13}$$

The second rule is the "asymptotically exact" type:

$$e^k \to 0 \quad \text{as} \quad k \to \infty. \tag{3.14}$$

The first rule is more constructive, as it essentially means that the *relative* error (the ratio between the size of the error term e^k and the size of the step $x_M^k - x_S^{k+1}$) in solving the subproblems needs to be small enough but can be fixed by the value of the parameter σ (which need not tend to zero); see [78, 57] and references therein for discussions of the advantages of this relative-error approach. That said, verifying (3.12) clearly requires the explicit knowledge of e^k in (3.11). Below we explain how e^k can be constructed and the conditions (3.11) and (3.12) checked explicitly in the case of continuous $\hat{F}^k$, if we have access to the iterates of the method applied to solve $\mathrm{VI}(\hat{F}^k, S_g)$. Of course, the latter is not the case when a "black-box" solver is used. In that sense, an advantage of the "asymptotical exactness" rule (3.14) is that it can be argued that in this case the explicit knowledge of e^k in (3.11) is not necessary. The algorithm used to solve subproblems (3.4) can be truncated according to any suitable internal criteria, provided the precision is progressively tightened along the iterations of the outer Algorithm 3.1.1. This would generate, at each step, some unknown error term e^k in (3.11). But as long as the inexactness in solving the subproblems (however it is measured) asymptotically vanishes, it seems valid that the error written in any other form, for example (3.11), must also tend to zero.

Suppose now that $\hat{F}^k$ is continuous (single-valued) and strongly monotone (recall that a monotone approximation F^k of F always exists even if F is non-monotone, and strong monotonicity of $\hat{F}^k$ can be induced by adding the proximal regularization with Q_k positive definite when needed). Let an Algorithm A (any suitable algorithm for solving $\mathrm{VI}(\hat{F}^k, S_g)$) generate a sequence $\{y^{k,i}\}$ which, if continued infinitely, is known to converge to the exact solution $\bar{x}_S^{k+1}$ of the subproblem $\mathrm{VI}(\hat{F}^k, S_g)$ as $i \to \infty$. This solution is unique, because $\hat{F}^k$ is strongly monotone. As is well known, it holds that

$$\bar{x}_S^{k+1} = P_{S_g}(\bar{x}_S^{k+1} - \hat{F}^k(\bar{x}_S^{k+1})).$$

Define the auxiliary sequences $\{z^{k,i}\}$ and $\{e^{k,i}\}$ by

$$z^{k,i} = P_{S_g}(y^{k,i} - \hat{F}^k(y^{k,i})),$$

$$e^{k,i} = (z^{k,i} - \hat{F}^k(z^{k,i})) - (y^{k,i} - \hat{F}^k(y^{k,i})).$$

Since, by continuity of $\hat{F}^k$ and of the projection operator, $\{z^{k,i}\}$ also converges to $\bar{x}_S^{k+1}$ as $i \to \infty$, it holds that

$$\lim_{i \to \infty} e^{k,i} = 0.$$

Observe now that

$$z^{k,i} = P_{S_g}(y^{k,i} - \hat{F}^k(y^{k,i})) = P_{S_g}(z^{k,i} - (\hat{F}^k(z^{k,i}) + e^{k,i})),$$

which means that, at each iteration of Algorithm A, $z^{k,i}$ solves the problem $\mathrm{VI}(\hat{F}^k + e^{k,i}, S_g)$. In other words, the condition (3.11) holds at every iteration i for $x_S^{k+1} = z^{k,i}$ and $e^k = e^{k,i}$ with the known $e^{k,i}$ defined above. Since

$\{e^{k,i}\} \to 0$ as $i \to \infty$, for any reasonable criterion of measuring approximations Algorithm A would yield in a finite number of iterations an approximate solution $x_S^{k+1} = z^{k,i}$ for the subproblem VI($\hat{F}^k, S_g$) with the known, and thus controllable, error $e^k = e^{k,i}$.

The only computational issue with the presented construction is the projection onto S_g to construct the auxiliary points $z^{k,i}$. However, this projection can be explicit for some problems (e.g., onto a box). Also, it may be already computed by Algorithm A in the course of its iterative procedure anyhow. For example, one of the most natural stopping conditions for VI($\hat{F}^k, S_g$) is precisely to check whether $\|y^{k,i} - z^{k,i}\|$ is small (this is the so-called natural residual of VI [27, Chapter 1.5]; if the natural residual is zero then $y^{k,i} = z^{k,i} = \bar{x}_S^{k+1}$ is the exact solution). In particular, most (if not all) projection methods for VIs (see, e.g., [27, Chapter 12.1], [79]) compute the right-hand side in the definition of $z^{k,i}$ as part of the iterates update (perhaps scaled with a stepsize, but this can be easily accounted for) and/or compute the natural residual $(y^{k,i} - z^{k,i})$ for the stopping test. Thus, within projection methods, $z^{k,i}$ and then $e^{k,i}$ are readily available. That said, solving subproblems with increasing accuracy makes iterations progressively more expensive, of course. An interesting proposal in the context of projection methods is presented in [44], where a fixed number of projection steps is performed throughout, with verifiable error bounds. Naturally, this leads to errors which are bounded but do not tend to zero, and would require a different type of analysis from the one to be presented below. Given its clear practical importance, the issue of how to handle in our setting asymptotically nonvanishing inaccuracy in the subproblem solution is an interesting subject of future research.

3.1.3 Managing the feasible set of the master problem

The basic choice is to take $X^{k+1} = \{x_S^0, \ldots, x_S^{k+1}\}$. As already mentioned, to ensure feasibility of the master problems (3.6) it should hold that $x_M^0 = x_S^0 \in S_h \cap S_g$, and for the existence of Lagrange multipliers $h(x_S^0) < 0$ if h is not affine.

However, when F is monotone and the projection onto S_g is cheap (and this is indeed the case for many applications of interest), we can generate at negligible computational cost some additional "improved" points that are closer to the solution set than the past iterates. This procedure is based on separation/projection ideas, e.g., [46, 75].

In this strategy we explicitly state that $X^k \subset X^{k+1}$ for all k, i.e., no points are ever deleted from the feasible set of the previous master problem; points can only be added. Then since x_M^k solves VI($F, S_h \cap \mathrm{conv}\, X^k$) and $x_M^j \in S_h \cap \mathrm{conv}\, X^k$ for $j \leq k$ (since $X^j \subset X^{j+1}$), we have that for the associated $w_M^k \in F(x_M^k)$ it holds that

$$\langle w_M^k, x_M^j - x_M^k \rangle \geq 0 \quad \text{for} \quad j = 1, \ldots, k. \tag{3.15}$$

On the other hand, if $\bar{x}$ is any solution of VI($F, S_h \cap S_g$), since $x_M^k \in S_h \cap S_g$ it holds that $\langle \bar{w}, x_M^k - \bar{x} \rangle \geq 0$ where $\bar{w} \in F(\bar{x})$. Then monotonicity of F

(actually, the weaker pseudo-monotonicity property is enough here) implies that $\langle w_M^k, x_M^k - \bar{x}\rangle \geq 0$. Hence, for every k, the solution set of $\mathrm{VI}(F, S_h \cap S_g)$ lies in the halfspace

$$\{x : \langle w_M^k, x - x_M^k\rangle \leq 0\}.$$

Thus, in view of (3.15), all the previous master problem solutions are separated from the solution set of $\mathrm{VI}(F, S_h \cap S_g)$ by the hyperplane $\{x : \langle w_M^k, x - x_M^k\rangle = 0\}$. In fact, as there seem to be no reasons for the inequality (3.15) to hold as equality, the separation should be expected to be strict for most points. It is then clear that projecting onto the separating hyperplane (can also be with under- or over-relaxation), would move the previous iterates closer to the solution set, thus giving better approximations to the solution [75].

In addition, previous solutions of subproblems could be considered too, i.e., the points with the property

$$\langle w_M^k, x_S^j - x_M^k\rangle \geq 0 \quad \text{for} \quad j = 1, \ldots, k \text{ such that } x_S^j \in S_h.$$

If there are such points then they can be projected/improved also. That said, since $x_S^j \in S_h$ need not hold in general, the existence of candidates to project of this kind is not a given (unlike the case with the previous master problem solutions for which the separation property always holds).

Summarizing, we can choose any subset

$$Z^k \subset \{z \in X^k \cup \{x_M^1, \ldots, x_M^k\} : \langle w_M^k, z - x_M^k\rangle > 0\}$$

and define

$$X^{k+1} = X^k \cup \{x_S^{k+1}\} \cup \left\{ P_{S_g}\left(z - \beta_z \frac{\langle w_M^k, z - x_M^k\rangle}{\|w_M^k\|^2}\, w_M^k\right) : z \in Z^k \right\},$$

where $\beta_z \in (0, 2)$ is over/under relaxation parameter ($\beta_z = 1$ corresponds to the projection onto the separating hyperplane). See [75] for formal justifications.

3.1.4 Stopping conditions

One reasonable stopping criterion for Algorithm 3.1.1 is based on monitoring, after solving the subproblem $\mathrm{VI}(\hat{F}^k, K_g)$, the quantity

$$\Delta_k = \langle w_M^k + [h'(x_M^k)]^\top \mu_M^k, x_S^{k+1} - x_M^k\rangle. \tag{3.16}$$

The motivation for (3.16) comes from the stopping test of the cutting-plane algorithm for maximizing the dual function (2.2) in the original Dantzig–Wolfe method for the linear program (2.1) (see Section 2.1). For this problem, using the linearity of the data and the fact that $\mu_M^k \perp h(x_M^k)$ (since these solve (2.3)), we have that

$$\begin{aligned}
\Delta_k &= \langle f'(x_M^k) + h'(x_M^k)^\top \mu_M^k, x_S^{k+1} - x_M^k\rangle \\
&= f(x_S^{k+1}) - f(x_M^k) + \langle \mu_M^k, h(x_S^{k+1}) - h(x_M^k)\rangle \\
&= f(x_S^{k+1}) + \langle \mu_M^k, h(x_S^{k+1})\rangle - f(x_M^k) \\
&= \theta(\mu_M^k) - \theta^k(\mu_M^k) \leq 0,
\end{aligned}$$

i.e., Δ_k measures how well the dual function θ is approximated by its cutting-plane model θ^k at the current dual iterate μ_M^k. It is standard to stop the cutting-plane method when Δ_k becomes small enough [5, Sec. 9.3.2].

Let us now go back to the variational setting. Suppose x_S^{k+1} is an inexact solution of the subproblem $\mathrm{VI}(\hat{F}^k, K_g)$ in the sense of (3.11). Since $x_M^k \in S_g$, it then holds that

$$\langle v_S^{k+1}, x_S^{k+1} - x_M^k \rangle \leq \langle e^k, x_M^k - x_S^{k+1} \rangle.$$

We can write

$$v_S^{k+1} = u_S^{k+1} + Q_k(x_S^{k+1} - x_M^k), \text{ where } u_S^{k+1} \in F^k(x_S^{k+1}) + [H^k(x_S^{k+1})]^\top \mu^k(x_S^{k+1}).$$

Then, for the inexactness criterion (3.12), we have that

$$\begin{aligned} \langle u_S^{k+1}, x_S^{k+1} - x_M^k \rangle &\leq \langle e^k, x_M^k - x_S^{k+1} \rangle - \langle Q_k(x_S^{k+1} - x_M^k), x_S^{k+1} - x_M^k \rangle \\ &\leq -(1-\sigma)\langle Q_k(x_S^{k+1} - x_M^k), x_S^{k+1} - x_M^k \rangle. \end{aligned}$$

Now, since $F^k(x) + [H^k(x)]^\top \mu^k(x)$ is monotone and

$$w_M^k + [h'(x_M^k)]^\top \mu_M^k \in F^k(x_M^k) + [H^k(x_M^k)]^\top \mu^k(x_M^k),$$

it holds that

$$\begin{aligned} \Delta_k &= \langle w_M^k + [h'(x_M^k)]^\top \mu_M^k, x_S^{k+1} - x_M^k \rangle \\ &\leq \langle u_S^{k+1}, x_S^{k+1} - x_M^k \rangle \\ &\leq -(1-\sigma)\langle Q_k(x_S^{k+1} - x_M^k), x_S^{k+1} - x_M^k \rangle \leq 0. \end{aligned} \tag{3.17}$$

If Q_k is positive definite, then $\Delta_k = 0$ implies $x_S^{k+1} = x_M^k$, and the latter point is a solution of $\mathrm{VI}(F, S_h \cap S_g)$ (see Proposition 3.2.2 below, which also deals with case when Q_k may be positive semidefinite if $\hat{F}^k$ is strictly monotone). A value of Δ_k close to zero means that the difference between the points x_S^{k+1} and x_M^k is small, which justifies the stopping test based on Δ_k.

3.1.5 Relaxing the constraints in the master problem

We now consider the option of relaxing the h-constraints by introducing slack variables. This feature can be useful when computing a feasible starting point in $S_h \cap S_g$ is nontrivial. A starting point in S_g is still needed, but recall that it is assumed to be a simple set in our context. A similar technique had been mentioned in [29, 12], but without any theoretical analysis.

Suppose that at an iteration $k \geq 0$ we have a finite subset X^{k+1} of S_g containing the subproblems solutions $\{x_S^0, \ldots, x_S^{k+1}\}$. We define the relaxed master feasible set

$$D_k = \{(x, z) \in \operatorname{conv} X^{k+1} \times \mathbb{R}^q : h(x) \leq z\},$$

and the function $F_M^k : \mathbb{R}^n \times \mathbb{R}^q \rightrightarrows \mathbb{R}^n \times \mathbb{R}^q$ by

$$F_M^k(x, z) = F(x) \times \{\zeta_k z\},$$

where $\zeta_k > 0$ is a scalar parameter. Then the relaxed master problem consists of solving

$$\mathrm{VI}(F_M^k, D_k). \tag{3.18}$$

Note that the set D_k is always nonempty and satisfies the Slater constraint qualification automatically. Also, F_M^k is (strongly) monotone, if so is F.

The new algorithm is given below.

Algorithm 3.1.2. (Relaxed Dantzig-Wolfe Decomposition)

1. Choose $x_M^0 \in S_g$, $w_M^0 \in F(x_M^0)$ and $\mu_M^0 \in \mathbb{R}_+^q$. Set $x_M^0 = x_S^0$ and $k := 0$.
2. **The Subproblem:** Choose the function $\hat{F}^k$ as in Algorithm 3.1.1 and find x_S^{k+1}, a solution of the problem $\mathrm{VI}(\hat{F}^k, S_g)$.
3. **The Master Problem:** Choose the set X^{k+1} as in Algorithm 3.1.1 and the parameter $\zeta_k > 0$. Find a solution (x_M^{k+1}, z^{k+1}) of the problem (3.18), with the associated $w_M^{k+1} \in F(x_M^{k+1})$ and a Lagrange multiplier μ_M^{k+1} associated to the h-constraint.
4. Set $k := k + 1$ and go to Step 2.

Note that solutions of the subproblems and of the master problems belong to the set S_g, but not necessarily to S_h. In Proposition 3.2.3 below, we shall see that the z-component of the master solution is actually uniquely defined and is of the form $z^{k+1} = \mu_M^{k+1}/\zeta_k$.

3.2 Convergence analysis

We first formalize the arguments that show that the algorithm is well-defined, i.e., that all the subproblems and all the master problems have solutions.

As discussed above, we can always choose F^k to be maximal monotone (even if F is not), so that $\hat{F}^k$ would be maximal monotone with its domain containing the domain of F. Moreover, $\hat{F}^k$ can always be made strongly monotone by taking the matrix Q_k positive definite when needed. Then maximal monotonicity and strong monotonicity of $\hat{F}^k$ ensure the existence of the unique solution to subproblem $\mathrm{VI}(\hat{F}^k, S_g)$ (by [65, Theorem 5]). In addition, as already discussed above, our choice of the starting points in both Algorithm 3.1.1 and Algorithm 3.1.2 ensures that all the master problems are always feasible and satisfy constraint qualifications [77], so that there exist Lagrange multipliers associated to the solutions (if any). Now, the master problems are variational inequalities either with a continuous mapping or with a maximal monotone one over nonempty compact feasible sets. Solutions to this type of problems exist, by [27, Corollary 2.2.5] and [65, Theorem 5], respectively.

We start our convergence analysis by establishing some key properties of the master problems solutions in Algorithm 3.1.1.

Proposition 3.2.1. *For a solution x_M^{k+1} of the master problem $VI(F, S_h \cap conv\, X^{k+1})$ in Algorithm 3.1.1, the following assertions hold:*

1. *If h is affine or if there exists $\hat{x} \in conv\, X^{k+1}$ such that $h(\hat{x}) < 0$, then there exists a Lagrange multiplier $\mu_M^{k+1} \in \mathbb{R}^q_+$ associated to the h-constraint.*

2. *For any such multiplier μ_M^{k+1} it holds that x_M^{k+1} solves $VI(F(\cdot) + [h'(\cdot)]^\top \mu_M^{k+1}, conv\, X^{k+1})$.*

3. *If for any $x \in S_g$ and any $v \in [F(x_M^{k+1}) + [h'(x_M^{k+1})]^\top \mu_M^{k+1}] \cap [-\mathcal{N}_{conv\, X^{k+1}}(x_M^{k+1})]$ it holds that $\langle v, x - x_M^{k+1} \rangle < 0$ then $x \notin conv\, X^{k+1}$.*

4. *If x_M^{k+1} is a solution of $VI(F(\cdot) + [h'(\cdot)]^\top \mu_M^{k+1}, S_g)$ then x_M^{k+1} solves $VI(F, S_g \cap S_h)$.*

5. *On the next iteration ($k := k + 1$), if x_M^k is a solution of the subproblem $VI(\hat{F}^k, S_g)$ then x_M^k solves $VI(F, S_g \cap S_h)$.*

Proof. 1. Since $\operatorname{conv} X^{k+1}$ is a polyhedral set, the linearity of h in the first case or the Slater constraint qualification in the second case guarantee that

$$\mathcal{N}_{S_h \cap \operatorname{conv} X^{k+1}}(x_M^{k+1}) = \mathcal{N}_{S_h}(x_M^{k+1}) + \mathcal{N}_{\operatorname{conv} X^{k+1}}(x_M^{k+1})$$

and

$$\mathcal{N}_{S_h}(x_M^{k+1}) = \{[h'(x_M^{k+1})]^\top \mu : \mu \in \mathbb{R}^q_+, \mu \perp h(x_M^{k+1})\}.$$

Then, since x_M^{k+1} solves $\mathrm{VI}(F, S_h \cap \operatorname{conv} X^{k+1})$, we have that

$$\begin{aligned} 0 \;&\in\; F(x_M^{k+1}) + \mathcal{N}_{S_h \cap \operatorname{conv} X^{k+1}}(x_M^{k+1}) \\ &=\; F(x_M^{k+1}) + \{[h'(x_M^{k+1})]^\top \mu : \mu \in \mathbb{R}^q_+, \mu \perp h(x_M^{k+1})\} \\ &\quad + \mathcal{N}_{\operatorname{conv} X^{k+1}}(x_M^{k+1}), \end{aligned}$$

which means the existence of the multiplier μ_M^{k+1} in question.

2. From the first part we have, in particular, that

$$0 \in F(x_M^{k+1}) + [h'(x_M^{k+1})]^\top \mu_M^{k+1} + \mathcal{N}_{\operatorname{conv} X^{k+1}}(x_M^{k+1}),$$

which means that x_M^{k+1} solves $\mathrm{VI}(F(\cdot) + [h'(\cdot)]^\top \mu_M^{k+1}, \operatorname{conv} X^{k+1})$, as claimed.

3. Note that any v in question can serve as an element associated to x_M^{k+1} which verifies that the latter is a solution of

$$\mathrm{VI}(F(\cdot) + [h'(\cdot)]^\top \mu_M^{k+1}, \operatorname{conv} X^{k+1}).$$

In other words, it holds that $\langle v, x - x_M^{k+1} \rangle \geq 0$ for all $x \in \operatorname{conv} X^{k+1}$. Thus, if this inequality does not hold for some $x \in \mathbb{R}^n$ (in particular, for some $x \in S_g$), it must be the case that $x \notin \operatorname{conv} X^{k+1}$.

4. Suppose now that x_M^{k+1} solves $\mathrm{VI}(F + [h'(\cdot)]^\top \mu_M^{k+1}, S_g)$, i.e.,

$$0 \in F(x_M^{k+1}) + [h'(x_M^{k+1})]^\top \mu_M^{k+1} + \mathcal{N}_{S_g}(x_M^{k+1}).$$

Since μ_M^{k+1} is a Lagrange multiplier associated to the h-constraint, we have

$$[h'(x_M^{k+1})]^\top \mu_M^{k+1} \in \mathcal{N}_{S_h}(x_M^{k+1}),$$

and hence,

$$0 \in F(x_M^{k+1}) + \mathcal{N}_{S_h}(x_M^{k+1}) + \mathcal{N}_{S_g}(x_M^{k+1}) \subset F(x_M^{k+1}) + \mathcal{N}_{S_h \cap S_g}(x_M^{k+1}),$$

which establishes the fourth assertion.

5. Finally, since by (3.7) it holds that

$$\hat{F}^k(x_M^k) \subset F(x_M^k) + [h'(x_M^k)]^\top \mu_M^k,$$

if x_M^k solves the subproblem $\mathrm{VI}(\hat{F}^k, S_g)$, then the previous item implies that it solves our problem $\mathrm{VI}(F, S_g \cap S_h)$.

□

Note that, by the third item of Proposition 3.2.1, it follows that for the gap defined in (3.16) whenever $\Delta_k < 0$ we have that $x_S^{k+1} \notin \operatorname{conv} X^k$. Thus, as long as $\Delta_k < 0$, the feasible set of the master problem keeps on growing, improving the approximation of the set S_g. If the subproblems are solved exactly then $x_S^{k+1} \in \operatorname{conv} X^k$ for some k implies that $\Delta_k \geq 0$ (actually $\Delta_k = 0$). Then, if $\hat{F}^k$ is at least strictly monotone (which can always be ensured by taking the matrix Q_k positive definite if needed) it holds that $x_M^k = x_S^{k+1}$ is a solution of the problem $\mathrm{VI}(F, S_g \cap S_h)$.

Proposition 3.2.2. *Let $\hat{F}^k$ be strictly monotone (e.g., Q_k is positive definite). Suppose that in Algorithm 3.1.1 it holds that $\Delta_k \geq \langle e^k, x_M^k - x_S^{k+1} \rangle$ for some iteration index k. Then $\Delta_k = 0$ and $x_M^k = x_S^{k+1}$ solves VI$(F, S_g \cap S_h)$.*

Proof. Since x_S^{k+1} is an approximate solution, with error e^k, of the subproblem $\mathrm{VI}(\hat{F}^k, S_g)$ in the sense of (3.11), and since $x_M^k \in S_g$, for the associated $v_S^{k+1} \in \hat{F}^k(x_S^{k+1})$ it holds that

$$\langle e^k, x_M^k - x_S^{k+1} \rangle \geq \langle v_S^{k+1}, x_S^{k+1} - x_M^k \rangle. \tag{3.19}$$

Then, by the definition of Δ_k in (3.16), by the monotonicity of $\hat{F}^k$, and by the fact that $w_M^k + [h'(x_M^k)]^\top \mu_M^k \in \hat{F}^k(x_M^k)$, it holds that

$$\begin{aligned} \Delta_k &= \langle w_M^k + [h'(x_M^k)]^\top \mu_M^k, x_S^{k+1} - x_M^k \rangle \\ &\leq \langle v_S^{k+1}, x_S^{k+1} - x_M^k \rangle \\ &\leq \langle e^k, x_M^k - x_S^{k+1} \rangle, \end{aligned} \tag{3.20}$$

where (3.19) was also used. But then the assumption that $\Delta_k \geq \langle e^k, x_M^k - x_S^{k+1}\rangle$ implies that

$$\Delta_k = \langle e^k, x_M^k - x_S^{k+1}\rangle.$$

Then by substituting the expression for Δ_k into the left-hand side of (3.19) we obtain that

$$\langle w_M^k + [h_2'(x_M^k)]^\top \mu_M^k - v_S^{k+1}, x_S^{k+1} - x_M^k\rangle \geq 0.$$

Strict monotonicity of $\hat{F}^k$ then implies that $x_S^{k+1} = x_M^k$. Obviously, it then holds that $\Delta_k = 0$. Also, since $x_S^{k+1} = x_M^k$ solves the subproblem $\mathrm{VI}(\hat{F}^k, S_g)$, Proposition 3.2.1 implies that this point is a solution of $\mathrm{VI}(F, S_g \cap S_h)$. □

We next establish the properties of solutions of the relaxed master problems.

Proposition 3.2.3. *For a solution (x_M^{k+1}, z^{k+1}) of the relaxed master problem $VI(F_M^k, D_k)$ in Algorithm 3.1.2, the following assertions hold:*

1. *There exists the unique Lagrange multiplier μ_M^{k+1} associated to the h-constraint. Moreover, it holds that*

$$\begin{gathered} z^{k+1} = \mu_M^{k+1}/\zeta_k, \\ 0 \leq \mu_M^{k+1} \perp h(x_M^{k+1}) - \mu_M^{k+1}/\zeta_k \leq 0. \end{gathered} \tag{3.21}$$

2. *The point x_M^{k+1} solves $VI(F(\cdot) + [h'(\cdot)]^\intercal \mu_M^{k+1}, conv\, X^{k+1})$.*

3. *If for any $x \in S_g$ and any $v \in [F(x_M^{k+1}) + [h'(x_M^{k+1})]^\top \mu_M^{k+1}] \cap [-\mathcal{N}_{conv\, X^{k+1}}(x_M^{k+1})]$ it holds that $\langle v, x - x_M^{k+1}\rangle < 0$, then $x \notin conv\, X^{k+1}$.*

Proof. We have that

$$0 \in F(x_M^{k+1}) \times \{\zeta_k z^{k+1}\} + \mathcal{N}_{D_k}(x_M^{k+1}, z^{k+1}).$$

Since $\operatorname{conv} X^{k+1}$ is a polyhedral set, it is easy to see that the constraints of D_k satisfy the Slater constraint qualification. Therefore,

$$\begin{aligned} \mathcal{N}_{D_k}(x_M^{k+1}, z^{k+1}) = &\left\{ \begin{bmatrix} [h'(x_M^{k+1})]^\top \\ -I \end{bmatrix} \mu : \begin{array}{c} \mu \in \mathbb{R}^q_+ \\ 0 \leq \mu \perp h(x_M^{k+1}) - z^{k+1} \leq 0 \end{array} \right\} \\ &+ \left\{ \begin{pmatrix} d \\ 0 \end{pmatrix} : d \in \mathcal{N}_{\operatorname{conv} X^{k+1}}(x_M^{k+1}) \right\}. \end{aligned}$$

In particular,

$$\zeta_k z^{k+1} - \mu = 0,$$

for any multiplier μ associated to the h-constraints. Hence, μ_M^{k+1} is uniquely defined and satisfies (3.21).

Also, it holds that

$$0 \in F(x_M^{k+1}) + [h'(x_M^{k+1})]^\top \mu_M^{k+1} + \mathcal{N}_{\operatorname{conv} X^{k+1}}(x_M^{k+1}),$$

which establishes the second assertion.

The last assertion follows from the same considerations as those used in Proposition 3.2.1 for Algorithm 3.1.1. □

We note that for Algorithm 3.1.2 the condition $x_S^{k+1} = x_M^k$ no longer implies that x_M^k solves VI$(F, S_h \cap S_g)$ (although $\Delta_k = 0$ still implies that $x_S^{k+1} = x_M^k$ when $\hat{F}^k$ is strictly monotone). This is the price to pay for the convenience of relaxing the h-constraint, as the master problem solutions x_M^k may no longer belong to S_h. Rather, the sequence approach this set asymptotically.

We are now in position to state the main convergence results for the Dantzig–Wolfe schemes described above.

Theorem 3.2.4. *Let the mapping F be (possibly set-valued) maximal monotone or single-valued continuous, and let the function h be convex and continuously differentiable. Suppose the sequence $\{(x_M^k, \mu_M^k, x_S^{k+1})\}$ generated by Algorithm 3.1.1 or Algorithm 3.1.2, with the subproblems* (3.4) *solved approximately in the sense of* (3.11) *with the associated error sequence $\{e^k\}$ satisfying* (3.12) *or* (3.14)*, is bounded. In the case of Algorithm 3.1.2, let $\{\zeta_k\} \to +\infty$ as $k \to \infty$.*

Then it holds that

1. *The sequence $\{\Delta_k\}$ converges to zero.*
2. *If the elements of $\{\hat{F}^k\}$ are strongly monotone uniformly with respect to k and the approximation rule* (3.14) *is used, or if the matrices in $\{Q_k\}$ are uniformly positive definite and either* (3.12) *or* (3.14) *is used, then $\lim_{k\to\infty} \|x_S^{k+1} - x_M^k\| = 0$.*
3. *If $\lim_{k\to\infty} \|x_S^{k+1} - x_M^k\| = 0$, the sequence $\{Q_k\}$ is bounded, the families of functions $\{F^k\}$, $\{H^k\}$ and $\{\mu^k\}$ are equicontinuous on compact sets, and in the case of the relative-error inexactness rule condition* (3.12) *is strengthened to* (3.13)*, then for every accumulation point $(\bar{x}, \bar{\mu})$ of the sequence $\{(x_M^k, \mu_M^k)\}$ the point $\bar{x}$ is a solution of VI$(F, S_h \cap S_g)$ while $\bar{\mu}$ is a multiplier associated to the h-constraint.*

Proof. 1. Using (3.17) in the case when subproblems are solved inexactly according to the rule (3.12), and (3.20) if the rule (3.14) is employed, we see that

$$\bar{\Delta} = \liminf_{k\to\infty} \Delta_k \le \limsup_{k\to\infty} \Delta_k \le 0. \tag{3.22}$$

Let $\{k_j\}$ be any subsequence of indices such that $\lim_{j\to\infty} \Delta_{k_j} = \bar{\Delta}$. Passing onto a further subsequence, if necessary, we can assume that $\{(x_M^{k_j}, \mu_M^{k_j}, x_S^{k_j+1})\} \to (\bar{x}, \bar{\mu}, \hat{x})$. Also, since under the stated assumptions F is locally bounded on S_g, the sequence $\{w_M^{k_j}\}$ is bounded and we can assume that $\{w_M^{k_j}\} \to \bar{w}$.

By definition (3.16), we have that

$$\lim_{j\to\infty} \Delta_{k_j} = \bar{\Delta} = \langle \bar{w} + [h'(\bar{x})]^\top \bar{\mu}, \hat{x} - \bar{x} \rangle.$$

Fix any index j. Then for every $i > j$ we have that $x_S^{k_j+1} \in X^{k_i}$. As a result, by the second item of Proposition 3.2.1 in case of Algorithm 3.1.1

or of Proposition 3.2.3 in case of Algorithm 3.1.2, it holds that

$$\langle w_M^{k_i} + [h'(x_M^{k_i})]^\top \mu_M^{k_i}, x_S^{k_j+1} - x_M^{k_i} \rangle \geq 0.$$

Passing onto the limit as $i \to \infty$ in the relation above, we conclude that

$$\langle \bar{w} + [h'(\bar{x})]^\top \bar{\mu}, x_S^{k_j+1} - \bar{x} \rangle \geq 0.$$

Now passing onto the limit as $j \to \infty$ in the latter relation, we obtain that

$$\langle \bar{w} + [h'(\bar{x})]^\top \bar{\mu}, \hat{x} - \bar{x} \rangle \geq 0.$$

Hence, $\bar{\Delta} \geq 0$. Together with (3.22) this proves the first assertion.

2. Since x_S^{k+1} solves approximately VI($\hat{F}^k, S_g$) in the sense of (3.11), there exists $v_S^{k+1} \in \hat{F}^k(x_S^{k+1}) \cap [-\mathcal{N}_{S_g}](x_S^{k+1})$ such that

$$\langle v_S^{k+1} + e^k, x_M^k - x_S^{k+1} \rangle \geq 0.$$

Then denoting $v_M^k = w_M^k + [h'(x_M^k)]^\top \mu_M^k \in \hat{F}^k(x_M^k)$ we have that

$$\begin{aligned} -\Delta_k &= \langle v_S^{k+1} - v_M^k, x_S^{k+1} - x_M^k \rangle - \langle v_S^{k+1}, x_S^{k+1} - x_M^k \rangle \\ &\geq \langle v_S^{k+1} - v_M^k, x_S^{k+1} - x_M^k \rangle + \langle e^k, x_S^{k+1} - x_M^k \rangle \\ &\geq c\|x_S^{k+1} - x_M^k\|^2 + \langle e^k, x_S^{k+1} - x_M^k \rangle, \end{aligned}$$

where $c > 0$ is the modulus of strong monotonicity of $\hat{F}^k$, independent of k. For the approximation rule (3.14) (that is $\{e^k\} \to 0$), since $\Delta_k \to 0$ as established above, it follows that $\|x_S^{k+1} - x_M^k\| \to 0$ as $k \to \infty$. The same conclusion holds for the choice of uniformly positive definite Q_k, as in that case the approximations $\hat{F}^k$ are uniformly strongly monotone. When the inexactness rule (3.12) is used, the assertion follows from (3.17) and the fact that $\Delta_k \to 0$.

3. Let $(\bar{x}, \bar{\mu})$ be an accumulation point of $\{(x_M^k, \mu_M^k)\}$ and let $\{(x_M^{k_j}, \mu_M^{k_j})\} \to (\bar{x}, \bar{\mu})$ be any associated convergent subsequence. By construction of the algorithm, the basic continuity argument implies that $\bar{x} \in S_g$ and $\bar{\mu} \in \mathbb{R}^q_+$. Since $x_S^{k_j+1}$ solves approximately VI($\hat{F}^{k_j}, S_g$) in the sense of (3.11), there exists $v_S^{k_j+1} \in \hat{F}^{k_j}(x_S^{k_j+1}) \cap [-\mathcal{N}_{S_g}(x_S^{k_j+1})]$ such that

$$\langle v_S^{k_j+1} + e^{k_j}, x - x_S^{k_j+1} \rangle \geq 0 \quad \text{for all } x \in S_g. \tag{3.23}$$

Since the families $\{F^k\}$, $\{H^k\}$ and $\{\mu^k\}$ are equicontinuous and the matrices Q_k are bounded, the family $\{\hat{F}^k\}$ remains equicontinuous on compact sets. Then, on an open ball containing the sequences $\{x_S^{k_j+1}\}$, $\{x_M^{k_j}\}$ and the point $\bar{x}$, for every $\varepsilon > 0$ there is $\delta > 0$ such that $\|x - y\| < \delta$ implies $d_H(\hat{F}^k(x), \hat{F}^k(y)) < \varepsilon$, for every k. Since $\|x_S^{k_j+1} - x_M^{k_j}\| \to 0$, there is an

index J such that for every $j > J$ the relation $\|x_S^{k_j+1} - x_M^{k_j}\| < \delta$ holds, and thus there exists $u_M^{k_j} \in F(x_M^{k_j})$ such that

$$\|u_M^{k_j} + [h'(x_M^{k_j})]^\top \mu_M^{k_j} - v_S^{k_j+1}\| < \varepsilon.$$

On the other hand, since under the stated assumptions F is locally bounded on S_g and outer semicontinuous, we can assume that the sequence $\{u_M^{k_j}\}$ converges to a point $\bar{u} \in F(\bar{x})$. Then

$$\lim_{k\to\infty} v_S^{k_j+1} = \bar{u} + [h'(\bar{x})]^\top \bar{\mu}. \tag{3.24}$$

Next, note that $\{e^k\} \to 0$. In the case of the inexactness rule (3.14) this is explicit. In the case of rule (3.13) it is an obvious consequence since $\{Q_k\}$ is bounded, $\|x_M^k - x_S^{k+1}\| \to 0$, and the right-hand side of (3.13) is quadratic in the latter quantity while the left-hand side is linear.

Now passing onto the limit in (3.23) as $j \to \infty$ and using (3.24), we obtain that

$$\langle \bar{u}, x - \bar{x}\rangle + \langle [h'(\bar{x})]^\top \bar{\mu}, x - \bar{x}\rangle \geq 0 \quad \text{for all } x \in S_g. \tag{3.25}$$

Since $\bar{\mu} \geq 0$, the convexity of h implies that

$$\begin{aligned} \langle \bar{\mu}, h(x) - h(\bar{x})\rangle &\geq \langle \bar{\mu}, h'(\bar{x})(x - \bar{x})\rangle \\ &= \langle [h'(\bar{x})]^\top \bar{\mu}, x - \bar{x}\rangle. \end{aligned}$$

Then, by (3.25), we obtain that

$$\langle \bar{u}, x - \bar{x}\rangle + \langle \bar{\mu}, h(x) - h(\bar{x})\rangle \geq 0 \quad \text{for all } x \in S_g.$$

It then holds that

$$\langle \bar{u}, x - \bar{x}\rangle \geq \langle \bar{\mu}, h(\bar{x})\rangle \quad \text{for all } x \in S_h \cap S_g. \tag{3.26}$$

For a sequence generated by Algorithm 3.1.1, $\{x_M^k\} \subset S_h \cap S_g$ and

$$\langle \mu_M^k, h(x_M^k)\rangle = 0$$

for all k. Hence, by continuity, $\bar{x} \in S_h \cap S_g$ and $\langle \bar{\mu}, h(\bar{x})\rangle = 0$. For a sequence generated by Algorithm 3.1.2, taking the limit in (3.21) as $k \to \infty$ and recalling the parameter choice $\{\zeta_k\} \to +\infty$, it again follows that $\bar{x} \in S_h \cap S_g$ and $\langle \bar{\mu}, h(\bar{x})\rangle = 0$.

In either case, we have $\bar{x} \in S_h \cap S_g$ and $\bar{u} \in F(\bar{x})$, with (3.26) yielding

$$\langle \bar{u}, x - \bar{x}\rangle \geq 0 \quad \text{for all } x \in S_h \cap S_g,$$

i.e., $\bar{x}$ is a solution of $\mathrm{VI}(F, S_h \cap S_g)$, as stated. The fact that $\bar{\mu}$ is a Lagrange multiplier associated to the h-constraint follows from (3.25).

□

It is clear that the families of functions $\{F^k_{\text{const}}\}$, $\{H^k_{\text{const}}\}$ and $\{H^k_{\text{exact}}\}$ defined above are equicontinuous on compact sets. For a bounded sequence $\{x^k_M\}$, both $\{F^k_{\text{N}}\}$ and $\{H^k_{\text{N}}\}$ are equicontinuous. When F is single-valued and continuous, the family $\{F^k_{\text{exact}}\}$ is equicontinuous. Finally, $\mu^k(x) = \mu^k_M$ is always equicontinuous while the augmented Lagrangian option $\tilde{\mu}^k(x)$ for linear constraints is equicontinuous if the sequence of penalization parameters $\{r_k\}$ is bounded.

3.3 Computational experiments

In this section, we describe a simplified game-theoretic model for electricity markets, and present our numerical results for computing the associated variational equilibria. For more sophisticated but related models we refer to [43, 44, 45]. For the purposes of our development here, the version considered is sufficient.

3.3.1 Energy markets as generalized Nash games

Let N_a agents generate electric energy for sale. The ith agent owns n_i plants whose total generation is represented by a vector $q^i \in \mathbb{R}^{n_i}_+$. The energy owned by this agent is the sum of the generation of all of the agent plants; accordingly, we define matrices S^i given by a row with n_i entries all equal to one, so that

$$S^i q^i = \sum_{k=1}^{n_i} q^i_k \,.$$

We suppose an *inverse-demand* function is available: the unitary energy price in the market depends on the total amount of energy produced by all agents. We model this relation by means of a quadratic concave function of one variable, that is the total energy, so $p : \mathbb{R} \to \mathbb{R}$. The exogenous coefficients defining this quadratic function are market-dependent and are given below.

The vector of all the agents' generation is denoted by $q^{-0} = (q^1, q^2, \ldots, q^{N_a}) \in \mathbb{R}^n$ (this peculiar notation will be clear soon). The total amount of energy available in the market, denoted below by $S^{-0}q^{-0}$, is the sum of the generation of all of the plants in the market:

$$S^{-0}q^{-0} = \sum_{i=1}^{N_a} S^i q^i = \sum_{i=1}^{N_a} \sum_{k=1}^{n_i} q^i_k \,.$$

Since the price depends on the total energy, the i-th agent will be paid

$$p\big(S^{-0}q^{-0}\big)S^i q^i.$$

If, to generate the amount q^i, the agent incurs an operating (convex) cost $c^i(q^i)$, the agent's profit is given by

$$p\big(S^{-0}q^{-0}\big)S^i q^i - c^i(q^i).$$

The profit of each agent depends on the generation level of all the agents in the market. In turn, each generation level is constrained by technological limitations of the power plants: for certain sets $X^i \subset \mathbb{R}^{n_i}$, the relation $q^i \in X^i$ must hold. In our simplified modeling, $X^i = [0, U^i]$ for some $U^i \in \mathbb{R}^{n_i}_+$, noting that in a realistic model the set X^i is given by complex relations expressing how different technologies (thermal, nuclear, hydraulic, eolic) generate power.

Remark 3.3.1. *The core difficulty for solution methods that do not use decomposition resides precisely in the fact that they handle the set* $\prod_{i=1}^{N_a} X^i$ *as a whole. From a numerical point of view, this usually means dealing* simultaneously *with mixed-integer variables and nonconvex relations. By contrast, a suitable decomposition method handles the difficulties by considering separately each technology (only thermal, only nuclear, etc.), dealing with each set* X^i *individually. As a result, an individual subproblem becomes "more computationally tractable"; for example, involving only affine functions and mixed-integer variables, or only nonlinear functions with continuous variables. Such separation of difficulties considerably simplifies the numerical solution of large problems. In a somewhat different context, this is also confirmed by our results below.*

An additional constraint for the generation levels q^i refers to the fact that agents are encouraged to satisfy the market demand $d > 0$. We let $q^0 \geq 0$ denote a scalar slack variable, measuring the deficit of energy in the market, sometimes called *load shedding*. Then, if for each agent i, the vector

$$q^{-i} = (q^0, q^1, \ldots, q^{i-1}, q^{i+1}, \ldots, q^{N_a})$$

denotes the generation level of all the other agents, including load shedding, the relation

$$q^i \in \mathcal{S}(q^{-i}) = X^i \cap \left\{ w^i \in \mathbb{R}^{n_i} : q^0 + \sum_{i \neq j = 1}^{N_a} \sum_{k=1}^{n_j} q^j_k + \sum_{k=1}^{n_i} w^i_k = d \right\}$$

must hold.

Summing up, the i-th agent tries to maximize profit by solving the (concave) problem

$$\max \quad p\big(S^{-0}q^{-0}\big)S^i q^i - c^i(q^i) \quad \text{s.t.} \quad q^i \in \mathcal{S}(q^{-i}).$$

The coordination, or regulation, of the market is done by the Independent System Operator (ISO), whose actions in the market are considered as those of an additional player (this is a so-called *bounded rationality* model). Accordingly, letting the ISO be player number 0, if the energy deficit is penalized with a price $P > 0$, the ISO tries to maximize the social welfare by solving

$$\max \quad p\big(S^{-0}q^{-0}\big)S^i q^{-0} - \sum_{i=1}^{N_a} c^i(q^i) - Pq^0 \quad \text{s.t.} \quad q^0 \in \mathcal{S}(q^{-0}),$$

where, having a maximal allowed level of load shedding U^0,

$$\mathcal{S}(q^{-0}) = \left\{ w^0 \in \mathbb{R} : 0 \leq w^0 \leq U^0, \quad w^0 + S^{-0}q^{-0} = w^0 + \sum_{i=1}^{N_a} S^i q^i = d \right\}.$$

As a result, for $i = 0, \dots, N_a$, the convex functions $f_i : \mathbb{R}^{1+n} \to \mathbb{R}$ given by

$$\begin{aligned} f_0(q) &= Pq^0 - \sum_{i=1}^{N_a} c^i(q^i) - p\big(S^{-0}q^{-0}\big)S^{-0}q^{-0} \\ f_i(q) &= c^i(q^i) - p\big(S^{-0}q^{-0}\big)S^i q^i, \qquad\qquad i = 1, \dots, N_a\,, \end{aligned}$$

define a generalized Nash game with N_a+1 players (the ISO and the N_a agents). In this game, each player tries to maximize profit by solving

$$\min \quad f_i(q^{-i}, q^i) \quad \text{s.t.} \quad q^i \in \mathcal{S}(q^{-i}).$$

With respect to the notation in (2.16), we have the relations

$$x^i \leftrightarrow q^i\,, \quad \text{and} \quad \mathcal{S} \leftrightarrow q^0 + \sum_{i=1}^{N} S^i q^i = d.$$

As discussed in Chapter 2, finding a generalized Nash equilibrium (GNE) of this game is equivalent to solving a quasi-variational inequality problem, see [25]. The same considerations as in Chapter 2 apply. We briefly recall them here. Quasi-variational problems are very hard to solve. Fortunately, in our case, it is possible to compute some GNE points (not all) by solving a variational inequality instead. These points are called *variational equilibria* and have some good/important properties from the economic point of view, see [26, 47] and again Chapter 2. For our problem, it is shown in [48] that variational equilibria are solutions to $\mathrm{VI}(F, S_g \cap S_h)$ where

$$F(q) = \big(\nabla_{q^0} f_0(q), \nabla_{q^1} f_1(q), \dots, \nabla_{q^{N_a}} f_{N_a}(q)\big)\,,$$

$$S_g = [0, U^0] \times \prod_{i=1}^{N_a} X^i \quad \text{and} \quad S_h = \{q \in \mathbb{R}^{1+n} : q^0 + S^{-0}q^{-0} = d\}\,.$$

In our model each function $f_i(q)$ is convex and differentiable in the variable q^i. So the function $\nabla_{q^i} f_i(q)$ is monotone in the i-th component of the variable q, but in general it is not monotone on the full variable q. Therefore, the singled-valued function F defining the variational problem (3.1) is nonmonotone. Observe that it also couples all the variables.

Our Dantzig-Wolfe strategy can be applied to nonmonotone single-valued functions F, simply by taking monotone approximations to F in the subproblems. In particular, any family F^k_{const}, $F^k_{\text{N-J}}$, or $F^k_{\text{exact-J}}$ can be used.

Another specificity of our game is that the demand constraint, that is the set S_h, couples all variables q^i. Without this constraint, the feasible set would be separable. This makes the considered game particularly suitable for application of our decomposition schemes.

Remark 3.3.2. *As discussed above, the model considered here is simplistic in some features; it is mostly meant to exhibit the interest of using decomposition*

schemes for problems with feasible sets having certain types of structure. In particular, the shared constraint in our model refers to satisfaction of the (exogenous) demand, but alternative joint constraints, like the ones in [38], could also be considered.

3.3.2 Battery of problems

We implemented Algorithm 3.1.1 in Matlab version 7.11(R2010b). The runs were done on a PC operating under Ubuntu 11.04 with a Core(TM)2 Duo 2.00GHz processor and 4GB of memory.

We created six market configurations of the generalized Nash game, by taking $N_a = 5$ agents and considering a mix with n power plants, for

$$n \in \{100, 250, 1000, 2500, 5000, 10000\}.$$

As n increases, the configurations become harder and harder to solve directly, without decomposition. Also the subproblems become harder, as we assume that each agent owns the same number of plants $n_i = n/5$, for $i = 1, \ldots, N_a = 5$.

Other values of the model parameters are as follows.

1. The entries of the maximum generation capacity vector U^i are random numbers in $[0, 10]$ while the maximum allowed deficit is fixed to $U^0 = 5$.

2. The demand is taken equal to $d = 0.8 \sum_{i=1}^{N_a} U^i$, corresponding to 80% of the market generation capacity.

3. The deficit price is set at $P = 120$.

4. The inverse-demand function defines the unitary price p as a quadratic concave function such that $p(0) = P$, $p'(0) = 0$ and $p(1.5d) = 0$.

5. The operating cost is of the form

$$c^i(q^i) = {b^i}^\top q^i + \frac{1}{2} {q^i}^\top M_i q^i$$

   where $b^i \in \mathbb{R}^{n_i}$ and $M_i \in \mathbb{R}^{n_i \times n_i}$ is a diagonal positive definite matrix. The corresponding values are generated randomly between $[30, 60]$ and $[0.4, 0.8]$, respectively.

With this data, the simplified model is set up in a manner ensuring that at a variational equilibrium of the game will have no deficit ($q^0 = 0$) and the price will be equal to $p(d) = P(1 - \frac{1}{1.5^2})$. As a way of ensuring correctness of the implementation, we checked that these values were obtained in all of our numerical results below.

3.3.3 Numerical results

For each of the six market configurations, we randomly generated 10 problem instances. For each instance, we apply Algorithm 3.1.1 using five approximations for F:

$$F_{\text{const}}\,, \quad F_{\text{N}}\,, \quad F_{\text{exact}} = F\,, \quad F_{\text{N-J}}\,, \quad F_{\text{J}}\,.$$

Since the h-constraint is linear, we used the exact family H_{exact} and likewise for the multipliers.

We also tried to solve the problem directly, without decomposition, using PATH [18, 28]. For the two largest configurations ($n = 5000, 10000$) PATH could no longer be used, stopping by lack of memory. With our computer and for the considered instances, when $n = 5000$ the solver stalled after about 4 hours. Also, since the larger configurations become time consuming, for these we only run the faster decomposition alternatives, in our case F_{const}, F_{J} and $F_{\text{N-J}}$.

Regarding specifics of the implementation of Algorithm 3.1.1, all the subproblems and master problems are themselves solved using PATH. Our focus here is on comparing various approximation options; for this reason we do not report on the variants with inexact solution of subproblems, with generating additional points via projections, or relaxing the master problems.

For the cases $n = 100, 250, 500, 1000$, and the options F_{const}, $F_{\text{N-J}}$, and $F_{\text{exact-J}}$, subproblems in variables q^i followed the decomposition pattern induced by the product $S_g = \prod_{i=0}^{5} S_{g_i}$, where $S_{g_0} = [0, U^0] \subset \mathbb{R}$ and $S_{g_i} = X^i \subset \mathbb{R}^{n/5}$, $i \geq 1$. With this decomposition, decision variables are precisely those of each player. For the larger configurations we used instead the product $S_g = \prod_{i=0}^{n/250} S_{g_i} = [0, U^0] \times \prod_{i=1}^{5} X^i$, with $S_{g_0} = [0, U^0] \subset \mathbb{R}$ and $S_{g_i} \subset \mathbb{R}^{250}$.

We use as stopping rules the following criteria. In PATH the stopping test employs the residual of the full problem based on the Fischer-Burmeister merit function [27, Chapter 1.5] with a default 10^{-6}. For the decomposition approaches the stopping criterion is

$$\frac{|\Delta_k|}{1 + |\Delta_1|} < 10^{-5},$$

where Δ_k is defined in (3.16). As discussed in Section 3.1.4, this is a natural stopping condition in the decomposition framework, as the access to the full problem, and thus to its residual, is not presumed.

All results are reported in Table 3.1, and interpreted in the two comparative Figures 3.1 and 3.2 below. For each configuration, we averaged over the 10 instances the results for each method. The table reports the average and maximal CPU times in seconds; the percentage of the total running time spent in the master and subproblem solution; the mean residual (the infinite-norm of the natural merit function [27, Chapter 1.5] for $\text{VI}(F(\cdot) + [h'(\cdot)]^\top \mu_M^k, S_g)$ at the master solution x_M^k); and the mean infinite-norm of the difference between x_M^k and x_S^{k+1} at termination. In particular, the latter distance and the residual are not a part of the decomposition stopping test (as they are not available within

the decomposition scheme anyway); these values were computed *a posteriori*, to confirm that an approximate solution of the problem was indeed obtained.

Regarding the running times, the main point we would like to stress is that for the configurations with $n \geq 5000$, applying PATH directly appears no longer possible even after relaxing the stopping tolerance from the default 10^{-6} to 10^{-2}. Of course, the exact threshold depends on the specific computer and implementation, yet there is always a threshold. On the other hand, some of the approximation options in the decomposition technique still succeed in solving the larger configurations in reasonable computational times.

The column reporting maximal CPU times in Table 3.1 gives an estimation on how the data dispersion affected each method. For $n = 2500$, for example, the percentage difference with respect to the mean CPU time was of 0.3, 7.5, 7.2, 15.8, 15.5, and 14.0%, respectively for PATH and the constant, Newton, Newton-Jacobi, and Jacobi approximations. As expected, the impact of varying data on the constant approximation is much smaller than for the other approximations, which incorporate more information. The situation is similar for $n = 10000$, where the percentages are 5.8, 19.1, and 7.2 for the only three approaches that could solve such large instances, respectively F_{const}, $F_{\text{N-J}}$, and F_{J}.

In order to see the benefit of decomposing, we took as reference the CPU time taken by a direct application of PATH and computed the ratio between the CPU times of each decomposition method and the reference one. Figure 3.1 shows the corresponding ratios. We should remark though that Figure 3.1 is intended merely to illustrate the dynamics of the comparison as the size grows and should not be taken literally. The reason is that, being a Newton-type method, when PATH works (i.e., for problems not too large), it provides highly accurate solutions. Generally, a comparable level of accuracy cannot be expected from the decomposition approach.

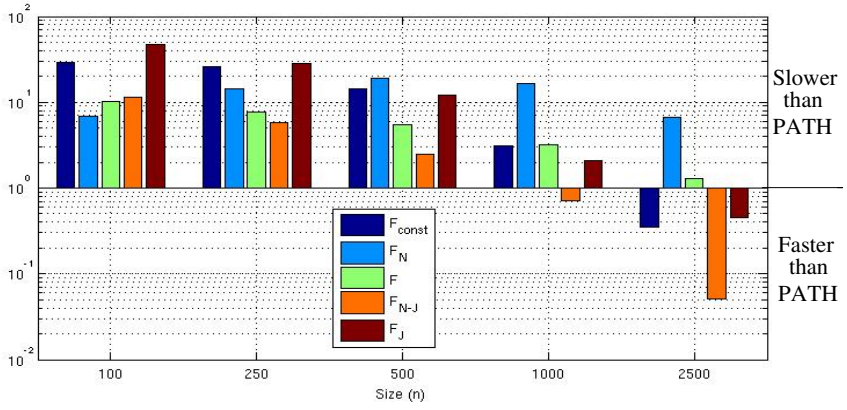

Figure 3.1: Time ratios (decomposition divided by PATH), configurations with $n \leq 2500$

Size and Model	CPU_{MEAN} (s)	CPU_{MAX} (s)	Master (% time)	SubPbm (% time)	Residual	$\|x_S - x_M\|_\infty$
$n = 100$						
PATH	0.162	0.235	-	-	-	-
F_{const}	5.199	8.081	92	8	0.001	3.006
F_{N}	1.170	1.331	72	27	0.016	0.048
F	1.973	2.276	46	53	0.015	0.051
$F_{\text{N-J}}$	2.010	4.923	89	10	0.083	0.069
F_{J}	8.785	10.027	17	83	0.080	0.068
$n = 250$						
PATH	0.303	0.326	-	-	-	-
F_{const}	9.002	12.617	94	6	0.007	6.114
F_{N}	4.397	4.819	23	77	0.016	0.043
F	2.530	2.691	41	59	0.015	0.045
$F_{\text{N-J}}$	1.915	2.274	80	20	0.070	0.125
F_{J}	9.561	11.107	16	84	0.075	0.123
$n = 500$						
PATH	1.413	1.713	-	-	-	-
F_{const}	19.413	30.356	95	5	0.008	8.071
F_{N}	29.079	31.620	6	93	0.017	0.035
F	7.229	8.359	25	75	0.017	0.036
$F_{\text{N-J}}$	3.313	4.067	64	36	0.087	0.099
F_{J}	14.285	20.401	15	85	0.086	0.100
$n = 1000$						
PATH	13.807	13.962	-	-	-	-
F_{const}	54.202	66.884	97	3	0.009	7.641
F_{N}	236.751	250.975	2	98	0.015	0.029
F	47.541	49.154	9	91	0.015	0.030
$F_{\text{N-J}}$	11.263	13.878	40	60	0.067	0.074
F_{J}	40.398	48.539	11	89	0.066	0.073
$n = 2500$						
PATH	693.439	695.626	-	-	-	-
F_{const}	255.843	275.134	98	2	0.025	8.918
F_{N}	4590.224	4919.049	0	100	0.022	0.037
F	923.633	1069.350	2	98	0.022	0.038
$F_{\text{N-J}}$	37.028	42.773	53	47	0.058	0.101
F_{J}	335.984	383.007	6	94	0.057	0.085
$n = 5000$						
PATH	-	-	-	-	-	-
F_{const}	1043.489	1257.464	99	1	0.038	9.535
$F_{\text{N-J}}$	114.774	123.597	83	17	0.032	0.056
F_{J}	2239.637	2408.383	4	96	0.031	0.055
$n = 10000$						
PATH	-	-	-	-	-	-
F_{const}	4204.447	4450.205	99	1	0.066	9.857
$F_{\text{N-J}}$	483.478	575.581	72	28	0.029	0.046
F_{J}	13891.376	14891.010	2	98	0.029	0.045

Table 3.1: Detailed list of all results

The ordinate in Figure 3.1 uses a logarithmic scale, for convenience. In the figure, when for a given method the value represented by a bar lies above the 0-ordinate (corresponding to 10^0 in the logarithmic scale), the decomposition method took longer than the direct approach with PATH. By contrast, when the bar is below the ordinate 10^0, the decomposition method was faster than PATH. The plot in Figure 3.1 shows a natural behavior. A direct application of PATH is very efficient for the smaller to medium sized problems. But as the size grows, decomposition becomes more and more competitive. For $n = 1000$, the $F_{\text{N-J}}$ decomposition already outperforms the direct approach. And for $n = 2500$, three of the decomposition approaches become faster than PATH, with the $F_{\text{N-J}}$ approximation being the best one, contrasting with the F_{N} approximation that has the worst performance because of the indecomposable structure of the subproblems involved in the algorithm.

When the percentage distribution of time between master and subproblem solution in Table 3.1 does not add up to 100%, this is due to some time spent in intermediate tasks, such as communicating with PATH mex-interface. We observe that for the larger configurations in general the best approach (Newton-Jacobi) spends less time in solving subproblems than in dealing with the master problems. In view of our comments in Remark 3.3.1, we conjecture that if we were to consider difficult sets X^i, the percentage distribution of time would result in higher figures for the subproblems. Since the Jacobi-like approximations are amenable to parallelization (thus making the subproblem solution quicker), for such decompositions the more CPU time is spent in solving subproblems in our current serial implementation, the faster would be the overall procedure in parallel implementation. Moreover, it is also likely that more intricate sets X^i would make decomposition preferable over a direct solution with PATH even for the smaller instances (always keeping in mind that this is a problem dependent issue).

Regarding solution quality, the two last columns in Table 3.1 report, respectively, the value of the *a posteriori* computed residual and of the gap between x_S and x_M. We observe that while the constant approximation gives systematically the most distant x_S and x_M's, for smaller instances this approximation also has the lowest residual. This tendency starts changing at $n = 2500$ and for $n \geq 5000$ the Newton-Jacobi and Jacobi approximations become more accurate, and practically equally so. Since the approximation $F_{\text{N-J}}$ is the fastest one, it appears as the best option for large configurations, both in terms of speed and accuracy.

We finish our analysis by considering scalability issues. Recall once again that PATH applied to the full problem stalls for $n \geq 5000$, while some decomposition approaches still work in reasonable time. Figure 3.2 compares the performance of F_{const}, $F_{\text{N-J}}$, F_{J}, which are the options able to handle the larger configurations. The plot shows the corresponding mean CPU times in minutes for each configuration, ranging from $n = 100$ to $n = 10000$. We observe that $F_{\text{N-J}}$ shows the best scalability with respect to the problem size, suggesting once more this is the best option for larger models of the type considered here.

Finally, we remark that the Dantzig-Wolfe decomposition algorithm is useful

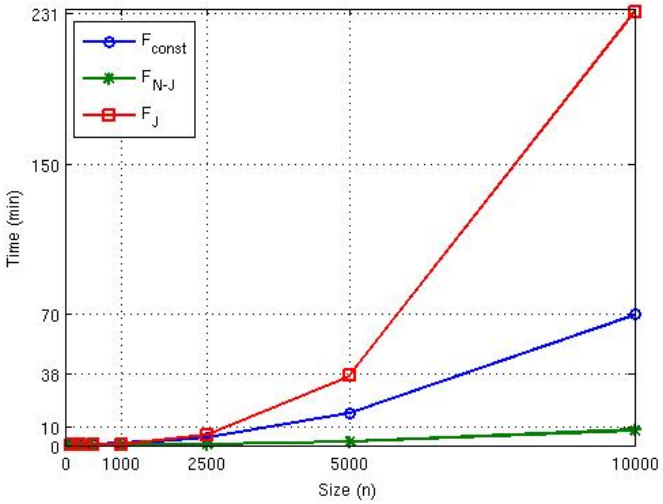

Figure 3.2: Scalability of the best decomposition options

for solving VI problems whose feasible set $\prod_i X^i \cap \mathcal{S}$ has the structure above, which is typical for GNEPs. By using this algorithm we reduce the computational effort from dealing directly with $\prod_i X^i \cap \mathcal{S}$ to dealing with each X^i separately. The resulting improvement may, in a number of cases, be very significant. However, in some applications, like those described in Subsection 2.4, Chapter 5, and Chapter 6, the setting is slightly different. Namely, the set X^i may consist of points (z, q) that satisfy some constraints like $Zz + Qq \geq b$, with q having large dimension, but with Q having a block decomposable structure, similar to the one discussed in Section 2.2 when describing the Benders algorithm for linear programming. The next chapter is devoted to VI problems whose feasible sets have a form amenable to Benders decomposition.

Chapter 4

Benders Decomposition for Variational Inequalities

In this chapter we follow the strategy described in Section 2.2 in the case of linear programming for deriving, via duality, Benders method from the Dantzig-Wolfe method in the case of VIs of a certain structure. In our presentation, we follow the general lines of [51].

Consider the problem

$$\mathrm{VI}(F_P, S_P) \tag{4.1}$$

with a certain special structure; specifically, $F_P : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ has the form

$$F_P(x, y) = F(x) \times G(y),$$

with $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, $G : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ and

$$S_P = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : Ax + By \leq d\},$$

where A and B are matrices of appropriate dimensions.

We further assume that after fixing the value of the variable y, the problem $\mathrm{VI}(F, S_P(y))$, with $S_P(y) = \{x \in \mathbb{R}^n : Ax + By \leq d\}$, is much easier to solve than (4.1). As in the linear programming case of Section 2.2, this happens when A has block decomposable structure, something in fact rather common in applications. For convergence analysis, we shall assume that F and G are outer semicontinuous. The latter holds, in particular, if they are maximal monotone (possibly multi-valued)[7, Proposition 4.2.1]. For solvability of iterative problems involved in the construction, one of F and G should also either be surjective or its inverse has to be surjective. The latter is a technical assumption needed to ensure the maximal monotonicity of the mapping of the dual problem ((4.3) further below). When F or G has domain or image bounded, then this technical assumption is automatic [7, Corollary 4.5.1].

The only other Benders type method for VIs that we are aware of is the one proposed in [30]. The differences between our development and [30], and our

contributions, are summarized as follows. First, we consider VI with a multi-valued mapping, whereas in [30] the mapping is single-valued and, even more importantly, it has a rather specific form, like $(F(x), c, d)$ with F continuous and invertible and c, d constant vectors. Clearly, our setting is much more general and covers far more applications. Moreover, as it will be seen below this generalization does not complicate too much the iterations of the algorithm, i.e., the subproblems will still be computationally tractable. In fact, the iterative subproblems in our method related to the multi-valued part, that replaces the constant part in the previous work, are independent of the variable x and can be solved relatively easily. Second, the existence of solutions of primal subproblems in the previous work was an assumption. Here, we use regularization to ensure solvability, and thus dispense with such assumptions. Finally, we allow approximations for the dual subproblem VI mapping, as in the Dantzig-Wolfe algorithm of Chapter 3, while previous work requires the use of exact information. Apart from general importance of using approximations in many real-world applications, this is of special significance here because it allows us to express the corresponding primal subproblem as a simple minimization problem (instead of a VI) if an appropriate type of approximation is chosen.

4.1 The dual problem

As in Section 2.2 for linear programming, the construction of Benders method for VI is via applying the Dantzig-Wolfe technique to an appropriately defined dual. In fact, since (4.1) has a especial structure, it is possible to define the dual problem by rearranging the KKT conditions. This dual problem fits well on the Dantzig-Wolfe decomposition scheme. It is worth to mention that so far there were proposed a number of approaches for binding a dual problem to a given VI considering various degree of abstractness (e. g. [56, 1, 40, 2]) that however do not fit very well to our the decomposition scheme. We thus start with associating to (4.1) the following dual problem:

$$\mathrm{VI}(F_D, S_D), \tag{4.2}$$

where $F_D : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ is given by

$$F_D(w, \zeta, \mu) = F^{-1}(w) \times G^{-1}(\zeta) \times \{d\} \tag{4.3}$$

and

$$S_D = \left\{ \begin{array}{r} w + A^\top \mu = 0 \\ (w, \zeta, \mu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p : \ \zeta + B^\top \mu = 0 \\ \mu \geq 0 \end{array} \right\}.$$

Again, if A has block-decomposable structure, the constraint $w + A^\top \mu = 0$ also has this property, which means that the set S_D would be easier to deal with if the constraint $\zeta + B^\top \mu = 0$ were to be removed. We then immediately recognize that (4.2) is amenable to the Dantzig-Wolfe decomposition developed in Chapter 3.

The following proposition shows that (4.1) and (4.2) are equivalent in a certain sense.

Proposition 4.1.1. *For the data defined above, the following holds.*

1. *If $(\bar{x}, \bar{y})$ together with $\bar{w} \in F(\bar{x})$ and $\bar{\zeta} \in G(\bar{y})$ solve the primal problem (4.1) with multiplier $\bar{\mu}$ then $(\bar{w}, \bar{\zeta}, \bar{\mu})$ together with $\bar{x} \in F^{-1}(\bar{w})$ and $\bar{y} \in G^{-1}(\bar{\zeta})$ solve the dual problem (4.2) with multipliers $(-\bar{x}, -\bar{y}, -A\bar{x} - B\bar{y} + d)$.*

2. *If $(\bar{w}, \bar{\zeta}, \bar{\mu})$ together with $\bar{x} \in F^{-1}(\bar{w})$ and $\bar{y} \in G^{-1}(\bar{\zeta})$ solve the dual problem (4.2) with multipliers $(-\bar{\alpha}, -\bar{\beta}, -\bar{\gamma})$, then*

   (a) *$\bar{\alpha} = \bar{x}$, $\bar{\beta} = \bar{y}$ and $\bar{\gamma} = A\bar{x} + B\bar{y} - d$,*

   (b) *$(\bar{x}, \bar{y})$ together with $\bar{w} \in F(\bar{x})$ and $\bar{\zeta} \in G(\bar{y})$ solve the primal problem (4.1) with multiplier $\bar{\mu}$.*

Proof. Since the constraints defining S_P are linear, we have that the Karush-Kuhn-Tucker (KKT) conditions hold:

$$\bar{w} + A^\top \bar{\mu} = 0, \tag{4.4a}$$

$$\bar{\zeta} + B^\top \bar{\mu} = 0, \tag{4.4b}$$

$$0 \leq \bar{\mu} \perp A\bar{x} + B\bar{y} - d \leq 0. \tag{4.4c}$$

This shows that $(\bar{w}, \bar{\zeta}, \bar{\mu}) \in S_D$.

On the other hand, the following inclusion is immediate:

$$0 \in \begin{bmatrix} F^{-1}(\bar{w}) \\ G^{-1}(\bar{\zeta}) \\ d \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ A \end{bmatrix} (-\bar{x}) + \begin{bmatrix} 0 \\ I \\ B \end{bmatrix} (-\bar{y}) + \begin{bmatrix} 0 \\ 0 \\ -I \end{bmatrix} (d - A\bar{x} - B\bar{y}). \tag{4.5}$$

The latter, together with (4.4c), proves the first item.

The proof of the second item is similar. □

Next, we describe the iteration subproblems that the Dantzig-Wolfe algorithm would have been solving if applied to the dual problem, and their primal counterparts that give the Benders' approach.

4.2 The iteration subproblem

At iteration k, given $(w_M^k, \zeta_M^k, \mu_M^k) \in S_D$ with $x_M^k \in F^{-1}(w_M^k)$ and $y_M^k \in G^{-1}(\zeta_M^k)$ and a Lagrange multiplier estimate $(-y_M^k - \theta_M^k)$ associated to the constraint $\zeta + B^\top \mu = 0$. The Lagrange multiplier estimate is expressed as sum of two terms by technical reasons concerning its relationship with the decision variables of (4.25), and so the way we write that problem.

The Dantzig-Wolfe subproblem at the kth iteration consists in solving

$$\mathrm{VI}(\hat{F}_D^k, S_{D_S}), \tag{4.6}$$

where the feasible set S_{D_S} is defined by

$$S_{D_S} = \left\{ (w, \zeta, \mu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p : \begin{array}{r} w + A^\top \mu = 0 \\ \mu \geq 0 \end{array} \right\}, \tag{4.7}$$

and the VI mapping $\hat{F}_D^k$ is defined using some approximations F_k^{-1} and G_k^{-1} of F^{-1} and G^{-1}, and some positive (semi)definite matrices P_k, Q_k and R_k, by

$$\hat{F}_D^k(w, \zeta, \mu) = \begin{bmatrix} F_k^{-1}(w) \\ G_k^{-1}(\zeta) \\ d \end{bmatrix} + \begin{bmatrix} 0 \\ I \\ B \end{bmatrix} (-y_M^k - \theta_M^k) + \begin{bmatrix} Q_k(w - w_M^k) \\ R_k(\zeta - \zeta_M^k) \\ P_k(\mu - \mu_M^k) \end{bmatrix}. \tag{4.8}$$

The approximating functions F^k and G^k are chosen according to the requirements of the Dantzig-Wolfe scheme in Section 3.1.1. In particular,

$$x_M^k \in F_k^{-1}(w_M^k) \subset F^{-1}(w_M^k),$$

and

$$y_M^k \in G_k^{-1}(\zeta_M^k) \subset G^{-1}(\zeta_M^k).$$

Next, note the following two important structural features. In the feasible set (4.7) the variable ζ is unconstrained, and in the VI mapping (4.8) the entry corresponding to this variable is independent of the other variables. It then follows that the subproblem (4.6) equivalently splits into the following two:

1. Find $\zeta \in \mathbb{R}^m$ such that

$$0 \in G_k^{-1}(\zeta) - y_M^k - \theta_M^k + R_k(\zeta - \zeta_M^k). \tag{4.9}$$

2. Solve

$$\mathrm{VI}(\hat{F}_{D2}^k, S_{D_{S2}}), \tag{4.10}$$

where

$$\hat{F}_{D2}^k(w, \mu) = \begin{bmatrix} F_k^{-1}(w) + Q_k(w - w_M^k) \\ d - B(y_M^k + \theta_M^k) + P_k(\mu - \mu_M^k) \end{bmatrix}, \tag{4.11}$$

and

$$S_{D_{S2}} = \left\{ (w, \mu) \in \mathbb{R}^n \times \mathbb{R}^p : \begin{array}{r} w + A^\top \mu = 0 \\ \mu \geq 0 \end{array} \right\}. \tag{4.12}$$

The two problems above have the disadvantage in that they are defined in terms of the inverse functions that in practice could be difficult to deal with. In fact, in most applications those inverses would not be known explicitly. For this reason, we solve (4.9) and (4.10) via their dual problems. The following propositions show those dual relations.

Proposition 4.2.1. *If y_S^{k+1} together with $\zeta_S^{k+1} \in G_k(y_S^{k+1})$ solve the problem*

$$0 \in y - y_M^k - \theta_M^k + R_k(G_k(y) - \zeta_M^k) \tag{4.13}$$

then ζ_S^{k+1} solves (4.9) with $y_S^{k+1} \in G_k^{-1}(\zeta_S^{k+1})$.

Also, the existence of solutions of (4.9) implies existence of solutions of (4.13). In particular, solutions exist if G_k is maximal monotone and R_k is positive definite.

Proof. The first assertion is obtained by direct inspection. The existence of solutions is by [65, Theorem 5]. □

Proposition 4.2.2. *Let (x_S^{k+1}, u_S^{k+1}) be any solution of*

$$VI(\hat{F}_P^k, S_P(y_M^k, \theta_M^k)), \tag{4.14}$$

where

$$\hat{F}_P^k(x, \mu) = \begin{pmatrix} F_k(x) \\ (AQ_kA^\top + P_k)\mu \end{pmatrix}, \tag{4.15}$$

and

$$S_P(y_M^k, \theta_M^k) = \{(x, \mu) : Ax + B(y_M^k + \theta_M^k) \leq d + (AQ_kA^\top + P_k)(\mu - \mu_M^k)\}. \tag{4.16}$$

Let $w_S^{k+1} \in F_k(x_S^{k+1})$ and multiplier μ_S^{k+1} be the quantities associated to (x_S^{k+1}, u_S^{k+1}) as a solution of (4.14) (i.e., the quantities that together with (x_S^{k+1}, u_S^{k+1}) verify KKT conditions for (4.14)).

Then the following holds.

1. $(AQ_kA^\top + P_k)(\mu_S^{k+1} - u_S^{k+1}) = 0$.
2. (x_S^{k+1}, μ_S^{k+1}) *solves VI($\hat{F}_P^k, S_P(y_M^k, \theta_M^k)$) with $w_S^{k+1} \in F^k(x_S^{k+1})$ and multiplier μ_S^{k+1}.*
3. (w_S^{k+1}, μ_S^{k+1}) *solves (4.10) with $x_S^{k+1} \in F_k^{-1}(w_S^{k+1})$ and multipliers $(-x_S^{k+1} - Q_k(w_S^{k+1} - w_M^k), d + (AQ_kA^\top + P_k)(\mu_S^{k+1} - \mu_M^k) - Ax_S^{k+1} - B(y_M^k + \theta_M^k))$.*

Also, the existence of solutions of (4.10) implies the existence of solutions of (4.14). In particular, solutions exist if F_k is maximal monotone with $\{-A^\top\mu : \mu \geq 0\} \cap int(F_k(\mathbb{R}^n)) \neq \emptyset$ and P_k, Q_k are positive definite.

Proof. Writing KKT conditions corresponding to (4.14), which hold by the linearity of constraints in this problem, we have that

$$0 = w_S^{k+1} + A^\top\mu_S^{k+1}, \tag{4.17a}$$

$$0 = (AQ_kA^\top + P_k)u_S^{k+1} - (AQ_kA^\top + P_k)^\top\mu_S^{k+1}, \tag{4.17b}$$

$$0 \leq \mu_S^{k+1} \perp Ax_S^{k+1} + B(y_M^k + \theta_M^k) - d - (AQ_kA^\top + P_k)(u_S^{k+1} - \mu_M^k) \leq 0. \tag{4.17c}$$

Then (4.17b) implies the first item of the proposition, which together with the system above shows also the second item. The third item is obtained in a way similar to Proposition 4.1.1; we omit the details.

The existence assertion follows by applying [65, Theorem 5] to (4.10). □

Clearly, this general approach can be used in particular with the exact information, i.e., $G_k = G$ and $F_k = F$. However, using suitable approximations, the primal subproblems could be much easier to solve. For example:

1. If we use the constant approximation $(G_k^{\text{const}})^{-1}(\zeta) = y_M^k$, the solution ζ_S^{k+1} of (4.9) is given *explicitly* by $\zeta_M^k + R_k^{-1}\theta_M^k$.

   Also, if we use $(F_k^{\text{const}})^{-1}(w) = x_M^k$, (4.10) is equivalent to

   $$\text{VI}(d - Ax_M^k - B(y_M^k + \theta_M^k) + (AQ_kA^\top + P_k)(\mu - \mu_M^k), \mathbb{R}_+^p), \quad (4.18)$$

   which is in fact a *strongly convex* (if P_k and Q_k are symmetric positive definite) *quadratic programing problem* with simple bounds, with a wealth of powerful software to apply. In this case, $w_S^{k+1} = -A^\top \mu_S^{k+1}$.

2. If we take $F_k(x) = w_M^k + H_{F_k}(x - x_M^k)$ for some H_{F_k} nonsingular, then $F_k^{-1}(w) = x_M^k + H_{F_k}^{-1}(w - w_M^k)$ and (4.10) is equivalent to

   $$\text{VI}(d - Ax_M^k - B(y_M^k + \theta_M^k) + (A(H_{F_k}^{-1} + Q_k)A^\top + P_k)(\mu - \mu_M^k), \mathbb{R}_+^p), \quad (4.19)$$

   again a simple quadratic program as the one just above, if the matrices are chosen accordingly. Also, $w_S^{k+1} = -A^\top \mu_S^{k+1}$.

It is worth to note that we can always choose the matrices Q_k, P_k and H_{F_k} in a convenient way. For example, as diagonal matrices or having a convenient block-decomposable structure. This is especially important when the matrix A has also block-decomposable structure, which we would like to exploit. In this case the matrices $AQ_kA^\top + P_k$ and $A(Q_k + H_{F_k}^{-1})A^\top + P_k$ also would have decomposable structures. This implies, for example, that the set (4.16) also splits according to the given pattern, which makes solving (4.14) much easier. Of course, in (4.15) the mapping $F_k(x)$ may not be decomposable for some choices (the entry corresponding to μ is clearly decomposable). However, even in that case we can still use special methods in order to take advantage of the structure of the feasible set (e.g., the parallel variable distribution coupled with sequential quadratic programming [71], if we are in the optimization setting). And in any case, with the specific choices of F_k that lead to the problems 4.18 and 4.19, the latter are always decomposable if A is (and the parameter matrices are made to follow the pattern).

4.3 The iteration master problem

In the last section, we showed how to solve, using primal information, the subproblem required by the Dantzig-Wolfe algorithm applied to the dual problem (4.2) of (4.1). Now, we deal with the next step of Dantzig-Wolfe algorithm, the Master problem. At iteration $k \geq 0$, given $X^{k+1} = \{(w_S^i, \zeta_S^i, \mu_S^i)\}_{i=0}^{k+1}$, the kth master problem consists in solving $\text{VI}(F_D, S_{D_M})$, where

$$S_{D_M} = \{(w, \zeta, \mu) \in \operatorname{conv} X^{k+1} : 0 = \zeta + B^\top \mu\}. \quad (4.20)$$

Under the technical assumptions stated in the beginning of this chapter, F_D is maximal monotone, and since S_{D_M} is compact, the master problem (4.20) is solvable [65, Theorem 5].

Let us remind that the points $(w_S^i, \zeta_S^i, \mu_S^i)$, for $i \geq 1$, are computed by solving the subproblems described in Section 4.2 above, whereas $(w_S^0, \zeta_S^0, \mu_S^0)$ is a feasible point of (4.2) chosen at the beginning of the algorithm (as prescribed by the Dantzig-Wolfe framework in Chapter 3).

Using the matrices

$$W_{k+1} = [w_S^0 | w_S^1 | \cdots | w_S^{k+1}],$$

$$Z_{k+1} = [\zeta_S^0 | \zeta_S^1 | \cdots | \zeta_S^{k+1}],$$

$$M_{k+1} = [\mu_S^0 | \mu_S^1 | \cdots | \mu_S^{k+1}],$$

problem (4.20) can be reformulated as

$$\mathrm{VI}(F_{D_\triangle}, S_{D_\triangle}), \tag{4.21}$$

where

$$F_{D_\triangle}(w, \zeta, \alpha) = F^{-1}(w) \times G^{-1}(\zeta) \times \{M_{k+1}^\top d\},$$

and $S_{D_\triangle}$ is the set of points $(w, \zeta, \alpha) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{k+1}$ that satisfy the following constraints:

$$w - W_{k+1}\alpha = 0, \tag{4.22a}$$

$$\zeta + B^\top M_{k+1}\alpha = 0, \tag{4.22b}$$

$$\zeta - Z_{k+1}\alpha = 0, \tag{4.22c}$$

$$1 - \mathbf{1}^\top \alpha = 0, \tag{4.22d}$$

$$\alpha \geq 0, \tag{4.22e}$$

where $\mathbf{1}$ is the vector of ones of the appropriate dimension.

It is worth to note that since all points $(w_S^i, \zeta_S^i, \mu_S^i)$ are feasible for (4.6), we have that $W_{k+1} = -AM_{k+1}$.

For same reasons as above (i.e., the involvement of the inverse functions), we shall avoid solving directly (4.21) and solve instead its dual problem, described next.

Proposition 4.3.1. *Defining*

$$F_P(x, y, \beta, \theta) = F(x) \times G(y) \times \{-1\} \times \{0\}, \tag{4.23}$$

and

$$S_{P_{k+1}} = \{(x, y, \beta, \theta) : M_{k+1}^\top [Ax + By - d] \leq -\mathbf{1}\beta - [Z_{k+1}^\top + M_{k+1}^\top B]\theta\}, \tag{4.24}$$

we have that if $(x_M^{k+1}, y_M^{k+1}, \beta_M^{k+1}, \theta_M^{k+1})$ *solves*

$$VI(F_P, S_{P_{k+1}}), \tag{4.25}$$

with some $w_M^{k+1} \in F(x_M^{k+1})$, $\zeta_M^{k+1} \in G(y_M^{k+1})$ *and some Lagrange multiplier* α_M^{k+1}, *then* $(w_M^{k+1}, \zeta_M^{k+1}, \alpha_M^{k+1})$ *solves VI*$(F_{D_\triangle}, S_{D_\triangle})$ *with* $x_M^{k+1} \in F^{-1}(w_M^{k+1})$, $y_M^{k+1} \in G^{-1}(\zeta_M^{k+1})$ *and Lagrange multipliers* $-x_M^{k+1}$, $-y_M^{k+1} - \theta_M^{k+1}$, θ_M^{k+1}, β_M^{k+1} *and* $M_{k+1}^\top[d - Ax_M^{k+1} - By_M^{k+1}] - [Z_{k+1}^\top + M_{k+1}^\top B]\theta_M^{k+1} - \mathbf{1}\beta_M^{k+1}$.

Also, VI$(F_P, S_{P_{k+1}})$ *has solutions if, and only if, so does VI*$(F_{D_\triangle}, S_{D_\triangle})$.

Proof. The proof is analogous to that of Proposition 4.1.1. □

4.4 Convergence analysis

We now state formally the algorithm and then analyze its convergence properties. The Benders decomposition for VI (4.1) follows the following pattern.

Algorithm 4.4.1. (Benders Decomposition)

1. Choose $\mu_S^0 \geq 0$ and $\zeta_S^0 \in \mathbb{R}^m$. Set $w_S^0 = -A^\top \mu_S^0$, $\mu_M^0 = \mu_S^0$, $\zeta_M^0 = \zeta_S^0$ and $w_M^0 = w_S^0$. Choose $x_M^0 \in F(w_M^0)$, $y_M^0 \in G(\zeta_M^0)$ and $\theta_M^0 \in \mathbb{R}^m$. Set $k = 0$.
2. **The Subproblem:** Choose approximations $F^k : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $G^k : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ of $F(\cdot)$ and $G(\cdot)$, and positive (semi)definite matrices $Q_k \in \mathbb{R}^{n\times n}$, $R_k \in \mathbb{R}^{m\times m}$ and $P_k \in \mathbb{R}^{p\times p}$. Find the primal-dual points (x_S^{k+1}, y_S^{k+1}) and $(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})$ by solving the problems (4.9) or (4.13), and (4.14) or (4.18) or (4.19), according to the approximation functions chosen.
3. **The Master Problem:** Find $(x_M^{k+1}, y_M^{k+1}, \theta_M^{k+1})$ and $(w_M^{k+1}, \zeta_M^{k+1}, \mu_M^{k+1})$ by solving (4.25), with a Lagrange multiplier μ_M^{k+1} associated to the constraint in (4.24).
4. Set $k := k + 1$ and go to Step 2.

We proceed to analyze convergence properties of Algorithm 4.4.1. We start with the associated convergence gap quantity.

The *gap of convergence* Δ_k, defined by (3.16) for the Dantzig-Wolfe method, in the present setting has the form

$$\begin{aligned}\Delta_k &= \Big\langle (x_M^k, -\theta_M^k, d - By_M^k - B\theta_M^k), (w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k) \Big\rangle \\ &= \langle x_M^k, w_S^{k+1} - w_M^k \rangle + \langle -\theta_M^k, \zeta_S^{k+1} - \zeta_M^k \rangle + \langle d - By_M^k - B\theta_M^k, \mu_S^{k+1} - \mu_M^k \rangle, \end{aligned} \tag{4.26}$$

where we have used the fact that

$$(x_M^k, y_M^k, d) + (0, -y_M^k - \theta_M^k, -By_M^k - B\theta_M^k) \in \hat{F}_D^k(w_M^k, \zeta_M^k, \mu_M^k),$$

which follows from (4.8).

Proposition 4.4.2. *If for each* k *the function* $\hat{F}_D^k$ *is chosen* c_k*-strongly monotone, then*

$$c_k \|(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\|^2 + \Delta_k \leq 0. \tag{4.27}$$

In particular, $\Delta_k \leq 0$ *for every* k.

Proof. Since $(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) \in S_{D_S}$ and $(w_M^k, \zeta_M^k, \mu_M^k) \in S_D$, we have that

$$\begin{aligned}\Delta_k &= \langle Ax_M^k, \mu_M^k - \mu_S^{k+1}\rangle + \langle -\theta_M^k, \zeta_S^{k+1} - \zeta_M^k\rangle + \langle d - By_M^k - B\theta_M^k, \mu_S^{k+1} - \mu_M^k\rangle \\ &= \langle Ax_M^k + By_M^k - d, \mu_M^k - \mu_S^{k+1}\rangle + \langle -\theta_M^k, \zeta_S^{k+1} - \zeta_M^k\rangle - \langle B\theta_M^k, \mu_S^{k+1} - \mu_M^k\rangle \\ &= \langle Ax_M^k + By_M^k - d, \mu_M^k - \mu_S^{k+1}\rangle - \langle \theta_M^k, \zeta_S^{k+1} + B^\top \mu_S^{k+1}\rangle.\end{aligned}$$

Using now c_k-strong monotonicity of $\hat{F}_D^k$, since

$$(x_M^k, -\theta_M^k, d - By_M^k - B\theta_M^k) \in \hat{F}_D^k(w_M^k, \zeta_M^k, \mu_M^k)$$

and since $(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})$ solves $\mathrm{VI}(\hat{F}_D^k, S_{D_S})$ with some

$$z_S^{k+1} \in \hat{F}_D^k(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}),$$

we have that

$$\begin{aligned}&c_k\|(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\|^2 \leq \\ &\langle z_S^{k+1} - (x_M^k, -\theta_M^k, d - By_M^k - B\theta_M^k), (w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\rangle.\end{aligned} \tag{4.28}$$

We then further obtain

$$\begin{aligned}& c_k\|(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\|^2 \\ \leq\ & \langle z_S^{k+1}, (w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\rangle - \Delta_k.\end{aligned}$$

Using the latter relation and $(w_M^k, \zeta_M^k, \mu_M^k) \in S_{D_S}$, (4.27) follows. □

We are now in position to state convergence properties of Benders decomposition of VIs with the given structure. The theorem below assumes that F and G are outer semicontinuous, which is the only property used in the proof, if the existence of solutions/iterations is a given. Outer semicontinuity is in fact automatic from the initial assumptions stated in the beginning of this chapter, which also guarantee the existence of solutions of all the problems along the iterations. We also assume the equicontinuity of the approximating families $\{F_k^{-1}\}$ and $\{G_k^{-1}\}$. Again, there is a number of ways to ensure the latter. Constant approximations is one option. If first-order (Newtonian) approximations are used for single-valued smooth data, choosing bounded $\{H_{F_k}^{-1}\}$ does the job. Finally, if the exact information F and G is employed (no approximations) in the single-valued case, the continuity of those functions is sufficient. Note also that most of the proof is similar to the proof of Theorem 3.2.4. We repeat the arguments because the dual problem does not satisfy all hypotheses required in the analysis of the Dantzig–Wolfe algorithm made in Chapter 3; specifically, those related to the domain of the dual problem operator.

Theorem 4.4.3. *Suppose that F and G are outer semicontinuous (which holds, in particular, if they are maximal monotone). For the iterative sequences generated by Algorithm 4.4.1, the following holds.*

1. *If the sequences* $\{\mu_S^{k+1}\}$, $\{y_S^{k+1}\}$, $\{y_M^k\}$ *and* $\{\theta_M^k\}$ *are bounded and if the family of matrices* $\{R_k\}$ *is uniformly positive definite, then the sequences* $\{(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})\}$ *and* $\{(w_M^k, \zeta_M^k, \mu_M^k)\}$ *are bounded.*

2. *If the sequences* $\{x_M^k\}$, $\{\mu_S^{k+1}\}$, $\{y_S^{k+1}\}$, $\{y_M^k\}$ *and* $\{\theta_M^k\}$ *are bounded, the family of matrices* $\{R_k\}$ *is uniformly positive definite, and the approximations* $\{\hat{F}_D^k\}$ *are chosen monotone, then*

$$\lim_{k\to\infty} \Delta_k = 0.$$

In particular, if the elements of $\{\hat{F}_D^k\}$ *are chosen uniformly strongly monotone, then*

$$\lim_{k\to\infty} \|(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\| = 0. \tag{4.29}$$

3. *Suppose that the sequences* $\{P_k\}$,$\{Q_k\}$,$\{R_k\}$, $\{(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})\}$ *and* $\{(x_S^{k+1}, y_S^{k+1})\}$ *are bounded and that* (4.29) *holds. Then, if the approximations* $\{F_k^{-1}\}$ *and* $\{G_k^{-1}\}$ *are equicontinuous on compact sets, every cluster point of* $\{(x_S^{k+1}, y_S^{k+1})\}$ *is a solution of VI* (4.1).

Proof. 1. Let $\{\mu_S^{k+1}\}$ be bounded. Since $\mu_M^k \in \operatorname{conv}(\{\mu_S^j\})$, it follows that the sequence $\{\mu_M^k\}$ is bounded. Furthermore, as $w_S^{k+1} = -A^\top \mu_S^{k+1}$, $w_M^k = -A^\top \mu_M^k$ and $\zeta_M^k = -B^\top \mu_M^k$, it follows that the sequences $\{w_S^{k+1}\}$ and $\{(w_M^k, \zeta_M^k, \mu_M^k)\}$ are bounded. Now boundedness of $\{\zeta_S^{k+1}\}$ follows from

$$y_S^{k+1} - y_M^k - \theta_M^k + R_k(\zeta_S^{k+1} - \zeta_M^k) = 0$$

and the uniform positive definite property of the family $\{R_k\}$.

2. By the first item, we have that the sequences $\{(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})\}$ and $\{(w_M^k, \zeta_M^k, \mu_M^k)\}$ are bounded.

   Using (4.27) with $c_k = 0$ (i.e., monotonicity instead of strong monotonicity), we have that $\Delta_k \le 0$. Hence,

$$\bar{\Delta} = \liminf_{k\to\infty} \Delta_k \le \limsup_{k\to\infty} \Delta_k \le 0.$$

   We take a subsequence $\{\Delta_{k_j}\}$ such that $\lim_{j\to\infty} \Delta_{k_j} = \bar{\Delta}$. Without loss of generality, we can assume convergence of the corresponding subsequences: $\{(w_M^{k_j}, \zeta_M^{k_j}, \mu_M^{k_j})\} \to (\bar{w}, \bar{\zeta}, \bar{\mu})$, $\{(x_M^{k_j}, y_M^{k_j}, \theta_M^{k_j})\} \to (\bar{x}, \bar{y}, \bar{\theta})$ and $\{(w_S^{k_j+1}, \zeta_S^{k_j+1}, \mu_S^{k_j+1})\} \to (\hat{w}, \hat{\zeta}, \hat{\mu})$. Then from (4.26), we have that

$$\lim_{j\to\infty} \Delta_{k_j} = \bar{\Delta} = \Big\langle (\bar{x}, -\bar{\theta}, d - B(\bar{y} + \bar{\theta})), (\hat{w}, \hat{\zeta}, \hat{\mu}) - (\bar{w}, \bar{\zeta}, \bar{\mu}) \Big\rangle.$$

   As in the second item of Proposition 3.2.1, we consider the problem that results after relaxing the constraint $\zeta + B^\top \mu = 0$ using the multiplier

$(-y_M^{k_j} - \theta_M^{k_j})$ in VI(F_D, S_{D_M}) (4.20). Fix any index j. Then for every $i > j$ we have that $(w_S^{k_j+1}, \zeta_S^{k_j+1}, \mu_S^{k_j+1})$ is a feasible point and $(w_M^{k_i}, \zeta_M^{k_i}, \mu_M^{k_i})$ is a solution in (4.20). Therefore,

$$\left\langle (x_M^{k_i}, -\theta_M^{k_i}, d - B(y_M^{k_i} + \theta_M^{k_i})), (w_S^{k_j+1}, \zeta_S^{k_j+1}, \mu_S^{k_j+1}) - (w_M^{k_i}, \zeta_M^{k_i}, \mu_M^{k_i}) \right\rangle \geq 0,$$

and after passing on to limit as $i \to \infty$, we get

$$\left\langle (\bar{x}, -\bar{\theta}, d - B(\bar{y} + \bar{\theta})), (w_S^{k_j+1}, \zeta_S^{k_j+1}, \mu_S^{k_j+1}) - (\bar{w}, \bar{\zeta}, \bar{\mu}) \right\rangle \geq 0.$$

Passing onto the limit again, now as $j \to \infty$, we obtain that

$$\bar{\Delta} = \left\langle (\bar{x}, -\bar{\theta}, d - B(\bar{y} + \bar{\theta})), (\hat{w}, \hat{\zeta}, \hat{\mu}) - (\bar{w}, \bar{\zeta}, \bar{\mu}) \right\rangle \geq 0,$$

which shows that $\lim_{k\to\infty} \Delta_k = 0$. Finally, (4.27) implies the last assertion of this item.

3. Suppose that $(\bar{x}, \bar{y})$ is an accumulation point of $\{(x_S^{k+1}, y_S^{k+1})\}$ and that the subsequence $\{(x_S^{k_j+1}, y_S^{k_j+1})\}$ converges to it as $j \to \infty$. Since

$$y_S^{k_j+1} - y_M^{k_j} - \theta_M^{k_j} + R_{k_j}(\zeta_S^{k_j+1} - \zeta_M^{k_j}) = 0,$$

using the stated hypotheses we conclude that

$$\lim_{j\to\infty} (y_M^{k_j} + \theta_M^{k_j}) = \bar{y}.$$

On the other hand, taking into account (4.29) and passing onto further subsequences if necessary, we can assume that

$$\begin{aligned}
\lim_{j\to\infty} w_S^{k_j+1} &= \lim_{j\to\infty} w_M^{k_j} = \bar{w}, \\
\lim_{j\to\infty} \zeta_S^{k_j+1} &= \lim_{j\to\infty} \zeta_M^{k_j} = \bar{\zeta}, \\
\lim_{j\to\infty} \mu_S^{k_j+1} &= \lim_{j\to\infty} \mu_M^{k_j} = \bar{\mu}.
\end{aligned}$$

Since the families of approximations $\{F_k^{-1}\}$ and $\{G_k^{-1}\}$ are equicontinuous on compact sets, there exists, for each t, some k_{j_t} such that for every k it holds that

$$d_H\left(F_k^{-1}(w_S^{k_{j_t}+1}), F_k^{-1}(w_M^{k_{j_t}})\right) < \frac{1}{t}.$$

In particular, there exists $\hat{x}_M^{k_{j_t}} \in F_k^{-1}(w_M^{k_{j_t}}) \subset F^{-1}(w_M^{k_{j_t}})$ such that

$$|x_S^{k_{j_t}+1} - \hat{x}_M^{k_{j_t}}| < \frac{1}{t}.$$

Hence,

$$\lim_{t\to\infty} \hat{x}_M^{k_{j_t}} = \bar{x}.$$

Then, since F is outer semicontinuous, we conclude that $\bar{x} \in F^{-1}(\bar{w})$. In a similar way we can conclude that $\bar{y} \in G^{-1}(\bar{\zeta})$. Also, note that since $(w_S^{k_j+1}, \mu_S^{k_j+1}) \in S_{D_S}$, we have that $(\bar{w}, \bar{\zeta}) \in S_{D_S}$, and since $(w_M^{k_j}, \mu_M^{k_j}) \in S_{D_M}$, also $(\bar{w}, \bar{\zeta}) \in S_{D_M}$. Thus $(\bar{w}, \bar{\zeta}, \bar{\mu}) \in S_D$.

Finally, since $(w_S^{k_j+1}, \mu_S^{k_j+1})$ solves (4.10) with $x_S^{k_j+1} \in F_{k_j}^{-1}(w_S^{k_j+1})$, we have that for every $(w, \mu) \in S_{D_{S2}}$ the following inequality holds:

$$\begin{aligned}\langle x_S^{k_j+1} + Q_{k_j}(w_S^{k_j+1} - w_M^{k_j}), w - w_S^{k_j+1}\rangle \\ + \langle d - B(y_M^{k_j} - \theta_M^{k_j}) + P_{k_j}(\mu_S^{k_j+1} - \mu_M^{k_j}), \mu - \mu_S^{k_j+1}\rangle \geq 0. \quad (4.30)\end{aligned}$$

Thus, passing onto the limit as $j \to \infty$, we obtain that

$$\langle \bar{x}, w - \bar{w}\rangle + \langle d - B\bar{y}, \mu - \bar{\mu}\rangle \geq 0,$$

that is,

$$\langle \bar{x}, w - \bar{w}\rangle + \langle d, \mu - \bar{\mu}\rangle + \langle \bar{y}, -\bar{\zeta} - B^\top \mu\rangle \geq 0.$$

So, for each $(w, \zeta, \mu) \in S_D$ we have that

$$\langle \bar{x}, w - \bar{w}\rangle + \langle d, \mu - \bar{\mu}\rangle + \langle \bar{y}, \zeta - \bar{\zeta}\rangle \geq 0.$$

Since also $\bar{x} \in F^{-1}(\bar{w})$ and $\bar{y} \in G^{-1}(\bar{\zeta})$, we conclude that $(\bar{w}, \bar{\zeta}, \bar{\mu})$ is a solution of VI(F_D, S_D). The latter implies, by Proposition 4.1.1, that $(\bar{x}, \bar{y})$ is a solution of (4.1).

□

This chapter finishes the decomposition part of this work for VIs. In summary, whereas the Dantzig-Wolfe decomposition algorithm deals with coupling constraints (a setting that fits naturally VI problems arising from GNEPs), the Benders decomposition algorithm is intended for VIs where we recognize the existence of "coupling variables". This kind of problems arises naturally when we model a process in which the decision variable is divided in two parts. One part is intended for investment decisions (say, z), and the other for operational decisions (say, q). The later may have a decomposable structure that is coupled by the investment variable (say a constraint of the form $Zz + Qq \geq b$, with Q decomposable by blocks).

We can use the Benders approach to reduce the solution of this type of VI, solving instead a number of subproblems with variables in lower dimensions. This is especially useful when modeling stochastic markets where the "wait-and-see" variables are large but have a decomposable structure. In the following two chapters we study the problem of finding deterministic and stochastic equilibria by means of GNEP models whose associated VI problems fit the use of the two decomposition algorithms described in Chapters 3 and 4.

Chapter 5

Finding Equilibria in Energy Markets

In spite of an undeniable worldwide trend of liberalization, industries dealing with energy networks (and to a lesser extent with water supply) continue to be subject to regulation in price, entry, and service quality of the network. Regarding electricity and natural gas transmission and distribution, the specific mechanism chosen for regulation impacts significantly competition and affects the network prices, investment and reliability.

In general, good performance of the regulatory framework results in lower operation and transmission costs, better service quality, and investment to expand the network and face future changes in demand and supply. Regulation plays an important role too in the presence of environmental concerns, for example encouraging carbon trading to reduce CO2 emissions.

It is therefore important to fully understand the interaction of competing agents in a market of energy that is subject to various regulatory interventions. Due to the presence of relatively few companies generating power in a given region, electricity markets are naturally set in an oligopolistic competition framework. A similar situation arises in the natural gas industry.

In a centralized environment the paradigm of cost minimization defines energy prices based on marginal costs or shadow prices obtained by optimization. In a liberalized setting, by contrast, prices are computed through equilibrium models aimed at ensuring profit maximization for all the agents. These type of models can be formulated in different manners, for example by means of mixed complementarity problems, bi-level programming, mathematical programs with equilibrium constraints. There are many references, among which we mention [37], [22], [61], [39],[84], [24], [13], [85], without the claim of being exhaustive.

In this chapter we follow the exposition in [50] to explore the relations between mixed complementarity, variational inequality, and game-theoretical formulations of energy markets both in deterministic and stochastic settings. Our analysis shows that the profit-maximization complementarity formulation

is equivalent to a game with agents minimizing costs if the setting is deterministic or risk neutral. By contrast, when the agents in the market exhibit risk aversion, a natural phenomenon in this type of markets, the equivalence no longer holds. More precisely, the risk-averse game becomes equivalent to a complementarity model where agents maximize the expected remuneration and hedge risk only in the cost.

In the development that follows we consider a stylized energy market that is general enough to cover the generation capacity expansion model [23] as well as the European natural gas market model in [32].

5.1 A simple network of agents

Our market is composed by producers, traders, and one end-consumption sector. Producers generate some kind of good (electricity, natural gas) that is sold to traders. Traders take the producers output and sell it to consumers after transporting and possibly modifying the product. The model can easily incorporate pipeline and storage operators, marketers, and other outsourcing agents like in [32]. For simplicity, and without loss of generality, in our presentation we analyze a network with only producers and traders that captures the main properties of the market model. Differently from [32], where it is considered only operational decision variables (e. g. production), we consider a setup suitable for [23], where it is allowed investment variables (e. g. production capacity expansion along time), in which decision variables are separated in two stages. For producers, for instance, some investment to increase capacity has to be decided at stage 0, in order to decide how much to generate at stage 1. Another example is, in the presence of uncertainty, when the second stage variables are a recourse to correct first stage decisions, taken before knowing the realization of uncertainty; [17].

In what follows, at equilibrium, all variables are denoted with a bar, for instance $\bar{\pi}$ stands for an equilibrium price.

5.1.1 Producers

There are N_P producers, each one with decision variable (z_P^i, q_P^i). As mentioned, the variable z_P^i could refer to decisions concerning capacity or technological investments with a smooth concave cost $I_P^i(z_P^i)$. The variable q_P^i is related to operational activities involving a (smooth concave) cost $c_P^i(q_P^i)$. All the producer decision variables are taken in some set X_P^i which represents technological and resource constraints. After transformation of the raw materials, expressed by a matrix S_P^i of suitable dimensions, the producer has the quantity $S_P^i q_P^i$ for sale. In our model, we suppose that producers are of the *price taker* type: they assume that there exists a market price that they can not influence directly. So, for a given price π_P (exogenous to the players) each producer tries to maximize

its profit by solving the following problem:

$$\begin{cases} \max & \left\langle S_P^i q_P^i, \pi_P \right\rangle - c_P^i(q_P^i) - I_P^i(z_P^i) \\ \text{s.t.} & (z_P^i, q_P^i) \in X_P^i \,. \end{cases} \tag{5.1}$$

5.1.2 Traders

There are N_T traders, the j-th trader has decision variable (z_T^j, q_T^j). Given a transformation matrix B_T^j of suitable size, the trader buys $B_T^j q_T^j$ from the producers at price π_P. After modifying and/or transporting the product via a matrix S_T^j of suitable dimensions, the quantity $S_T^j q_T^j$ is sold to consumers at price π_T. The trader may have some additional (smooth concave) operational expenses $c_T^j(q_T^j)$ along the process and maximizes its revenue by solving the following problem:

$$\begin{cases} \max & \left\langle S_T^j q_T^j, \pi_T \right\rangle - \left\langle B_T^j q_T^j, \pi_P \right\rangle - c_T^j(q_T^j) - I_T^j(z_T^j) \\ \text{s.t.} & (z_T^j, q_T^j) \in X_T^j \,. \end{cases} \tag{5.2}$$

We will see below that, as in [32], traders have a special role in the market, and can exert market power by withholding supply from end costumers.

5.1.3 Market clearing and consumers modeling

When the market is at equilibrium, there is no excess of generation and the producers supply meets the traders demand:

$$\sum_{i=1}^{N_P} S_P^i \bar{q}_P^i - \sum_{j=1}^{N_T} B_T^j \bar{q}_T^j = 0 \qquad \text{(mult. } \bar{\pi}_P\text{)}. \tag{5.3}$$

The rightmost notation means that the producers are remunerated at a price that clears the market: $\bar{\pi}_P$ is the multiplier corresponding to (5.3) at an equilibrium.

An environmentally responsible regulator can also impose a CO2 clearing condition, similar to (5.3), but involving different emission factors, depending on the technology employed to generate energy. The essential feature of such constraints is that they *couple* the actions of different agents, and in this sense (5.3) suffices for our development.

The representation of the end-consumption sector can be done in different ways, depending on the manner price-taking producers operate in an imperfectly competitive market. Market imperfections can originate in regulatory measures such as price caps, and/or in traders exerting market power. We now review some alternatives that fit our general modeling.

Consumers via inverse-demand function

When a price-sensitive demand curve is available (as in the example 3.3.1), the consumers needs are represented implicitly by their inverse-demand function.

We model it by an affine function $P \cdot + d_0$, depending on given intercept d_0 and matrix P. The dimension of d_0 is the same as of the traders selling price (π_T in (5.2)); the matrix P is of order $|\pi_T| \times |S_T^j q_T^j|$. At equilibrium the constraint

$$\sum_{j=1}^{N_T} P S_T^j \bar{q}_T^j + d_0 - \bar{\pi}_T = 0 \tag{5.4}$$

must be satisfied.

The inverse-demand function is useful to model the influence that the traders may exert on the market, a typical phenomenon in oligopolies. Instead of selling all the goods at price π_T (exogenous, hence not controllable), the trader sells a portion δ^j at price $\sum_{k=1}^{N_T} P S_T^k q_T^k + d_0$ (that depends on the amount of product the trader offers to the market). The factor $\delta^j \in [0,1]$ determines the strength of the influence the trader can have on the market. Accordingly, now the trader's problem (5.2) is

$$\left\{ \begin{array}{ll} \max & \left\langle S_T^j q_T^j, \delta^j (\sum_{k=1}^{N_T} P S_T^k q_T^k + d_0) + (1-\delta^j)\pi_T \right\rangle \\ & \qquad - \left\langle B_T^j q_T^j, \pi_P \right\rangle - c_T^j(q_T^j) - I_T^j(z_T^j) \\ \text{s.t.} & (z_T^j, q_T^j) \in X_T^j \, . \end{array} \right. \tag{5.2}_{\delta^j}$$

For future use, note that the initial problem (5.2) amounts to setting $\delta^j = 0$ for all the traders. Like for (5.2), both prices π_P and π_T are exogenous for the traders.

Consumers via explicit demand constraint

Sometimes there is instead a load duration curve segmented into blocks defining a vector D, which represents the consumers demand. Accordingly, letting q^0 denote a nonnegative variable, at the equilibrium the constraint

$$\sum_{j=1}^{N_T} S_T^j \bar{q}_T^j + \bar{q}^0 - D = 0 \qquad \text{(mult. } \bar{\pi}_T\text{)} \tag{5.5}$$

should be satisfied. To prevent traders from exerting market power, and following [23], the deficit variable is related in a dual manner to a price cap imposed by the regulating agency:

$$\bar{\pi}_T \leq PC \qquad \text{(mult. } \bar{q}^0\text{)}$$

for PC a maximum allowed price. Note that, in view of their definitions, the variables q^0 and π_T have the same dimension.

In what follows, we refer to the model (5.1),$(5.2)_{\delta^j}$,(5.3),(5.4) as *implicit* model; while (5.1),(5.2),(5.3),(5.5), and the price-cap condition define the *explicit* model.

5.2 Equilibrium: mixed complementarity formulation

For both consumers models, the equilibrium problem consists in computing prices $\bar{\pi}$ and decision variables $(\bar{z}, \bar{q})$ such that:

- for the i-th producer, problem (5.1) written with price $\pi_P := \bar{\pi}_P$ is solved by $(\bar{z}_P^i, \bar{q}_P^i)$; and
- for the j-th trader, problem $(5.2)_{\delta^j}$ written with prices $(\pi_P, \pi_T) := (\bar{\pi}_P, \bar{\pi}_T)$ is solved by $(\bar{z}_T^i, \bar{q}_T^i)$, keeping in mind that if the explicit model is used $\delta^j = 0$ for all the traders.
- The market is cleared and (5.3) holds.
- Regarding the price at which the traders sell the final product,
  - if the implicit model is used, the relation (5.4) holds;
  - if the explicit model is used, both (5.5) and the price cap conditions (cf. (5.8) below) hold.

For the sake of clarity we derive first the mixed complementarity problem (MCP) when the consumers model is explicit, i.e., the trader's problem is (5.2) and both (5.5) and the price cap condition hold.

5.2.1 Explicit demand constraint

We start by writing down the KKT conditions for the profit maximization problems of the producers and traders. Let in (5.1) and (5.2) the feasible sets X_P^i and X_T^j be polyhedra of the form

$$Z_P^i z_P^i + Q_P^i q_P^i \geq b_P^i \quad \text{and} \quad Z_T^j z_T^j + Q_T^j q_T^j \geq b_T^j,$$

respectively, and let μ_P^i and μ_T^j denote the corresponding multipliers. The KKT conditions for the producers problems (5.1), dropping the superindices i to alleviate notation, are

$$\begin{array}{rcl} 0 & = & I_P'(z_P) - Z_P^\top \mu_P \\ 0 & = & c_P'(q_P) - Q_P^\top \mu_P - S_P^\top \pi_P \\ 0 & \leq & Z_P z_P + Q_P q_P - b_P \perp \mu_P \geq 0\,. \end{array} \tag{5.6}$$

Similarly for the traders, dropping the superindices j, we write

$$\begin{array}{rcl} 0 & = & I_T'(z_T) - Z_T^\top \mu_T \\ 0 & = & c_T'(q_T) - Q_T^\top \mu_T + B_T^\top \pi_P - S_T^\top \pi_T \\ 0 & \leq & Z_T z_T + Q_T q_T - b_T \perp \mu_T \geq 0\,. \end{array} \tag{5.7}$$

It remains to complete the system with (5.3), (5.5), and the price cap inequality, which we now write as follows:

$$0 \leq PC - \pi_T \perp q^0 \geq 0\,. \tag{5.8}$$

We use the primal and dual variables defined by

$$\begin{aligned} \mathfrak{p} &:= \left((z_P^i)_{i=1}^{N_P}, (q_P^i)_{i=1}^{N_P}, (z_T^j)_{j=1}^{N_T}, (q_T^j)_{i=1}^{N_T}, q^0 \right) \\ \text{and } \mathfrak{d} &:= \left((\mu_P^i)_{i=1}^{N_P}, (\mu_T^j)_{j=1}^{N_T}, \pi_P, \pi_T \right) \end{aligned}$$

over the sets

$$\mathfrak{P} := \mathbb{R}^{\sum_{i=1}^{N_P}(|z_P^i|+|q_P^i|)+\sum_{j=1}^{N_T}(|z_T^j|+|q_T^j|)} \times \mathbb{R}_{\geq 0}^{|q^0|} \tag{5.9}$$

$$\text{and } \mathfrak{D} := \mathbb{R}_{\geq 0}^{\sum_{i=1}^{N_P}|\mu_P^i|+\sum_{j=1}^{N_T}|\mu_T^j|} \times \mathbb{R}^{|\pi_P|+|\pi_T|} \tag{5.10}$$

where $|q^0| = |\pi_T|$, by construction.

To write the associated generalized equation (GE) in a compact form, we introduce the following notation for a sequence of matrices $M^k, k = 1, \ldots, K$:

$$\texttt{diag}(M^k) := \begin{pmatrix} M^1 & & \\ & \ddots & \\ & & M^K \end{pmatrix}, \quad \texttt{col}(M^k) := \begin{pmatrix} M^1 \\ \vdots \\ M^K \end{pmatrix},$$

$$\texttt{row}(M^k) := \begin{bmatrix} M^1 & \ldots & M^K \end{bmatrix}.$$

The transpose of the last matrix is column-like matrix aligning the transposed matrices and, hence,

$$[\texttt{row}(M^k)]^\top = \texttt{col}(M^{k\,\top}).$$

We use the new notation in the matrix below, with row and column dimensions given by the cardinality of $\mathfrak{D}$ and $\mathfrak{P}$, respectively:

$$B := \begin{pmatrix} \texttt{diag}(Z_P^i) & \texttt{diag}(Q_P^i) & 0 & 0 & 0 \\ 0 & 0 & \texttt{diag}(Z_T^j) & \texttt{diag}(Q_T^j) & 0 \\ 0 & \texttt{row}(S_P^i) & 0 & -\texttt{row}(B_T^j) & 0 \\ 0 & 0 & 0 & \texttt{row}(S_T^j) & I \end{pmatrix}, \tag{5.11}$$

where I is an identity matrix of order $|\pi_T| = |q^0|$.

The GE that results from putting together the relations in (5.6), (5.7), (5.3), (5.5), and (5.8) is

$$0 \in \begin{bmatrix} 0 & -B^\top \\ B & 0 \end{bmatrix} \begin{pmatrix} \mathfrak{p} \\ \mathfrak{d} \end{pmatrix} + \begin{pmatrix} (I_P^{i\,\prime}(z_P^i))_{i=1}^{N_P} \\ (c_P^{i\,\prime}(q_P^i))_{i=1}^{N_P} \\ (I_T^{j\,\prime}(z_T^j))_{j=1}^{N_T} \\ (c_T^{j\,\prime}(q_T^j))_{j=1}^{N_T} \\ PC \\ -(b_P^i)_{i=1}^{N_P} \\ -(b_T^j)_{j=1}^{N_T} \\ 0 \\ -D \end{pmatrix} + \mathcal{N}_{\mathfrak{P}\times\mathfrak{D}}(\mathfrak{p}, \mathfrak{d}),$$

where in the last term we use the normal cone to the primal and dual feasible sets. This GE can be rewritten in a more compact form, by introducing an operator acting on primal variables only and a dual vector, as follows:

$$F(\mathfrak{p}) := \begin{pmatrix} (I_P^{i\,\prime}(z_P^i))_{i=1}^{N_P} \\ (c_P^{i\,\prime}(q_P^i))_{i=1}^{N_P} \\ (I_T^{j\,\prime}(z_T^j))_{j=1}^{N_T} \\ (c_T^{j\,\prime}(q_T^j))_{j=1}^{N_T} \\ PC \end{pmatrix} \quad \text{and} \quad b := \begin{pmatrix} (b_P^i)_{i=1}^{N_P} \\ (b_T^j)_{j=1}^{N_T} \\ 0 \\ D \end{pmatrix}, \tag{5.12}$$

yielding the GE

$$0 \in \begin{bmatrix} 0 & -B^\top \\ B & 0 \end{bmatrix} \begin{pmatrix} \mathfrak{p} \\ \mathfrak{d} \end{pmatrix} + \begin{pmatrix} F(\mathfrak{p}) \\ -b \end{pmatrix} + \mathcal{N}_{\mathfrak{P}\times\mathfrak{D}}(\mathfrak{p},\mathfrak{d})\,. \tag{5.13}$$

5.2.2 Inverse-demand function

Since in this case the traders conditions are more involved when there is market power, we shall not drop the superindices j as this could lead to confusion, and write the optimality system for the traders as follows:

$$\begin{aligned} 0 &= I_T^{j\,\prime}(z_T^j) - {Z_T^j}^\top \mu_T^j \\ 0 &= c_T^{j\,\prime}(q_T^j) - {Q_T^j}^\top \mu_T^j + {B_T^j}^\top \pi_P - (1-\delta^j){S_T^j}^\top \pi_T \\ &\qquad -\delta^j {S_T^j}^\top (\sum_{k=1}^{N_T} P S_T^k q_T^k + d_0) - \delta^j {S_T^j}^\top P^\top S_T^j q_T^j \\ 0 &\le Z_T^j z_T^j + Q_T^j q_T^j - b_T^j \perp \mu_T^j \ge 0\,. \end{aligned} \tag{5.14}$$

As before, the KKT conditions (5.6) and (5.14), together with the market clearing condition (5.3) and the implicit representation of consumers via (5.4), give a GE on both primal and dual variables. There are a few differences with (5.13), though:

- There is no deficit variable q^0, so the primal variables and set are now

$$\begin{aligned} \tilde{\mathfrak{p}} &:= \left((z_P^i)_{i=1}^{N_P}, (q_P^i)_{i=1}^{N_P}, (z_T^j)_{j=1}^{N_T}, (q_T^j)_{j=1}^{N_T}\right) \\ \text{and } \tilde{\mathfrak{P}} &:= \mathbb{R}^{\sum_{i=1}^{N_P}(|z_P^i|+|q_P^i|)+\sum_{j=1}^{N_T}(|z_T^j|+|q_T^j|)}\,. \end{aligned}$$

  Accordingly, instead of the matrix B from (5.11), we consider the submatrix $\tilde{B}$ obtained by eliminating from B the last row and column. Dual variables are still the same, so the GE will use $\tilde{B}$ and an additional row, to represent the relations in (5.4).

- The market power terms in the third line in (5.14) enter the primal oper-

ator, which becomes

$$\tilde{F}(\tilde{\mathfrak{p}}) := \begin{pmatrix} I_P^{i\,\prime}(z_P^i) \\ c_P^{i\,\prime}(q_P^i) \\ I_T^{j\,\prime}(z_T^j) \\ c_T^{j\,\prime}(q_T^j) \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \\ \delta^j {S_T^j}^\top (\sum_{k=1}^{N_T} P S_T^k q_T^k + d_0) + \delta^j {S_T^j}^\top P^\top S_T^j q_T^j \,. \end{pmatrix}$$

To alleviate the writing we omitted the superindices ranges: $i = 1, \ldots, N_P$ and $j = 1, \ldots, N_T$, which are clear from the context; see (5.12).

- Replacing (5.5) by (5.4) modifies the dual vector as follows

$$\tilde{b} := \begin{pmatrix} b_P^i \\ b_T^j \\ 0 \\ -d_0 \end{pmatrix}$$

where, once again, i and j run in their respective ranges, as in (5.12).

Finally, the GE with the implicit model is

$$0 \quad \in \quad \begin{bmatrix} & 0 & & -\tilde{B}^\top \begin{pmatrix} 0 \\ 0 \\ 0 \\ -\mathtt{col}((1-\delta^j){S_T^j}^\top) \end{pmatrix} \\ & \tilde{B} & & 0 \\ \begin{bmatrix} 0 & 0 & \mathtt{row}(PS_T^j) & 0 \end{bmatrix} & & 0 & -I \end{bmatrix} \begin{pmatrix} \tilde{\mathfrak{p}} \\ \mathfrak{d} \end{pmatrix}$$

$$+ \begin{pmatrix} \tilde{F}(\tilde{\mathfrak{p}}) \\ -\tilde{b} \end{pmatrix} + \mathcal{N}_{\tilde{\mathfrak{P}} \times \mathfrak{D}}(\mathfrak{p}, \mathfrak{d}) \,. \tag{5.15}$$

When compared to (5.13), the GE above does not involve a skewed symmetric linear part. In particular, the last line in the matrix relates primal and dual elements.

We shall see in Section 5.3 that GEs of the form (5.13) can be reduced to variational inequalities in smaller dimensions, which can in turn be interpreted in terms of a Nash game with shared constraints. GE (5.15), on the other hand, cannot be reformulated the same way directly. We next rewrite (5.15) in an equivalent form that does have the desired properties.

5.2.3 Inverse-demand function and an extra variable

Taking inspiration from the explicit model, we introduce a new primal variable p^0, gathering the portion of supply that the traders cannot influence by exerting

market power. Thus, we require that the relation

$$\sum_{j=1}^{N_T}(1-\delta^j)S_T^j\bar{q}_T^j - \bar{p}^0 = 0 \tag{5.16}$$

be satisfied when the market is at an equilibrium point. In view of its definition, this new variable has the same dimension as the deficit variable q^0 from (5.5) in the explicit model (and, hence, $|p^0| = |\pi_T|$).

The KKT conditions (5.6) and (5.14), together with (5.3), (5.4), and (5.16) now employ the primal variables and set

$$\begin{aligned} \hat{\mathfrak{p}} &:= \left((z_P^i)_{i=1}^{N_P}, (q_P^i)_{i=1}^{N_P}, (z_T^j)_{j=1}^{N_T}, (q_T^j)_{i=1}^{N_T}, p^0\right) \\ \text{and } \widehat{\mathfrak{P}} &:= \mathbb{R}^{\sum_{i=1}^{N_P}(|z_P^i|+|q_P^i|)+\sum_{j=1}^{N_T}(|z_T^j|+|q_T^j|)+|p^0|}, \end{aligned} \tag{5.17}$$

noting that the dual variables remain the same from the explicit model, given in (5.10).

The primal sets in the implicit and explicit models, from (5.17) and (5.9) respectively, only differ in the last component (q^0 and p^0, respectively). Specifically, while in the explicit model the deficit is nonnegative (as a multiplier of the price cap (5.8)), in the implicit model the new primal variable is unconstrained. So the normal cone to q^0 will be the null vector and we can require satisfaction of the inverse-demand relation (5.4) in the corresponding new component of the GE. This eliminates the primal-dual coupling in the last line of the linear term in (5.15). Similarly, recalling that the π_T-component of the dual feasible set is the whole space, satisfaction of (5.16) will be ensured by the last component of the GE.

Define the matrix $\hat{B}$ with size given by the cardinality of $\widehat{\mathfrak{P}}$ and $\mathfrak{D}$ by

$$\hat{B} := \begin{pmatrix} \tilde{B} & 0 \\ \begin{bmatrix} 0 & 0 & 0 & \texttt{row}((1-\delta^j)S_T^j) \end{bmatrix} & -\hat{I} \end{pmatrix}, \tag{5.18}$$

where the identity matrix $\hat{I}$ has order $|\pi_T| = |p^0|$. The resulting GE is

$$0 \in \begin{bmatrix} 0 & -\hat{B}^\top \\ \hat{B} & 0 \end{bmatrix} \begin{pmatrix} \hat{\mathfrak{p}} \\ \mathfrak{d} \end{pmatrix} + \begin{pmatrix} \hat{F}(\hat{\mathfrak{p}}) \\ -\hat{b} \end{pmatrix} + \mathcal{N}_{\widehat{\mathfrak{P}}\times\mathfrak{D}}(\mathfrak{p},\mathfrak{d}), \tag{5.19}$$

for the primal operator and dual vector given below:

$$\widehat{F}(\hat{\mathfrak{p}}) := \begin{pmatrix} I_P^{i\,\prime}(z_P^i) \\ c_P^{i\,\prime}(q_P^i) \\ I_T^{j\,\prime}(z_T^j) \\ c_T^{j\,\prime}(q_T^j) - \delta^j S_T^{j\,\top}(\sum_{k=1}^{N_T} P S_T^k q_T^k + d_0) - \delta^j S_T^{j\,\top} P^\top S_T^j q_T^j \\ -\sum_{k=1}^{N_T} P S_T^k q_T^k - d_0 \end{pmatrix} \text{ and}$$

$$\hat{b} := \begin{pmatrix} b_P^i \\ b_T^j \\ 0 \\ 0 \end{pmatrix}. \tag{5.20}$$

In both vectors superindices are to be understood as $i = 1, \ldots, N_P$ and $j = 1, \ldots, N_T$, like in (5.12).

5.3 Equivalent formulations for the mixed complementarity problem

Both GEs (5.13) and (5.19) are defined using very simple normal cones, and have a very specific primal-dual structure. The size of both GEs is the same: the respective primal and dual sets only differ is that the last primal component, which is nonnegative in the explicit model (q^0), and unconstrained in the implicit one (p^0); see Table 5.1.

Table 5.1: Primal and dual dimensions of the complementarity formulations

Gen. Equation	Primal Set and Dimension	Dual Set and Dimension
(5.13)	$\mathfrak{P}$ from (5.9)	$\mathfrak{D}$ from (5.10)
(explicit)	$\sum_{i=1}^{N_P}(\lvert z_P^i\rvert + \lvert q_P^i\rvert) + \sum_{j=1}^{N_T}(\lvert z_T^j\rvert + \lvert q_T^j\rvert) + \lvert q^0\rvert$	$\sum_{i=1}^{N_P}\lvert\mu_P^i\rvert + \sum_{j=1}^{N_T}\lvert\mu_T^j\rvert + \lvert\pi_P\rvert + \lvert\pi_T\rvert$
(5.19)	$\widehat{\mathfrak{P}}$ from (5.17)	$\mathfrak{D}$ from (5.10)
(implicit+1var.)	same as explicit model ($\lvert p^0\rvert = \lvert q^0\rvert$)	same as explicit model

To establish the relation of the MCP models with a game-theoretical formulation, we state a result from [62]; see also [34]. Our GEs are a particular case of the setting covered by the reduction method in [62], as the linear part is skew-symmetric and the primal sets $\mathfrak{P}$ are cones in both models. Here, the relation with a game could actually be also shown directly, by comparing the KKT conditions of the MCP model with those for a game (stated in Section 2.4). We prefer to state the more general result, because it includes a nice characterization of dual variables as solutions to a certain linear programming problem,

defined *a posteriori*, once the primal solution is available. As dual variables have economical meaning as prices, this is an interesting feature; we comment more on this after the theorem and in Remark 5.3.4.

Theorem 5.3.1. *The following statements are equivalent:*

Primal-Dual GE: *the primal-dual pair* $(\bar{\mathfrak{p}}, \bar{\mathfrak{d}})$ *satisfies* (5.13).

Primal GE + Dual LP: *the primal variable* $\bar{\mathfrak{p}}$ *solves the generalized equation*

$$0 \in F(\mathfrak{p}) + \mathcal{N}_{\mathfrak{P}^0}(\mathfrak{p}) \tag{5.21}$$

where $\mathfrak{P}^0 := \mathfrak{P} \cap \mathcal{S}$ *and*

$$\mathcal{S} := \left\{ \mathfrak{p} : b - B\mathfrak{p} \in \mathbb{R}^{\sum_{i=1}^{N_P} |\mu_P^i| + \sum_{j=1}^{N_T} |\mu_T^j|}_{\leq 0} \times \{0 \in \mathbb{R}^{|\pi_P| + |\pi_T|}\} \right\}.$$

As for the dual variable, $\bar{\mathfrak{d}}$ *solves the linear programming problem*

$$\begin{cases} \min & \langle B\bar{\mathfrak{p}} - b, \mathfrak{d} \rangle \\ \text{s.t.} & B^\top \mathfrak{d} - F(\bar{\mathfrak{p}}) \in \mathcal{N}_{\mathfrak{P}}(\bar{\mathfrak{p}}) \\ & \mathfrak{d} \in \mathfrak{D}\,. \end{cases} \tag{5.22}$$

Proof. The statement is just a rewriting of Propositions 1 and 2 in [62] in our notation. Specifically, the respective correspondence for primal elements is $(p, d(p), P) = (\mathfrak{p}, -F(\mathfrak{p}), \mathfrak{P})$, for the dual ones $(y, Y) = (\mathfrak{d}, \mathfrak{D})$, and for the matrix and vector $(A, f) = -(B^\top, b)$. Our coupling set $\mathcal{S}$ corresponds to the set Z in Proposition 2, using the fact that in our setting the polar cone therein, $Y^0 = \mathfrak{D}^0$, has a very simple expression. □

Existence of solutions to the generalized equation (5.21) can be guaranteed under mild assumptions, such as continuity of F and convexity and compactness of $\mathfrak{P}^0$, [27, Corollary 2.2.5]. These conditions are natural in our context: components of F consist of derivatives of smooth convex functions and the feasible set $\mathfrak{P}^0$ represents limited resources. Furthermore, the existence of solutions of (5.21) implies the existence of solutions of the (bounded) linear program (5.22), whose optimal value is zero.

The interest of Theorem 5.3.1 is twofold. First, the GE (5.21) is in primal variables only, stated over a set that (for both of our models) is a simple polyhedron. It is therefore a VI with linear constraints. We shall see that in some cases the multipliers corresponding to the constraints provide the equilibrium prices. Once a primal solution is at hand, the dual component of the MCP solution can be found by solving an easy linear program. This feature is attractive to identify (undesirable) situations in which equilibrium prices are not unique, even if the primal part of the equilibrium points is unique (the linear program solution will not be unique in this case; see Remark 5.3.4 below). A second advantage of the equivalent formulation is that, in addition to providing a mechanism for ensuring existence of solutions of the game, the equivalent formulation reveals the particular structure of the set $\mathfrak{P}^0$, amenable to decomposition. More precisely,

without the coupling constraints (some components in $b - B\mathfrak{p}$), the feasible set is decomposable (like $\mathfrak{P}$ from (5.9), (5.17)). This is often the case in generalized Nash games, that can then be exploited by decomposition methods, like the Dantzig-Wolfe and Benders' algorithms developed in Sections 3 and 4.

We now explore the relations between MCP and game formulations for the markets in Section 5.1

5.3.1 Game for the explicit model

Instead of viewing the agents as maximizing revenue, like in the complementarity model, we consider a GNEP as in Section 2.4, where the players minimize costs. The coupling constraints in the game are (5.3) and (5.5). In addition to the traders and producers, there is an additional player, indexed by number "0", in charge of the price caps. Specifically, given a primal point

$$\mathfrak{p} = ((\tilde{z}_P^i)_{i=1}^{N_P}, (\tilde{q}_P^i)_{i=1}^{N_P}, (\tilde{z}_T^j)_{j=1}^{N_T}, (\tilde{q}_T^j)_{i=1}^{N_T}, \tilde{q}^0),$$

the purpose of the game is to solve the following minimization problems:

$$\textbf{Producers} \quad \begin{cases} \min & I_P^i(z_P^i) + c_P^i(q_P^i) \\ \text{s.t.} & (z_P^i, q_P^i) \in X_P^i \\ & S_P^i q_P^i + \sum\limits_{\substack{k=1 \\ k \neq i}}^{N_P} S_P^k \tilde{q}_P^k - \sum\limits_{j=1}^{N_T} B_T^j \tilde{q}_T^j = 0. \end{cases} \tag{5.23}$$

$$\textbf{Traders} \quad \begin{cases} \min & I_T^j(z_T^j) + c_T^j(q_T^j) \\ \text{s.t.} & (z_T^j, q_T^j) \in X_T^j\,. \\ & -B_T^j \tilde{q}_T^j + \sum\limits_{i=1}^{N_P} S_P^i \tilde{q}_P^i - \sum\limits_{\substack{k=1 \\ k \neq j}}^{N_T} B_T^k \tilde{q}_T^k = 0 \\ & S_T^j q_T^j + \sum\limits_{\substack{k=1 \\ k \neq i}}^{N_T} S_T^k \tilde{q}_T^k + \tilde{q}^0 - D = 0 \end{cases} \tag{5.24}$$

$$\textbf{Consumers' representative} \quad \begin{cases} \min & \langle PC, q^0 - D \rangle \\ \text{s.t.} & q^0 \geq 0\,. \\ & \sum\limits_{j=1}^{N_T} S_T^j \tilde{q}_T^j + q^0 - D = 0 \end{cases} \tag{5.25}$$

In the GNEP (5.23)-(5.25), the market between producers and traders is cleared, and demand is satisfied up to certain deficit, q^0. The deficit is minimized by the action of the additional player, who tries to reduce the impact of imposing a price cap. In Corollary 5.3.2 below it is shown that (the negative of) the multiplier

of the coupling constraint (5.5) is precisely the traders' remuneration in (5.2). We shall also see that in the game formulation, the price cap is maintained in an indirect manner, via (5.25).

As already explained, in the game the solution of each individual problem depends on the decisions of the other agents in the market. For example, (5.24) is an optimization problem on the j-th trader variables (say, $\mathfrak{p}_j$), that depends on actions of other traders (say, on $\mathfrak{p}_{-j}$). A primal point

$$\bar{\mathfrak{p}} = ((\check{z}_P^i)_{i=1}^{N_P}, (\check{q}_P^i)_{i=1}^{N_P}, (\check{z}_T^j)_{j=1}^{N_T}, (\check{q}_T^j)_{i=1}^{N_T}, \check{q}^0)$$

is a Nash equilibrium for the game (5.23)–(5.25) when each player's optimal decision (say, $\bar{\mathfrak{p}}_j$) is obtained by solving the individual problem (say, (5.24)) after fixing the other players' decisions to the corresponding entries on $\bar{\mathfrak{p}}$ (say $\bar{\mathfrak{p}}_{-j}$). As this notion is so general that it includes points contradicting the natural intuition of what an equilibrium must be, it is further specialized as follows.

Note that the value function for the producers, defined as

$$v_P^i(x) := \left\{ \begin{array}{ll} \min & I_P^i(z_P^i) + c_P^i(q_P^i) \\ \text{s.t.} & (z_P^i, q_P^i) \in X_P^i \\ & S_P^i q_P^i + \sum\limits_{\substack{k=1 \\ k \neq i}}^{N_P} S_P^k \tilde{q}_P^k - \sum\limits_{j=1}^{N_T} B_T^j \tilde{q}_T^j = x\,, \end{array} \right.$$

is convex. Furthermore, because in (5.23) all constraints are linear and the objective function is differentiable, there exists a Lagrange multiplier $\check{\pi}_P^i$ associated to the equality constraint. This multiplier represents represents a marginal cost, since it satisfies the inclusion $-\check{\pi}_P^i \in \partial v_P^i(0)$, [36, Theorem VII.3.3.2]. The issue with a generic Nash equilibrium like $\bar{\mathfrak{p}}$ is that it may have multipliers associated to coupling constraints of the players' problems that are different for the different players. In economical terms, this means that the equilibrium is "unfair", because it benefits some players more than others. To avoid this undesirable feature, we shall solve a VI derived from the game and find a variational equilibrium of the GNEP (5.23)–(5.25), ensuring that the multipliers associated with the coupling constraints are the same.

By Theorem 5.3.1, the GE (5.13) is equivalent to solving the GE (5.21), written with the data from Subsection 5.1.3. Putting together (5.11), (5.9) and (5.12) yields for (5.21) the following:

$$0 \in \begin{pmatrix} (I_P^{i\,\prime}(z_P^i))_{i=1}^{N_P} \\ (c_P^{i\,\prime}(q_P^i))_{i=1}^{N_P} \\ (I_T^{j\,\prime}(z_T^j))_{j=1}^{N_T} \\ (c_T^{j\,\prime}(q_T^j))_{j=1}^{N_T} \\ PC \end{pmatrix} + \mathcal{N}_{\mathfrak{P}^0}(\mathfrak{p}),$$

where

$$\mathfrak{P}^0 := \prod_{i=1}^{N_P} X_P^i \times \prod_{j=1}^{N_T} X_T^j \times \mathbb{R}_{\geq 0}^{|q^0|} \cap \mathcal{S},$$

$$\mathcal{S} := \{(z_P^i, q_P^i, z_T^j, q_P^j, q^0) : (5.3) \text{ and } (5.5) \text{ hold}\}.$$

The equivalence between the MCP formulation and the generalized Nash game follows from applying Theorem 5.3.1.

Corollary 5.3.2 (Game formulation for the explicit model)**.** *The MCP in Subsection 5.1.3 and the game* (5.23)-(5.25) *are equivalent, in the following sense. Suppose the game has a variational equilibrium*

$$\bar{\mathfrak{p}} := \left((\check{z}_P^i)_{i=1}^{N_P}, (\check{q}_P^i)_{i=1}^{N_P}, (\check{z}_T^j)_{j=1}^{N_T}, (\check{q}_T^j)_{i=1}^{N_T}, \check{q}^0 \right),$$

with $(\check{\mu}_P^i)_{i=1}^{N_P}, (\check{\mu}_T^j)_{j=1}^{N_T}$ *being the corresponding multipliers for the constraints in* (5.23) *and* (5.24)*, and let* $\check{\pi}_P$ *and* $\check{\pi}_T$ *be the multipliers associated to the coupling constraints* (5.3) *and* (5.5)*.*

Then the primal-dual pair $(\bar{\mathfrak{p}}, \bar{\mathfrak{d}})$ *with* $\bar{\mathfrak{d}} := (\check{\mu}_P, \check{\mu}_T, -\check{\pi}_P, -\check{\pi}_T)$ *solves the MCP given by* (5.1)-(5.3)*,* (5.5)*, and* (5.8)*.*

Proof. By Theorem 5.3.1, for the result to hold $\bar{\mathfrak{d}}$ needs to solve the linear program therein. For the objects in (5.13), and for the normal cone to the primal set $\mathfrak{P}$ from (5.9), this linear program is

$$\left\{ \begin{array}{ll} \min\limits_{\substack{\mu_P, \mu_T \geq 0 \\ \text{any } \pi_P, \pi_T}} & \sum\limits_{i=1}^{N_P} \left\langle Z_P^i \check{z}_P^i + Q_P^i \check{q}_P^i - b_P^i, \mu_P^i \right\rangle + \sum\limits_{j=1}^{N_T} \left\langle Z_T^j \check{z}_T^j + Q_T^j \check{q}_T^j - b_T^j, \mu_T^j \right\rangle \\ \text{s.t.} & Z_P^{i\top} \mu_P^i = {I_P^i}'(\check{z}_P^i)\,,\ Z_T^{j\top} \mu_T^j = {I_T^j}'(\check{z}_T^j) \\ & Q_P^{i\top} \mu_P^i + S_P^{i\top} \pi_P = {c_P^i}'(\check{q}_P^i) \\ & Q_T^{j\top} \mu_T^j - B_T^{j\top} \pi_P + S_T^{j\top} \pi_T = {c_T^j}'(\check{q}_T^j) \\ & \pi_T \leq PC \text{ and } \pi_T^k = PC^k \text{ whenever } \check{q}^{0k} > 0\,. \end{array} \right. \tag{5.26}$$

The optimality conditions for problems (5.23) and (5.24) amount to $\check{\mu}_P, \check{\mu}_T, -\check{\pi}_P$ and $-\check{\pi}_T$ satisfy the first four equalities in the feasible set of (5.26). Note also that, by complementarity, the (nonnegative) objective function attains its minimum value at $\check{\mu}_P, \check{\mu}_T$. The last line in (5.26), written with $-\check{\pi}_T$, is $\check{q}^0 \perp PC + \check{\pi}_T \geq 0$; since these relations result from the optimality condition of (5.25), the desired result follows. □

5.3.2 Game for the implicit model

We now apply Theorem 5.3.1 to the GE (5.19). Writing (5.21) with the data from Subsection 5.1.3, that is, using (5.18), (5.17) and (5.20), we have:

$$
\begin{aligned}
0 \quad \in \quad & \begin{pmatrix}
I_P^{i\,\prime}(z_P^i) \\
c_P^{i\,\prime}(q_P^i) \\
I_T^{j\,\prime}(z_T^j) \\
c_T^{j\,\prime}(q_T^j) - \delta^j {S_T^j}^\top (\sum_{k=1}^{N_T} P S_T^k q_T^k + d_0) - \delta^j {S_T^j}^\top P^\top S_T^j q_T^j \\
-\sum_{k=1}^{N_T} P S_T^k q_T^k - d_0
\end{pmatrix} + \mathcal{N}_{\mathfrak{P}^0}(\mathfrak{p}) \\
& \text{where } \mathfrak{P}^0 := \textstyle\prod_{i=1}^{N_P} X_P^i \times \prod_{j=1}^{N_T} X_T^j \times \mathbb{R}^{|p^0|} \cap \mathcal{S} \\
& \text{and } \mathcal{S} := \{(z_P^i, q_P^i, z_T^j, q_P^j, p^0) : (5.3) \text{ and } (5.16) \text{ hold}\}.
\end{aligned}
$$

The MCP formulation of (5.1), $(5.2)_{\delta^j}$, (5.3) and (5.4) is now equivalent to the following generalized Nash game.

Producers same as (5.23)

$$\textbf{Traders} \quad \begin{cases} \min\limits_{z_T, q_T} & I_T^j(z_T^j) + c_T^j(q_T^j) \\ & -\delta^j \left\langle \sum\limits_{k=1}^{N_T} P S_T^k q_T^k + d_0, S_T^j q_T^j \right\rangle \\ \text{s.t.} & (z_T^j, q_T^j) \in X_T^j \,. \\ & \sum\limits_{i=1}^{N_P} S_P^i q_P^i - \sum\limits_{k=1}^{N_T} B_T^k q_T^k = 0 \\ & \sum\limits_{k=1}^{N_T} (1-\delta^k) S_T^k q_T^k - p^0 = 0 \end{cases} \tag{5.27}$$

$$\textbf{Consumers' representative} \quad \begin{cases} \max\limits_{p^0} & \left\langle \sum\limits_{k=1}^{N_T} P S_T^k q_T^k + d_0, p^0 \right\rangle \\ \text{s.t.} & \sum\limits_{j=1}^{N_T} (1-\delta^j) S_T^j q_T^j - p^0 = 0 \end{cases} \tag{5.28}$$

The game (5.23),(5.27)-(5.28) can be interpreted as follows. The additional player tries to maximize the revenue of all the traders, perceiving their remuneration in terms of the inverse-demand function. The traders see their influence on the market as a way of reducing costs, or of increasing their income (the negative δ^j term in the objective function from (5.27)). Transactions between producers and traders are cleared, as before. Regarding the traders remuneration π_T (that is, the multiplier of constraint (5.16)), we now show that the additional player controls it in a manner ensuring satisfaction of (5.4).

Corollary 5.3.3 (Game formulation for the implicit model)**.** *The MCPs in Subsections 5.1.3 and 5.2.3 and the game* (5.23),(5.27)-(5.28) *are equivalent in the following sense. Suppose the game has a variational equilibrium*

$$\bar{\mathfrak{p}} := \left((\check{z}_P^i)_{i=1}^{N_P}, (\check{q}_P^i)_{i=1}^{N_P}, (\check{z}_T^j)_{j=1}^{N_T}, (\check{q}_T^j)_{i=1}^{N_T}, \check{p^0} \right),$$

with $(\check{\mu}_P^i)_{i=1}^{N_P}, (\check{\mu}_T^j)_{j=1}^{N_T}$ *being the corresponding multipliers for the constraints in* (5.23) *and* (5.27)*, and let* $\check{\pi}_P$ *and* $\check{\pi}_T$ *be the multipliers associated to the coupling constraints* (5.3) *and* (5.16)*.*

Then the primal-dual pair $(\bar{\mathfrak{p}}, \bar{\mathfrak{d}})$ *with* $\bar{\mathfrak{d}} := (\check{\mu}_P, \check{\mu}_T, -\check{\pi}_P, -\check{\pi}_T)$ *solves the MCP* (5.19)*, which is equivalent to* (5.15)*.*

Proof. Like for Corollary 5.3.2, we only need to show that $\bar{\mathfrak{d}}$ solves the linear program in Theorem 5.3.1. In this case, the normal cone to the primal set $\widehat{\mathfrak{P}}$

from (5.17) is just the null vector and, hence, the linear program is

$$\begin{cases} \min\limits_{\substack{\mu_P\,,\,\mu_T\,\geq\,0\\ \text{any }\pi_P\,,\,\pi_T}} & \displaystyle\sum_{i=1}^{N_P}\left\langle Z_P^i\check{z}_P^i + Q_P^i\check{q}_P^i - b_P^i, \mu_P^i\right\rangle + \sum_{j=1}^{N_T}\left\langle Z_T^j\check{z}_T^j + Q_T^j\check{q}_T^j - b_T^j, \mu_T^j\right\rangle \\ \text{s.t.} & Z_P^{i\,\top}\mu_P^i = {I^i}_P'(\check{z}_P^i)\,, \quad Z_T^{j\,\top}\mu_T^j = {I^j}_T'(\check{z}_T^j) \\ & Q_P^{i\,\top}\mu_P^i + S_P^{i\,\top}\pi_P = {c^i}_P'(\check{q}_P^i) \\ & Q_T^{j\,\top}\mu_T^j - B_T^{j\,\top}\pi_P + (1-\delta^j)S_T^{j\,\top}\pi_T = {c^j}_T'(\check{q}_T^j) \\ & \qquad\qquad -\delta^j S_T^{j\,\top}\Big(\sum_{k=1}^{N_T} PS_T^k\check{q}_T^k + d_0\Big) - \delta^j S_T^{j\,\top}P^\top S_T^j\check{q}_T^j \\ & \pi_T = \displaystyle\sum_{k=1}^{N_T} PS_T^k\check{q}_T^k + d_0\,. \end{cases} \tag{5.29}$$

It is easy to see that all the relations in KKT conditions of this problem are verified by $\bar{\mathfrak{d}}$, except for the last equality, corresponding to (5.4). For the latter, observe that since $\check{\pi}_T$ is the multiplier of the coupling constraint (5.16), and the variable p^0 is unconstrained in the problem of the extra player (5.28), the p^0-component of the optimality conditions for the game gives that $0 = -\sum_{k=1}^{N_T} PS_T^k\check{q}_T^k - d_0 - \check{\pi}_T$, and the result follows. □

Remark 5.3.4 (Uniqueness of prices)**.** *With the explicit model, equilibrium prices will be unique if the linear program* (5.26) *has the unique solution. Likewise for the implicit model, which depends on the linear program* (5.29)*. This problem can be further simplified, by eliminating the variable π_T, as follows:*

$$\begin{cases} \min\limits_{\substack{\mu_P\,,\,\mu_T\,\geq\,0\\ \text{any }\pi_P}} & \displaystyle\sum_{i=1}^{N_P}\left\langle Z_P^i\check{z}_P^i + Q_P^i\check{q}_P^i - b_P^i, \mu_P^i\right\rangle + \sum_{j=1}^{N_T}\left\langle Z_T^j\check{z}_T^j + Q_T^j\check{q}_T^j - b_T^j, \mu_T^j\right\rangle \\ \text{s.t.} & Z_P^{i\,\top}\mu_P^i = {I^i}_P'(\check{z}_P^i)\,, \quad Z_T^{j\,\top}\mu_T^j = {I^j}_T'(\check{z}_T^j) \\ & Q_P^{i\,\top}\mu_P^i + S_P^{i\,\top}\pi_P = {c^i}_P'(\check{q}_P^i) \\ & Q_T^{j\,\top}\mu_T^j - B_T^{j\,\top}\pi_P + S_T^{j\top}\Big(\sum_{k=1}^{N_T} PS_T^k\check{q}_T^k + d_0\Big) \\ & \qquad\qquad = {c^j}_T'(\check{q}_T^j) - \delta^j S_T^{j\top}P^\top S_T^j\check{q}_T^j\,. \end{cases}$$

5.4 The European gas network

Consider a market with a third kind of player, called *outsourcer*, in charge of modifying or transporting the product before the traders supply it to the end consumers.

In the network, producers only deal with traders. Therefore, they solve the problems (5.1), which are nothing but problems (2.21) in the current setting.

Traders now deal also with the outsourcer players, who charge a unitary price π_O for their activity. The exchange between the trader and the oursourcer player involves transformation of the product, represented by matrices $S^j_{T\to O}$,

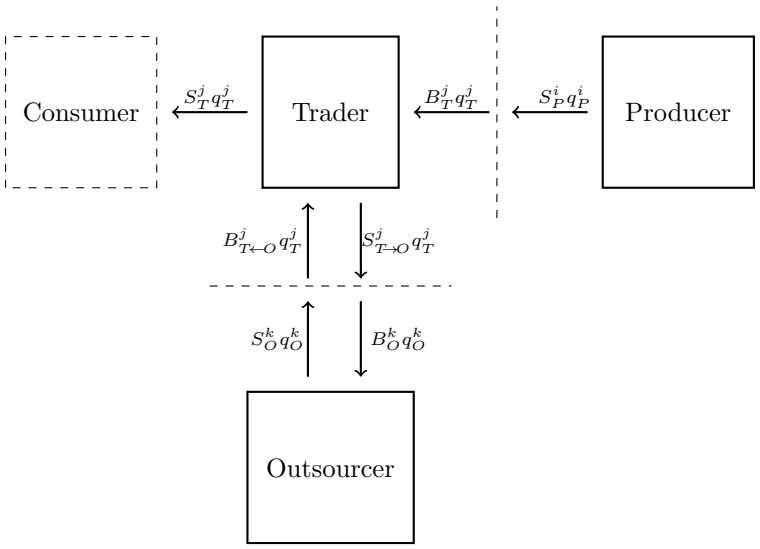

Figure 5.1: Market flow.

$B^j_{T\leftarrow O}$, S_O^k, B_O^k as schematically represented in Figure 5.1, with the product flow. The jth trader problem (5.2) is modified accordingly:

$$\begin{cases} \max & \left\langle S_T^j q_T^j, \pi_T \right\rangle - \left\langle B_T^j q_T^j, \pi_P \right\rangle - \left\langle S^j_{T\to O} q_T^j, \pi_O \right\rangle - c_T^j(q_T^j) - I_T^j(z_T^j) \\ \text{s.t.} & (z_T^j, q_T^j) \in X_T^j \,. \end{cases}$$

As for the outsourcing players, denoting once more the investment-operational decision variables of the kth agent by (z_O^k, q_O^k) and similarly for the costs and feasible set, the corresponding maximization problem is

$$\begin{cases} \max & \left\langle B_O^k q_O^k, \pi_O \right\rangle - c_O^k(q_O^k) - I_O^k(z_O^k) \\ \text{s.t.} & (z_O^k, q_O^k) \in X_O^k \,. \end{cases} \tag{5.30}$$

Here, the Outsourcer charges the Trader for processing $B_O^k q_O^k$.

To clear the market, in addition to (5.3) and (5.4), the exchange between traders and outsourcing players should be balanced and, hence,

$$\sum_{j=1}^{N_T} S^j_{T\to O} \bar{q}_T^j - \sum_{k=1}^{N_O} B_O^k \bar{q}_O^k = 0 \,. \tag{5.31}$$

The additional balance $\sum_{k=1}^{N_O} S_O^k \bar{q}_O^k - \sum_{j=1}^{N_T} B^j_{T\leftarrow O} \bar{q}_T^j = 0$, is omitted, because it is often automatic from (5.31).

We are once again in the considered framework (which is a particularization of Subsection 2.4 to affine shared constraints).

5.4.1 Numerical assessment

We consider the full European gas network described in [21]. Its structure is outlined in Figure 5.2. For a network covering 54 countries and 36 markets, the market has 7 types of players representing producers, traders, and 5 different outsourcing activities. Specifically, there are 28 producers, 22 traders, 10 liquefiers, 15 regasifiers, 22 storage operators, 74 pipeline operators, and 36 marketers.

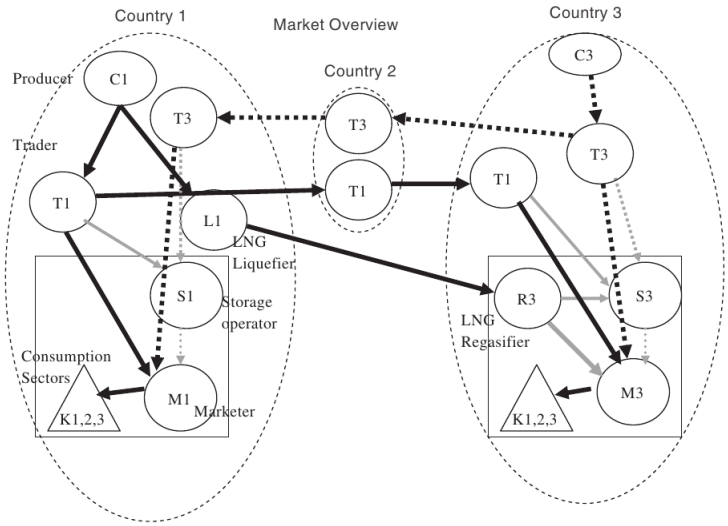

Figure 5.2: European Gas Network (as described in [21])

To illustrate the analysis that can be derived from the models presented above, we coded them in Matlab (R2012a), using PATH [18, 28] to solve the variational problems. The runs were done on a PC operating under Ubuntu 12.04-64 bit with a processor Intel Atom 1.80GHz $\times$ 4 and 2GB of memory.

The data in [21] yields for the game a problem with 4620 variables and 488 constraints. We solved the equilibrium problem of the implicit model, with and without market power. In the first instance, the trader's problem $(5.2)_{\delta^j}$ has $\delta^j \equiv 0$. In the second, $\delta^j = 0.75$ for Russia, Norway, the Netherlands, and Algeria; and $\delta^j = 0.25$ for the Caspian Sea, Denmark, and the UK.

To ensure that the implementation is error-free, we first ran both formulations, that is the game (5.23),(5.27),(5.28) and the MCP (5.15), and checked whether the corresponding output was alike. Table 5.2 summarizes the results. The ∞-norms of the differences of the primal solutions obtained with both approaches were very small in all the cases. We observed larger differences in the dual components, in percentages ranging up to 6% (for the competitive case, without market power). However, this is still an insignificant difference in this context, which allows us to conclude that the output of both formulations is indeed "the same" and the implementations are correct. An interesting infor-

Table 5.2: Output for the implicit model

Formulation	Market Power?	PATH Residual	CPU (seconds)
Game	no	2E-08	36.7
MCP	no	7E-08	47.3
Game	yes	7E-11	77.1
MCP	yes	2.5E-11	201.5

mation in Table 5.2 is the CPU times. In general PATH was very fast, but solving time increased significantly for the MCP formulation when there is market power. At this point, one could ask why this increase is of importance, given that the solution times were still within some minutes. The answer is that this increase, still significant in percentage terms, would blow up once stochasticity is introduced to the model. According to PATH final convergence report, when there is market power, the solver needed much more inner iterations to converge. We observed that when decreasing the solver precision from 10^{-8} to 10^{-6}, both formulations were again solved in about 70 seconds.

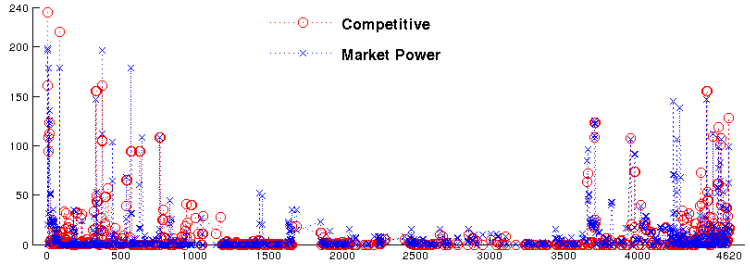

Figure 5.3: Comparison of the primal output

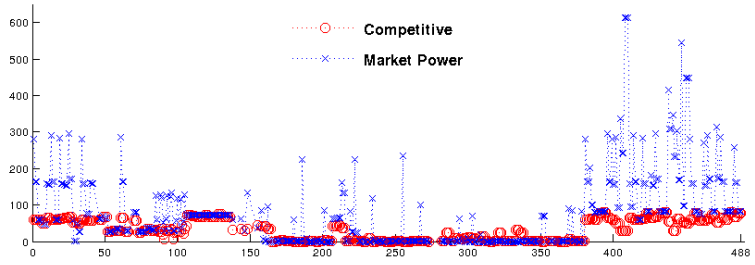

Figure 5.4: Comparison of the dual output

The primal and dual output with and without market power is plotted in Figures 5.3 and 5.4, respectively. The impact of market power is especially noticeable in the dual variables, corresponding to prices: in Figure 5.4 the red circles (competitive prices) are systematically lower than the blue crosses (market power).

5.5 Equilibrium for stochastic models

Realistic models for the energy industry often include *uncertainty*: for instance in (5.5), the actual electrical load may deviate from the predicted one due to random variations of temperature, switch off/on of local consumers, or daylight. Similarly in (5.1), for the generation costs $c_P^i(\cdot)$ or the available resources defining the feasible sets X_P^i. To reflect such variations, a stochastic model of uncertainty must be built and the risk-averse decision process must be put in a suitable setting.

5.5.1 Hedging risk: The setting

Consider the probability space defined by a measure $\mathbb{P}$ on a sample space Ω equipped with a sigma-algebra $\mathcal{F}$. Decision variables are now random functions in the space $L_p(\Omega, \mathcal{F}, \mathbb{P})$ for $p \in [1, +\infty)$, with dual $L_q(\Omega, \mathcal{F}, \mathbb{P})$ for $q \in (1, +\infty]$ such that $1/p + 1/q = 1$. We sometimes use the shorter notation L_p and L_q for these spaces, which are paired by the duality product, see [69], [17],

$$\langle x^*, x \rangle_{\mathbb{P}} = \int_\omega \langle x^*(\omega), x(\omega) \rangle \, d\mathbb{P}(\omega) \, .$$

In the presence of uncertainty, a natural reaction of agents in the market is to hedge against undesirable events. For the ith producer, in particular, aversion to volatility is expressed by a coherent risk measure $\rho^i(\cdot)$, assumed to be a proper function, as in [17, Chapter 6]. One possibility in the space $L_1(\Omega, \mathcal{F}, \mathbb{P})$ is to take the *Average Value-at-Risk* of level $1 - \varepsilon_i$, a recent renaming of the *Conditional Value-at-Risk* [66]. Namely, given a confidence level $0 < 1 - \varepsilon < 1$, if the random outcome $X \in L_p$ represents a loss (lower values are preferred), the measure is given by the expression

$$AV@R_{\,\varepsilon}(X) := \min_u \left\{ u + \frac{1}{1-\varepsilon} \mathbb{E}[X(\omega) - u]^+ \right\},$$

where $[\cdot]^+ := \max\{0, \cdot\}$ is the positive-part function and $\mathbb{E}(\cdot)$ denotes the expected-value function taken with respect to $d\mathbb{P}$.

It can also be useful to consider the more general risk measures of the form

$$\rho^i(X) := (1 - \kappa_i)\mathbb{E}(X) + \kappa_i \, AV@R_{\,\varepsilon_i}(X) \tag{5.32}$$

for the given risk-aversion parameter $\kappa_i \in [0, 1]$ and confidence level $1 - \varepsilon_i \in [0, 1)$. When the random outcome represents a reward the definition changes to $\varrho^i(X) := \rho^i(-X)$.

It is shown in [17, Theorem 6.4] that any proper coherent risk measure is in fact the support function of the domain of its conjugate; see also [67], [60]. In particular, for (5.32) we have the dual representation

$$\rho^i(X) = \sup_{x^* \in X^*} \langle x^*, X \rangle_{\mathbb{P}}, \text{ where} \tag{5.33}$$

$$X^* := \left\{ x^* \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}) : \begin{array}{l} 1 - \kappa_i \le x^*(\omega) \le 1 - \kappa_i + \kappa_i/\varepsilon_i \quad a.e.\ \omega \in \Omega \\ \mathbb{E}(x^*) = 1 \end{array} \right\},$$

see [17, Theorem 6.4, and (6.69) in Ex. 6.16].

In what follows, the random vectors q belong to the space $L_p(\omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$, that is $q(\omega) \in \mathbb{R}^m$ for all $\omega \in \Omega$. We start by stating a technical result that will be used to define the concept of stochastic variational equilibria and interpret game formulations in a stochastic complementarity framework.

Proposition 5.5.1. *Let ρ be a risk measure in the family* (5.33)*, $I : \mathbb{R}^n \to \mathbb{R}$ a smooth convex function, $f : \mathbb{R}^{n+m} \times \Omega \to \mathbb{R}$ a random finite-valued lower semicontinuous function, convex for almost every $\omega \in \Omega$. For nonempty closed convex sets $X(\omega) \subset \mathbb{R}^{n+m}$, a matrix S and a random vector $D_0(\omega)$ of suitable dimensions, consider the problem*

$$\left\{ \begin{array}{ll} \min & I(z) + \rho(f(z, q(\omega), \omega)) \\ \text{s.t.} & z \in \mathbb{R}^n, q \in L_1(\Omega, \mathcal{F}, \mathbb{P}), \\ & (z, q(\omega)) \in X(\omega) \quad \text{a.e. } \omega \in \Omega, \\ & Sq(\omega) = D_0(\omega) \quad \text{a.e. } \omega \in \Omega\,. \end{array} \right. \tag{5.34}$$

Suppose (5.34) *satisfies some appropriate constraint qualification condition.*

A necessary and sufficient condition for $(\bar{z}, \bar{q})$ to solve (5.34) *is the following:*

$$\left\{ \begin{array}{l} \exists\, \bar{x}^* \in X^* \text{ and } \bar{\pi} \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}) \text{ such that} \\ \bar{x}^* \text{ maximizes } \displaystyle\int_\Omega x^*(\omega) f(\bar{z}, \bar{q}(\omega), \omega) d\mathbb{P}(\omega) \text{ for } x^* \in X^*, \\ \text{and} \\ 0 = I'(\bar{z}) + \displaystyle\int_\Omega \Big(\nabla_z f(\bar{z}, \bar{q}(\omega), \omega) + \nu_z(\omega) \Big) \bar{x}^*(\omega) d\mathbb{P}(\omega) \\ 0 = \nabla_q f(\bar{z}, \bar{q}(\omega), \omega) - S^\top \bar{\pi}(\omega) + \nu_q(\omega) \end{array} \right. \quad \text{a.e. } \omega \in \Omega\,,$$

where we use the notation $(\nu_z(\omega), \nu_q(\omega)) \in \mathcal{N}_{X(\omega)}(\bar{z}, \bar{q}(\omega))$ for the normal elements.

Proof. By Propositions 6.5 and 6.7 in [17], the considered risk measure is continuous, sub-differentiable, and finite-valued. By [17, Theorem 6.6], the set X^* is convex, bounded and weakly* closed and, in view of (5.33), the second term in the objective function of (5.34) is finite-valued. Furthermore, this term is also convex, because risk measures are monotone and the composition with an increasing convex function preserves convexity; [36, Proposition IV.2.1.8].

By Proposition 6.37 in [17], (5.34) has the equivalent expression

$$\min_{z \in \mathbb{R}^n} \Big(I(z) + \rho(F(z, \omega)) \Big) \tag{5.35}$$

for the second-stage value function

$$F(z,\omega) := \left\{ \begin{array}{ll} \min & f(z,q,\omega) \\ \text{s.t.} & q \in \mathbb{R}^m, \\ & (z,q) \in X(\omega) \text{ and } Sq = D_0(\omega) \quad \text{a.e. } \omega \in \Omega\,. \end{array} \right. \tag{5.36}$$

Our assumptions ensure that this value function is a random lower semicontinuous function, convex in z almost everywhere. By Proposition 6.32 in [17], a necessary and sufficient condition for $(\bar{z},\bar{q})$ to solve (5.35) is that there exists $\bar{x}^* \in X^*$ such that $\bar{x}^* \in \partial\rho(\bar{F})$ for $\bar{F} = F(\bar{z},\cdot)$ and

$$0 \in \mathbb{E}_*(\partial f_\omega(\bar{z})) + I'(\bar{z})\,, \tag{5.37}$$

where $\mathbb{E}_*(X) := \langle \bar{x}^*, X\rangle_{\mathbb{P}}$ denotes the expectation taken with respect to $\bar{x}^* d\mathbb{P}$. The inclusion (5.37) means that there exists an integrable random vector $\eta(\omega) \in \partial f_\omega(\bar{z})$ such that

$$0 = \mathbb{E}_*(\eta) + I'(\bar{z})\,,$$

where we defined $f_\omega(\cdot) := F(\cdot,\omega)$. To interpret these relations in our setting, we first note that by the identity (6.43) in [17], the subgradient relation amounts to requiring that $\bar{x}^*$ maximizes $\langle x^*, F(\bar{z})\rangle_{\mathbb{P}}$ over X^*. As for (5.37), the marginal function $f_\omega(z) = F(z,\omega)$ has a subdifferential depending on the minimand in (5.36). More precisely, recalling that i_S denotes the indicator function of a set S, and setting

$$g(z,q,\omega) := f(z,q,\omega) + i_{X(\omega)}(z,q) + i_{\{q:Sq=D_0(\omega)\}}(q),$$

yields, by [36, Corollary VI.4.5.3], that

$$\partial f_\omega(\bar{z}) = \{s : (s,0) \in \partial_{z,q} g(\bar{z}, q_{\bar{z},\omega}, \omega)\} \text{ for any minimizer } q_{\bar{z},\omega} \text{ in (5.36).}$$

Since $\partial i_{\{q:Sq=D_0(\omega)\}}(q) = lin(S^\top)$, there exist random vectors $(\nu_z(\omega), \nu_q(\omega)) \in \mathcal{N}_{X(\omega)}(\bar{z},\bar{q}(\omega))$ with $\nu_z(\omega)$ integrable such that

$$(\eta(\omega),0) = (\nabla_z f(\bar{z},\bar{q}(\omega),\omega), \nabla_q f(\bar{z},\bar{q}(\omega),\omega)) + (0, -S^\top\bar{\pi}(\omega)) + (\nu_z(\omega),\nu_q(\omega))$$

from where we obtain the stated condition, and the result follows. □

The following corollary is straightforward.

Corollary 5.5.2. *Any solution to* (5.34) *in Proposition 5.5.1 solves the relaxed problem*

$$\left\{ \begin{array}{ll} \min & I(z) + \rho(f(z,q(\omega),\omega)) - \langle Sq(\omega), \bar{\pi}(\omega)\rangle_{\mathbb{P}} \\ \text{s.t.} & z \in \mathbb{R}^n\,, q \in L_1(\Omega,\mathcal{F},\mathbb{P})\,, \\ & (z,q(\omega)) \in X(\omega) \quad \text{a.e. } \omega \in \Omega\,. \end{array} \right. \quad \square \tag{5.38}$$

We shall use Corollary 5.5.2 in Definition 5.5.3 below to introduce the concept of *stochastic variational equilibrium*.

5.5.2 Risk-averse games

For convenience, we now make three simplifying assumptions:

- The stochastic counterparts of the producers' and traders' problems are set in a two-stage framework. For example, in (5.23) the "investment" variables z_P^i are of the "here-and-now" type, to be decided before the uncertainty realizes. By contrast the "generation" variables q_P^i are of type "wait-and-see": they are decided at a second stage, once ω becomes known, so q_P^i depends on ω.
- The concept of stochastic equilibrium and its connections with a game formulation is examined for a market without traders. Since there are only producers, in (5.23) and throughout all subindices P are dropped. Also, the market clearing relation (5.3) disappears, only a stochastic variant of (5.4) is in order.
- Finally, to ease the writing and without loss of generality, uncertainty comes from the generation costs c^i and the demand in (5.4) (the feasible sets X^i are assumed deterministic).

In this context, given a price cap $PC \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$, the complementarity formulation of stochastic equilibrium with risk aversion in [23] is

Find $(\bar{z}^i \in \mathbb{R}^{nz^i}, \bar{q}^i \in L_1, \bar{q}^0 \in L_1, \bar{\pi} \in L_\infty)$ such that

$$\textbf{Risk-averse producers} \quad \begin{cases} \min & I^i(z^i) + \rho^i\Big(c^i(q^i(\omega),\omega) - \big\langle \pi(\omega), S^i q^i(\omega)\big\rangle\Big) \\ \text{s.t.} & Z^i z^i + Q^i q^i(\omega) \geq b^i \quad \text{a.e. } \omega \in \Omega \end{cases} \tag{5.39}$$

$$\textbf{Coupling constraints} \quad \sum_{i=1}^N S^i q^i(\omega) + q^0(\omega) = D(\omega) \quad \text{a.e. } \omega \in \Omega \;\; (\text{mult.}\pi(\omega))$$

$$\textbf{Price cap} \quad 0 \leq q^0(\omega) \perp PC(\omega) - \pi(\omega) \geq 0 \quad \text{a.e. } \omega \in \Omega. \tag{5.40}$$

When compared to (5.1), the producers' problem is now set as a minimization, because the risk averse measure controls losses and not incomes. The objective function in (5.39) is in fact equivalent to the one considered in [23], taking into account that the investment functions I^i and the first-stage variables z^i are deterministic, recalling that risk measures are equivariant to translations.

In order to compare the risk-averse MCP above with a game formulation, we consider the following stochastic game, depending on given $N+1$ risk-averse functions $\rho^0, \rho^1, \ldots, \rho^N$ of the form (5.32)-(5.33):

$$\textbf{Risk-averse producers} \quad \begin{cases} \min & I^i(z^i) + \rho^i\Big(c^i(q^i(\omega),\omega)\Big) \\ \text{s.t.} & Z^i z^i + Q^i q^i(\omega) \geq b^i \quad \text{a.e. } \omega \in \Omega \end{cases} \tag{5.41}$$

$$\textbf{Risk-averse player representing consumers} \quad \begin{cases} \min & \rho^0\Big(\big\langle PC(\omega), q^0(\omega) - D(\omega)\big\rangle\Big) \\ \text{s.t.} & q^0(\omega) \geq 0 \quad \text{a.e. } \omega \in \Omega \end{cases} \tag{5.42}$$

$$\textbf{Coupling constraints} \quad \sum_{i=1}^{N} S^i q^i(\omega) + q^0(\omega) = D(\omega) \quad \text{a.e. } \omega \in \Omega. \tag{5.43}$$

For such a stochastic game, we now define the concept of variational equilibrium. Recall that when dealing with GNEP in the deterministic setting, we were not interested in arbitrary Nash equilibria, but rather in the game Variational Equilibria, defined as solutions to VIs derived from the game. In a general stochastic context like the one under consideration, instead of deriving an explicit VI and motivated by Corollary 5.5.2, we characterize the VE through the Lagrange multipliers of the coupling constraint of the game.

Definition 5.5.3 (Stochastic VE)**.** *For a stochastic GNEP like* (5.41)-(5.43)*, the point* $\Big((\bar{z}^i)_{i=1}^N, (\bar{q}^i)_{i=1}^N, \bar{q}^0\Big)$ *is a* variational equilibrium *if there exists a Lagrange multiplier* $\bar{\pi} \in L_\infty$ *associated to the coupling constraint, the same for all the players, such that* $\Big((\bar{z}^i)_{i=1}^N, (\bar{q}^i)_{i=1}^N, \bar{q}^0\Big)$ *still solves the agents' problems after relaxing the coupling constraints (as in Corollary 5.5.2).*

We now are in a position to find the equivalent MCP counterpart of our risk-averse game.

Theorem 5.5.4 (MCP formulation for the risk-averse game)**.** *Suppose the risk-averse GNEP* (5.41)-(5.43) *has a variational equilibrium* $\bar{\mathfrak{p}} := \Big((\bar{z}^i)_{i=1}^N, (\bar{q}^i)_{i=1}^N, \bar{q}^0\Big)$*, with* $(\bar{\mu}^i)_{i=1}^N$ *and* $\bar{\pi}$ *being the* L_∞*-multipliers for the constraints in* (5.41) *and the coupling constraints* (5.43)*, respectively.*

Then the primal-dual pair $(\bar{\mathfrak{p}}, \bar{\mathfrak{d}})$ *with* $\bar{\mathfrak{d}} := (\bar{\mu}, \bar{\pi})$ *solves the risk-averse MCP derived from*

$$\textbf{Risk-averse producers} \quad \begin{cases} \min & I^i(z^i) + \rho^i\Big(c^i(q^i(\omega), \omega)\Big) - \big\langle S^{i\,\top}\pi, q^i\big\rangle_{\mathbb{P}} \\ \text{s.t.} & Z^i z^i + Q^i q^i(\omega) \geq b^i \quad \text{a.e. } \omega \in \Omega \end{cases} \tag{5.44}$$

Coupling constraints *as in* (5.43)

Risk-averse Price cap $0 \leq q^0(\omega) \perp x_0^*(\omega) PC(\omega) - \pi(\omega) \geq 0 \quad \text{a.e. } \omega \in \Omega$

for x_0^* *solving*

$$\begin{cases} \min & \displaystyle\sum_{i=1}^{N} \int_\Omega \big\langle x_0^*(\omega) PC(\omega), S^i \bar{q}^i(\omega)\big\rangle \, d\mathbb{P}(\omega) \\ \text{s.t.} & \mathbb{E}(x_0^*) = 1 \\ & 1 - \kappa_0 \leq x_0^*(\omega) \leq 1 - \kappa_0 + \dfrac{\kappa_0}{\varepsilon_0} \quad \text{a.e. } \omega \in \Omega. \end{cases} \tag{5.45}$$

Proof. To derive a complementarity formulation, we write the ith risk-averse

problem of the producers in the form

$$\begin{cases} \min & I^i(z^i) + \rho^i\Big(c^i(q^i(\omega),\omega)\Big) \\ \text{s.t.} & (z^i, q^i(\omega)) \in X^i \quad \text{a.e. } \omega \in \Omega \\ & S^i q^i(\omega) = D_0(\omega) := D(\omega) - \sum_{i\neq j} S^j q^j(\omega) - q^0(\omega)\,. \end{cases}$$

Proposition 5.5.1 applies with $f(z,q,\omega) = c^i(q^i,\omega)$ and, hence (dropping the subindices i), there exists $\bar{x}^* \in X^*$ maximizing $\langle x^*, c(\bar{q}(\omega),\omega)\rangle_{\mathbb{P}}$ over X^* and $\bar{\pi} \in L_\infty$ such that

$$\begin{cases} 0 = I'(\bar{z}) + \mathbb{E}_*(\nu_z) \\ 0 = c'(\bar{q}(\omega),\omega) - S^\top \bar{\pi}(\omega) + \nu_q(\omega) \quad \text{a.e. } \omega \in \Omega. \end{cases}$$

These are the optimality conditions for (5.44). In the case of comsumers' representative (5.42), by Definition 5.5.3, $\bar{q}^0$ solves the relaxed problem

$$\begin{cases} \min & \rho^0\Big(\big\langle PC(\omega), q^0(\omega) - D(\omega)\big\rangle\Big) - \Big\langle \big\langle q^0(\omega), \bar{\pi}(\omega)\big\rangle \Big\rangle_{\mathbb{P}} \\ \text{s.t.} & q^0(\omega)) \geq 0\,. \end{cases} \tag{5.46}$$

Consider the affine operator $A : L_1 \to L_1$ defined by

$$[A(q^0)](\omega) := \big\langle PC(\omega), q^0(\omega) - D(\omega)\big\rangle\ .$$

Since $PC \in L_\infty$ and $D \in L_1$, we have that A is continuous and, hence, the optimality condition for (5.46) is

$$0 \in \partial\big(\rho^0 \circ A + i_{\geq 0}\big)(\bar{q}^0) - \bar{\pi} = \partial(\rho^0 \circ A)(\bar{q}^0) + \mathcal{N}_{\geq 0}(\bar{q}^0) - \bar{\pi}$$

By the definition of the normal cone above, there exist $g \in \partial(\rho^0 \circ A)(\bar{q}^0)$ and $\nu \in L_\infty$ such that

$$0 \leq \bar{q}^0(\omega) \perp -\bar{\nu}(\omega) \geq 0 \quad \text{a.e. } \omega \in \Omega$$

and

$$0 = g(\omega) + \bar{\nu}(\omega) - \bar{\pi}(\omega) \quad \text{a.e. } \omega \in \Omega\,.$$

To get an explicit expression for g above, we need to compute the subdifferential $\partial(\rho^0 \circ A)(\bar{q}^0)$, recalling that the mapping A is affine and continuous, and the risk measure is increasing and finite-valued. Therefore, by [86, Thm.2.83],

$$g \in \partial(\rho^0 \circ A)(\bar{q}^0) \iff g(\omega) = PC(\omega)s(\omega) \quad \text{for } s \in \partial\rho^0\Big(A(\bar{q}^0)\Big)\,.$$

The definitions of the subdifferential and of the conjugate function give the equivalence $s \in \partial\rho^0(A(\bar{q}^0)) \iff A(\bar{q}^0) \in \partial\rho^{0*}(s)$. By the dual representation (5.33), the conjugate of ρ^0 is the indicator function of the (convex and bounded) dual set X^*, that is $\rho^0 = i^*_{X^*}$. Then, $\rho^{0*} = i^{**}_{X^*} = i_{X^*}$. Since the subdifferential of the indicator function of a closed convex set is the normal cone of the set, by the definition of the normal cone, the subgradient $g \in \partial(\rho^0 \circ A)(\bar{q}^0)$ has components $g(\omega) = PC(\omega)s(\omega)$ for $s \in X^*$ satisfying $\big\langle A(\bar{q}^0), x^* - s\big\rangle_{\mathbb{P}} \leq 0$ for all $x^* \in X^*$. So s maximizes $\big\langle A(\bar{q}^0), x^*\big\rangle_{\mathbb{P}}$ over X^*, and in view of (5.43), $s = \bar{x}^*_0$ from (5.45). The risk-averse price cap condition follows from plugging $g(\omega) = PC(\omega)\bar{x}^*_0(\omega)$ in the optimality condition. □

Theorem 5.5.4 shows that, like in the deterministic framework, the stochastic game is equivalent to a complementarity model with risk aversion. The MCP model is *not of the form* (5.39), where agents *hedge individually their profit*. Instead, a variational equilibrium for the game (5.41)-(5.43) gives a stochastic equilibrium for a market that is cleared because (5.43) is satisfied, and where the risk-averse producers are *remunerated a price that is controlled by a risk-averse price cap*.

In the game, aversion to risk is peculiar in the sense that producers hedge volatility by controlling only variations in the generation costs. Compare the game problem (5.44), where the remuneration is taken in mean without hedging risk, with the MCP (5.39)-(5.40) where each producer tries to control the risk in their individual revenue. In the game the control of volatile prices is "delegated" to some higher instance. This is the same instance that caps the remunerations, only that now the cap is chosen adaptively, in a manner that is optimal for the market, in the sense of (5.45). By contrast, in the risk-averse MCP, the instance limiting prices only takes into account stochasticity but does not perceive the fact of capping prices as a risky action on the market.

We show below that the three models turn equivalent when all the agents in the market are risk-neutral.

Corollary 5.5.5 (Equivalence for risk-neutral agents)**.** *Suppose that for all the agents* $\rho^i = \mathbb{E}$*, the expected-value function. Then finding a variational equilibrium for the GNEP* (5.41),(5.42)-(5.43) *is equivalent to solving the MCP* (5.39)-(5.40) *which is in turn equivalent to the MCP* (5.44)-(5.45)*.*

Proof. Straightforward from Theorem 5.5.4, noting that the expected-value function is recovered by setting $\kappa_i = 0$ in (5.32), with a singleton dual set $X^* = \{x^* \equiv 1\}$ in (5.33). In particular, a risk-neutral representative of the consumers can only take $\bar{x}_0^* \equiv 1$, which yields the stochastic price cap from (5.40). The equivalence with the last MCP results from the linearity of the expected-value function. □

Some final comments and observations are now in order:

- Instead of handling uncertainty in two stages, a multistage setting can also be of interest. In a multi-stage modeling, for an optimization time horizon with T steps and a discrete-time stochastic process ω in a probability space $(\Omega, \{\mathcal{F}_t\}, \mathbb{P})$, information becomes progressively more uncertain as t increases from 1 to T: $\mathcal{F}_t$ is the sigma-algebra $\mathcal{F}_t := \sigma(\omega_j, j \leq t)$ and Ω is a sample space equipped with the filtration $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_T$. The sequence of decisions also forms a stochastic process that is non-anticipative (or adapted to the filtration of sigma-algebras): at time t decisions are taken on the basis of knowledge available up to time t. When compared to a two-stage modeling, the multistage setting is more accurate regarding uncertainty representation. This, keeping in mind that multistage risk aversion is a delicate subject, for which time-consistent models should be set. Last but not least, and as discussed in [73, Section 5], risk-averse

variants of sampling approaches like [59] and [74] lack of implementable stopping criteria. Multistage risk-averse models present numerous challenges already in an optimization framework, we refer to [70] for more details.

- Like it has been done in the deterministic case in Section 5.4.1, it would be interesting to analyze and compare the performances of the risk-averse game versus the risk-averse MCP on a numerical example. However, due to the positive-part function in (5.32), risk measures are not differentiable and for both models the generalized equation has a mapping F with multi-valued components. In this context, a direct application of a solver like PATH is no longer possible (and there is currently no other established software that can do the job). In [23], the MCP (5.39)-(5.40) is "solved" ignoring nondifferentiability issues and treating the mapping as if it were single-valued. This heuristic seems to produce sound results for the considered example, but cannot be regarded as a reliable solution method, of course. In order to handle nonsmoothness, some special technique should be used, for example the approximation procedure in [52]. This is the subject of the next, final, chapter.

Chapter 6

An Approximation Scheme for a Class of Stochastic Generalized Nash Equilibrium Problems

We shall mostly follow the approach of [52]. Like in Subsection 5.5.2, we consider a risk-averse stochastic counterpart of the GNEP described in Subsection 2.4, only that now, for numerical purposes, the probability space is finite. Specifically, we assume we are given an (approximate) discrete multivariate probability distribution of the game uncertainty (the demand D in (5.5), the costs $c^i(\cdot)$ in (5.23), etc), with support given by finitely many realizations, say K. To each realization $\omega_k = (D_k, c^i(\cdot;\omega_k), \ldots)$ corresponds a probability $p_k > 0$ for $k = 1, \ldots, K$. With this approximation, instead of functions $q^i(\omega) \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \Re^{m_i})$, the ith producer second-stage variables q^i are vectorial strategies concatenating the decisions $q^i_k := q^i(\omega_k) \in \Re^{m_i}$ for each scenario: $q^i = (q^i_1, q^i_2, \ldots, q^i_K)$. As for the constraints, we keep as before deterministic technological and resource sets X^i. By contrast, rather than having affine coupling constraints like in the previous chapter, we now return to the GNEP setting described in Subsection 2.4, and consider the shared constraints:

$$\sum_{i=1}^{N} [h^i_k(z^i) + g^i_k(q^i_k)] \leq 0 \quad \text{for } k = 1, \ldots, K\,, \tag{6.1}$$

where the functions $h^i_k(\cdot) := h(\cdot, \omega_k)$ and $g^i_k(\cdot) := g^i(\cdot, \omega_k)$ are convex for each realization ω_k. For fixed $\bar{q}$, the shorthand notation of the random cost will be used throughout:

$$c^i_\omega(q^i, \bar{q}^{-i}) := \Big(c^i(q^i_1, \bar{q}^{-i}_1; \omega_1), \ldots, c^i(q^i_K, \bar{q}^{-i}_K; \omega_K)\Big) \in \mathbb{R}^K\,.$$

6.1 Motivation

The work [33] proposes a modeling structure for risk-neutral stochastic VIs in which the expectation is estimated by simulation. In the setting of finding a variational equilibrium for the game, each player's problem is similar to (2.19), taking into account uncertainty:

$$\textbf{Risk-neutral problem} \quad \begin{cases} \min_{z^i,q^i} & I^i(z^i, z^{-i}) + \mathbb{E}\Big(c^i_\omega(q^i, q^{-i})\Big) \\ \text{s.t.} & \text{for } k = 1, \ldots, K \quad (z^i, q^i_k) \in X^i \text{ and} \\ & \sum_{j=1}^{N} [h^j_k(z^j) + g^j_k(q^j_k)] \leq 0. \end{cases} \tag{6.2}$$

Because all the involved functions are smooth and convex, and assuming that (6.1) satisfies some constraint qualification condition, the variational equilibrium is found by solving the

$$\text{VI}(F, \prod_{i=1}^{N} X^i \cap \mathcal{S}), \tag{6.3}$$

where $\mathcal{S}$ is the set of points satisfying the shared constraints (6.1) and the F-components of the operator are

$$F^i(z, q) := \nabla_{z^i} I^i(z) + \nabla_{q^i} \mathbb{E}\Big(c^i_\omega(q)\Big)$$

for $i = 1, \ldots, N$ and with $(z, q) = \{(z^j, q^j)\}_j$. Suppose we use a scalar product preserving the structure of $\Re^{n_i+m_i}$ as a product space, so that

$$\left\langle (z^i, q^i), (\tilde{z}^i, \tilde{q}^i) \right\rangle = \left\langle z^i, \tilde{z}^i \right\rangle + \left\langle q^i, \tilde{q}^i \right\rangle .$$

Then, in the writing of (2.14), and because the expectation and gradient operators commute, we look for a feasible point $(\bar{z}, \bar{q})$ such that

$$\left\langle \left\{ \nabla_{z^j} I^j(\bar{z}) \right\}_j, z - \bar{z} \right\rangle + \left\langle \mathbb{E}\Big(\left\{ \nabla_{q^j} c^j(\bar{q}) \right\}_j \Big), q - \bar{q} \right\rangle \geq 0 \tag{6.4}$$

for all $\{(z^j, q^j)\}_j \in \prod_{i=1}^{N} X^i \cap \mathcal{S}$.

The approximation in [33] is useful when the number of realizations K is extremely large (the sampling space can even be infinite; we present here a finite-dimensional version for simplicity). This is often the case in practice, because only a large number of realizations captures all the ingredients in the uncertainty. The idea is to define approximating (stochastic) VIs that have solutions with probability one and that in the limit solve (6.3). An example is to replace the probability space Ω by $\{\omega_1, \ldots, \omega_\ell\}$ at iteration ℓ and solve the associated VI. The work [33] also provides bounds for the closeness of the approximate solutions to solutions of the limit problem, in terms of quality of approximation.

The development below follows a similar path for risk-averse (and no longer risk-neutral) VIs. There is an important difference, though: in our problem

the VI under consideration is special in the sense that it comes from writing an equilibrium problem for a market with agents trying to optimize their activities. As such, the aversion to risk is set on the *optimization problems of the players*, and not on the VI operator. This is in contrast with some recent works on stochastic variational inequalities, such as [11] and [10] that endow with risk aversion the VI itself, and do not seem appropriate to the context of a special VI associated to a game. More precisely, these works employ the so-called D-gap function [27] associated to the stochastic VI to define a loss function that is then minimized, after composition with a risk measure. Since in our setting the stochastic VI has components

$$F^i(z, q; \omega) := \nabla_{z^i} I^i(z) + \nabla_{q^i(\omega)} c^i(q(\omega); \omega)\,,$$

the approach amounts to hedging risk on the *derivative* of the costs, instead of on the costs themselves. In this context, such an approach appears hard to interpret. As we do know the origin of our stochastic VI (this is not the case in the considered works), we exploit this knowledge and consider solution methods for a VI whose operator has components of the specific form

$$\nabla_{z^i} I^i(z) + \partial_{q^i(\omega)} \rho^i\Big(c^i(q(\omega); \omega)\Big)\,,$$

which, however, are multi-valued.

6.2 A stochastic approximation scheme

Given proper coherent risk measures $\rho^i : \mathbb{R}^K \to \mathbb{R}$, of the form (5.32), we are interested in finding a variational equilibrium for a game with risk-averse version of (6.2) given by

$$\textbf{Risk-averse problem} \quad \begin{cases} \min_{z^i, q^i} \quad I^i(z^i, z^{-i}) + \rho^i\Big(c^i_\omega(q^i, q^{-i})\Big) \\ \text{s.t.} \qquad \text{for } k = 1, \dots, K \quad (z^i, q^i_k) \in X^i \text{ and} \\ \qquad\qquad \sum_{j=1}^{N} [h^j_k(z^j) + g^j_k(q^j_k)] \leq 0. \end{cases} \tag{6.5}$$

In view of the analysis in Section 5.5, including the dual representation (5.33), the second term in the objective function is convex but not differentiable. So our GNEP with shared constraints yields a VI whose mapping is multi-valued (and whose full image at a given point would generally be hard to compute).

To overcome this difficulty, we construct a sequence of approximating GNEPs whose variational equilibria are easier to compute because they replace the risk measures by smooth differentiable functions, suitably chosen.

In order to present our approximation scheme, we first assume that for each player a sequence of approximating convex continuously differentiable functions $\rho^i_\ell : \mathbb{R}^K \to \mathbb{R}$ is given. We also assume that the sequence converges continuously to ρ^i (i.e., whenever a sequence $\{x_\ell\}$ converges to x, the sequence $\{\rho^i_\ell(x_\ell)\}$

converges to $\rho^i(x)$; see [68, Chapter 5]). We establish convergence of the approximating scheme in that case. We then show how to construct such approximations for the risk measures in (5.32), involving the $AV@R$, and assess the approximating procedure on a simple numerical example.

The approximation level will depend on an iteration index $\ell = 1, 2, \ldots$. At each iteration, the game has shared constraints (6.1) with

$$\textbf{ℓ-th Risk-averse problem} \quad \begin{cases} \min_{z^i, q^i} & I^i(z^i, z^{-i}) + \rho^i_\ell\Big(c^i_\omega(q^i, q^{-i})\Big) \\ \text{s.t.} & \text{for } k = 1, \ldots, K \quad (z^i, q^i_k) \in X^i \text{ and} \\ & \sum_{j=1}^{N} [h^j_k(z^j) + g^j_k(q^j_k)] \leq 0. \end{cases} \tag{6.6}$$

Since now the functions ρ^i_ℓ are differentiable, the GNEP variational equilibria (z_ℓ, q_ℓ) are solutions to

$$\mathrm{VI}(F_\ell, \prod_{i=1}^{N} X^i \cap \mathcal{S}) \text{ for } \mathcal{S} := \{(z, q) \text{ satisfying (6.1)}\}, \tag{6.7}$$

where the single-valued VI operator has components

$$F^i_\ell(z, q) := \Big(\nabla_{z^i} I^i(z), \nabla_{q^i} \rho^i_\ell\Big(c^i_\omega(q^i, q^{-i})\Big)\Big) .$$

The following approximation property is related to [33, Theorem 2], stated for general random objective functions in a more general setting. In particular, the theorem requires certain *coherent orientation* condition that is equivalent to strong regularity. In our finite-dimensional setting, we assume that some constraint qualification [77] holds for the whole set of constraints, while an individual CQ is required for the coupling constraint.

Theorem 6.2.1. *For each player $i = 1, \ldots, N$, let $\{\rho^i_\ell : \mathbb{R}^K \to \mathbb{R}\}$ be a sequence of convex continuously differentiable functions converging continuously to ρ^i as $\ell \to \infty$. If the sequence of solutions $\{(z_\ell, q_\ell)\}$ of (6.7) is bounded the following hold:*

1. *If the coupling constraints (6.1) satisfy the Slater condition (i.e., there exists a feasible point $(z_\circ, q_\circ)$ such that the inequalities in (6.1) hold strictly), then the sequence of Lagrange multipliers $\{\pi_\ell\}$ is bounded.*

2. *If the sequence $\{\pi_\ell\}$ is bounded, then for every accumulation point $(\bar{z}, \bar{q}, \bar{\pi})$ of the sequence $\{(z_\ell, q_\ell, \pi_\ell)\}$, its primal part is a variational equilibrium for the GNEP corresponding to (6.5), and $\bar{\pi}$ is a Lagrange multiplier associated to the coupling constraints (6.1).*

Proof. Problem (6.7) has shared constraint (6.1). After relaxing this constraint using an associated Lagrange multiplier $\pi_{k,\ell}$, and taking into account the decomposable structure of the set $\prod_{i=1}^{N} X^i$, we have that the ith components of a

solution (z_ℓ, q_ℓ) also solve the ith relaxed variational inequality

$$\mathrm{VI}\Big(F^i_\ell(z^i,q^i) + \sum_{k=1}^{K}[G^i_k(z^i,q^i)]^\top \pi_{\ell,k}, X^i\Big) \tag{6.8}$$

$$\begin{aligned} \text{for } F^i_\ell(z^i,q^i) &:= \nabla_{z^i} I^i(z^i, z_\ell^{-i}) + \nabla_{q^i}\rho^i_\ell\Big(c^i_\omega(q^i, q_\ell^{-i})\Big) \\ \text{and } G^i_k(z^i,q^i) &:= \nabla_{(z^i,q^i)}[h^i_k(z^i) + g^i_k(q^i_k)]; \end{aligned}$$

together with the complementarity conditions:

$$0 \le \pi_{\ell,k} \perp \sum_{i=1}^{N}[h^i_k(z^i_\ell) + g^i_k(q^i_{\ell,k})] \le 0 \quad \text{for } k = 1,2,\ldots,K. \tag{6.9}$$

We first show item (a). Using the Slater point $(z_\circ, q_\circ)$ as a feasible point in the relaxed VI (6.8) gives, for each i, the inequalities

$$\left\langle F^i_\ell(z^i_\ell,q^i_\ell) + \sum_{k=1}^{K}[G^i_k(z^i_\ell,q^i_\ell)]^\top \pi_{\ell,k}, (z^i_\circ - z^i_\ell, q^i_\circ - q^i_\ell)\right\rangle \ge 0\,.$$

Then, using also the convexity of I^i and of ρ^i_ℓ, we obtain that

$$\begin{aligned} &\sum_{i=1}^{N}\left[[I^i(z^i_\circ, z_\ell^{-i}) + \rho^i_\ell(c^i_{\omega_k}(q^i_\circ, q_\ell^{-i}))] - [I^i(z^i_\ell) + \rho^i_\ell(c^i_{\omega_k}(q^i_\ell))]\right] \\ &\quad\ge \sum_{i=1}^{N}\left\langle F^i(z^i_\ell,q^i_\ell), (z^i_\circ - z^i_\ell, q^i_\circ - q^i_\ell)\right\rangle \\ &\quad\ge -\sum_{i=1}^{N}\left\langle \sum_{k=1}^{K}[G^i_k(z^i_\ell,q^i_\ell)]^\top \pi_{\ell,k}, (z^i_\circ - z^i_\ell, q^i_\circ - q^i_\ell)\right\rangle. \end{aligned} \tag{6.10}$$

Using the convexity of h^i_k and of g^i_k, it follows that

$$[h^i_k(z^i_\circ) + g^i_k(q^i_{\circ,k})] - [h^i_k(z^i_\ell) + g^i_k(q^i_{\ell,k})] \ge G^i_k(z^i_\ell,q^i_\ell)(z^i_\circ - z^i_\ell, q^i_\circ - q^i_\ell).$$

and, hence, as multipliers are nonnegative,

$$\begin{aligned} &\left\langle [h^i_k(z^i_\circ) + g^i_k(q^i_{\circ,k})] - [h^i_k(z^i_\ell) + g^i_k(q^i_{\ell,k})], \pi_{\ell,k}\right\rangle \\ \ge\ &\langle [G^i_k(z^i_\ell,q^i_\ell)]^\top \pi_{\ell,k}, (z^i_\circ - z^i_\ell, q^i_\circ - q^i_\ell)\rangle. \end{aligned}$$

Using the latter relation and (6.9), we have that

$$\begin{aligned} &\left\langle \sum_{i=1}^{N}[h^i_k(z^i_\circ) + g^i_k(q^i_{\circ,k})], \pi_{\ell,k}\right\rangle \\ \ge\ &\sum_{i=1}^{N}\left\langle [G^i_k(z^i_\ell,q^i_\ell)]^\top \pi_{\ell,k}, (z^i_\circ - z^i_\ell, q^i_\circ - q^i_\ell)\right\rangle. \end{aligned}$$

Combining now the inequality above with (6.10), we obtain that

$$\begin{aligned}
&\sum_{i=1}^{N}\left[[I^i(z^i_\circ, z^{-i}_\ell)+\rho^i_\ell(c^i_{\omega_k}(q^i_\circ, q^{-i}_\ell))]-[I^i(z^i_\ell)+\rho^i_\ell(c^i_{\omega_k}(q^i_\ell))]\right]\\
\geq\ & -\sum_{i=1}^{N}\left\langle\sum_{k=1}^{K}[h^i_k(z^i_\circ)+g^i_k(q^i_{\circ,k})],\pi_{\ell,k}\right\rangle. \qquad (6.11)
\end{aligned}$$

Since the constraints in (6.1) satisfy the Slater condition, there exists $\varepsilon>0$ such that

$$\sum_{i=1}^{N}[h^i_k(z^i_\circ)+g^i_k(q^i_{\circ,k})]\leq -\varepsilon,\quad k=1,\ldots,K.$$

Hence, using once more the nonnegativity of the multipliers $\pi_{\ell,k}$, from (6.11) we obtain that

$$\sum_{i=1}^{N}\left[[I^i(z^i_\circ, z^{-i}_\ell)+\rho^i_\ell(c^i_{\omega_k}(q^i_\circ, q^{-i}_\ell))]-[I^i(z^i_\ell)+\rho^i_\ell(c^i_{\omega_k}(q^i_\ell))]\right]\geq\varepsilon\sum_{k=1}^{K}\|\pi_{\ell,k}\|_1.$$

Finally, by the continuity of I^i and of $c^i_{\omega_k}$, the boundedness of the sequence $\{(z_\ell,q_\ell)\}$ and the fact that $\{\rho^i_\ell\}$ converges continuously to ρ, the left-hand side of the last inequality above is bounded, which implies the first assertion.

To show (b), notice that the convexity of the functions involved in (6.8) implies that (z^i_ℓ, q^i_ℓ) solves the following problem:

$$\left\{\begin{array}{ll}\min_{z^i,q^i} & I^i(z^i,z^{-i}_\ell)+\rho^i_\ell(c^i_{\omega_k}(q^i,q^{-i}_\ell))\\ & \quad+\displaystyle\sum_{k=1}^{K}\left[h^i_k(z^i)+g^i_k(q^i_k)+\sum_{\substack{j=1\\ j\neq i}}^{N}[h^i_k(z^j_\ell)+g^i_k(q^j_{\ell,k})]\right]^\top\pi_{\ell,k}\\ \text{s.t.} & (z^i,q^i)\in X^i.\end{array}\right. \qquad (6.12)$$

Now, suppose that the subsequence $\{(z_{\ell_j},q_{\ell_j},\pi_{\ell_j})\}$ of the sequence $\{(z_n,q_n,\pi_n)\}$ converges to $(\bar z,\bar q,\bar\pi)$ as $j\to\infty$. Then, from (6.9), the complementarity condition still holds:

$$0\leq\bar\pi_k\perp\sum_{i=1}^{N}[h^i_k(\bar z^i)+g^i_k(\bar q^i_k)]\leq 0\quad\text{for } k=1,2,\ldots,K. \qquad (6.13)$$

Also, from (6.12), we have that for any $(z^i,q^i)\in X^i$ it holds that

$$\begin{aligned}
&I^i(z_{\ell_j})+\rho^i_{\ell_j}(c^i_{\omega_k}(q_{\ell_j}))+\sum_{k=1}^{K}[h^i_k(z^i_{\ell_j})+g^i_k(q^i_{\ell_j,k})]^\top\pi_{\ell,k_j}\\
\leq\ & I^i(z^i,z^{-i}_{\ell_j})+\rho^i_{\ell_j}(c^i_{\omega_k}(q^i,q^{-i}_{\ell_j}))+\sum_{k=1}^{K}[h^i_k(z^i)+g^i_k(q^i_k)]^\top\pi_{\ell_j,k}.
\end{aligned}$$

Then, passing onto the limit as $j \to \infty$ in the latter relation and using the fact that the subsequence $\{\rho^i_{\ell_j}\}$ converges continuously to ρ, we conclude that $(\bar{z}^i, \bar{q}^i)$ solves the problem

$$\begin{cases} \min_{z^i, q^i} & I^i(z^i, \bar{z}^{-i}) + \rho(c^i_{\omega_k}(q^i, \bar{q}^{-i})) \\ & \quad + \sum_{k=1}^{K} \Big[h^i_k(z^i) + g^i_k(q^i_k) + \sum_{\substack{j=1 \\ j \neq i}}^{N} [h^i_k(\bar{z}^j) + g^i_k(\bar{q}^j_k)] \Big]^\top \bar{\pi}_k \\ \text{s.t.} & (z^i, q^i) \in X^i. \end{cases} \tag{6.14}$$

This, together with the complementarity relation (6.13), establishes the second assertion. □

6.3 Approximating coherent risk measures

In this section we describe how to build suitable approximation functions for the risk measures in (5.32), involving the *AV@R*. Without loss of generality, we take the risk-aversion parameter $\kappa = 0$ in (5.32), so that

$$\rho(x) := AV@R_{\,\varepsilon}(x) = \min_u \left\{ u + \frac{1}{1-\varepsilon} \mathbb{E}\Big([x(\omega_k) - u]^+ \Big) \right\}. \tag{6.15}$$

To define continuously differentiable approximations for this function, we employ a sequence of scalar convex differentiable functions $\{\sigma_\ell\}$ that converges uniformly to the positive-part function as $\ell \to \infty$. The resulting approximating function ρ^i_ℓ is defined by

$$\rho^i_\ell(x) := \min_u \left\{ u + \frac{1}{1-\varepsilon} \mathbb{E}\Big(\sigma_\ell(x(\omega_k) - u \Big) \right\}. \tag{6.16}$$

We show below some useful properties of the approximating function.

Theorem 6.3.1. *Given a scalar convex differentiable function $\sigma_\ell(\cdot)$, suppose there exists $M > 0$ such that*

$$\Big| [a]^+ - \sigma_\ell(a) \Big| \leq M \text{ for all } a \in \mathbb{R}.$$

Then the function ρ^i_ℓ given by (6.16) is well defined, convex, differentiable, and satisfies

$$\Big| AV@R_{\,\varepsilon}(x) - \rho^i_\ell(x) \Big| \leq \frac{M}{1-\varepsilon} \text{ for all } x \in \mathbb{R}^K.$$

Proof. Given $u \in \mathbb{R}$ and $x \in \mathbb{R}^K$, we have that

$$\begin{aligned} & -\frac{M}{1-\varepsilon} + u + \frac{1}{1-\varepsilon} \mathbb{E}\Big([x(\omega_k) - u]^+ \Big) \\ \leq \;& u + \frac{1}{1-\varepsilon} \mathbb{E}\Big(\sigma_\ell(x(\omega_k) - u) \Big) \\ \leq \;& \frac{M}{1-\varepsilon} + u + \frac{1}{1-\varepsilon} \mathbb{E}\Big([x(\omega_k) - u]^+ \Big). \end{aligned} \tag{6.17}$$

Furthermore, for the minimand $u + \frac{1}{1-\varepsilon}\mathbb{E}\Big([x(\omega_k) - u]^+\Big)$ it holds that it is

$$\geq \begin{cases} u & \text{if } u > \max_k\{|x(\omega_k)|\}, \\ \left(1 - \dfrac{1}{1-\varepsilon}\right) u + \dfrac{1}{1-\varepsilon}\mathbb{E}\Big(x(\omega_k)\Big) & \text{if } u < -\max_k\{|x(\omega_k)|\}. \end{cases}$$

This shows that all the terms in (6.17) are coercive in the variable u, and thus the functions in (6.15) and (6.16) are well defined (the minimization problems therein have solutions).

Furthermore, taking the infimum values for all the expressions in the chain of inequalities (6.17), we obtain that

$$-\frac{M}{1-\varepsilon} + AV@R_{\,\varepsilon}(x) \leq \rho^i_\ell(x) \leq \frac{M}{1-\varepsilon} + AV@R_{\,\varepsilon}(x),$$

which is one of the assertions.

Convexity of ρ^i_ℓ can be checked directly by definition. Indeed, for any $t \in [0,1]$ and any x, y, let u_x and u_y be such that

$$\rho^i_\ell(x) = u_x + \frac{1}{1-\varepsilon}\mathbb{E}\Big(\sigma_\ell(x(\omega_k) - u_x)\Big)$$

and

$$\rho^i_\ell(y) = u_y + \frac{1}{1-\varepsilon}\mathbb{E}\Big(\sigma_\ell(y(\omega_k) - u_y)\Big).$$

Then

$$\begin{aligned} \rho^i_\ell(tx + (1-t)y) &\leq (tu_x + (1-t)u_y) \\ &\quad + \frac{1}{1-\varepsilon}\mathbb{E}\Big(\sigma_\ell\left(tx(\omega_k) + (1-t)y(\omega_k)\right) - (tu_x + (1-t)u_y)\Big) \\ &\leq t\left[u_x + \frac{1}{1-\varepsilon}\mathbb{E}\Big(\sigma_\ell(x(\omega_k) - u_x)\Big)\right] \\ &\quad + (1-t)\left[u_y + \frac{1}{1-\varepsilon}\mathbb{E}\Big(\sigma_\ell(y(\omega_k) - u_y)\Big)\right] \\ &= t\rho^i_\ell(x) + (1-t)\rho^i_\ell(y). \end{aligned}$$

It remains to prove the differentiability of ρ^i_ℓ. Take any $\bar{x} \in \mathbb{R}^k$ and any $w \in \partial\rho^i_\ell(\bar{x})$. Let u_ℓ be such that

$$\rho^i_\ell(\bar{x}) = u_\ell + \frac{1}{1-\varepsilon}\mathbb{E}\Big(\sigma_\ell(\bar{x}(\omega_k) - u_\ell)\Big).$$

The function

$$f_\ell(x) = u_\ell + \frac{1}{1-\varepsilon}\mathbb{E}\Big([\sigma_\ell(x(\omega_k) - u_\ell)\Big)$$

is convex and differentiable.

Note that for every $y \in \mathbb{R}^K$, it holds that

$$f_\ell(y) \geq \rho^i_\ell(y) \geq \rho^i_\ell(\bar{x}) + \langle w, y - \bar{x}\rangle = f_\ell(\bar{x}) + \langle w, y - \bar{x}\rangle,$$

which shows that

$$w \in \partial f_\ell(\bar{x}) = \{\nabla f_\ell(\bar{x})\}.$$

In particular, a subgradient w of ρ^i_ℓ at $\bar{x}$ is uniquely defined (is given by $\{\nabla f_\ell(\bar{x})\}$). This means that ρ^i_ℓ is differentiable with $\partial\rho^i_\ell(\bar{x}) = \{\nabla\rho^i_\ell(\bar{x})\}$ and

$$\nabla\rho^i_\ell(\bar{x}) = \frac{1}{1-\varepsilon}\nabla_x\mathbb{E}\Big(\sigma_\ell(\bar{x}(\omega_k) - u_\ell)\Big)$$

□

With Theorem 6.3.1 at hand, we immediately obtain the desired approximation result.

Corollary 6.3.2. *The functions defined in (6.16) provide a sequence of convex differentiable functions converging continuously to the function AV@R $_\varepsilon$.*

Proof. The fact that $\{\rho^i_\ell\}$ converges uniformly to *AV@R* $_\varepsilon$ is immediate from Theorem 6.3.1. Next, since the limit *AV@R* $_\varepsilon$ is itself a continuous function, the fact that convergence is continuous is by [68, Chapter 5]. □

Remark 6.3.3 (Approximating functions as risk measures)**.** *It can be shown that the approximating functions ρ_ℓ are risk measures themselves. Specifically, they satisfy the axioms of convexity, monotonicity and translation equivariance in [17, Chapter 6.3]. It would be of interest to study under which conditions they are also* coherent *measures. By Theorem 6.4 in [17], this requires showing positive homogeneity for ρ, or computing the conjugate function ρ^* to determine its domain. Since these calculations depend on the smoothing functions σ_ℓ, we do not analyze further this topic, and leave it for future research.*

Therefore, to build a sequence of approximating functions for *AV@R* $_\varepsilon$ it is enough to find a sequence of functions $\{\sigma_\ell\}$ that converges uniformly to the positive part-function $[\cdot]^+$. A general framework for building this kind of approximations is described in [8], together with a number of specific examples. Here we give a brief summary.

We begin by considering a nonnegative piecewise continuous function d with finite number of pieces such that

$$\int_{-\infty}^{+\infty} d(t)dt = 1 \quad \text{and} \quad \int_{-\infty}^{+\infty} |t|d(t)dt < \infty.$$

Then, given $\beta > 0$ we define an approximating function σ_β by

$$\sigma_\beta(x) := \int_{-\infty}^{x}\int_{-\infty}^{y} \frac{1}{\beta}d\big(\frac{t}{\beta}\big)dt\,dy. \tag{6.18}$$

It can be easily seen that (6.18) is well defined. Moreover, since d is nonnegative, we have that in fact (6.18) defines a convex function. Also, when d is strictly positive, σ_β is strictly convex.

We reproduce below, without a proof, part of [8, Proposition 2.2], which shows in particular that (6.18) provides a sequence of uniformly convergent functions to the positive-part function $[\cdot]^+$, as required for our applications.

Proposition 6.3.4. *The function σ_β defined in (6.18) satisfies the following properties.*

1. *It is nondecreasing, convex, and continuously differentiable. Moreover, if d is k-times continuously differentiable, then σ_β is $(k+2)$-times continuously differentiable.*

2. *For every x it holds that*

$$-D_2\beta \leq \sigma_\beta(x) - [x]^+ \leq D_1\beta,$$

where

$$D_1 = \int_{-\infty}^{0} |t| d(t)dt \quad \textit{and} \quad D_2 = \left[\int_{-\infty}^{+\infty} t d(t)dt\right]^+. \quad \square$$

Table 6.1 contains some families of smoothing functions, as well as their specific parameters D_1, D_2 and function d.

$d(t)$	$\sigma_\beta(x)$	D_1	D_2
$\dfrac{e^{-t}}{(1+e^{-t})^2}$	$x + \beta\log(1+e^{-\frac{x}{\beta}})$	$\log 2$	0
$\dfrac{2}{(t^2+4)^{\frac{3}{2}}}$	$\dfrac{x+\sqrt{x^2+4\beta^2}}{2}$	1	0
$\begin{cases} 1 & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2}{2\beta} & \text{if } 0 \leq x \leq \beta \\ x - \frac{\beta}{2} & \text{if } x > \beta \end{cases}$	0	$\frac{1}{2}$
$\begin{cases} 1 & \text{if } -\frac{1}{2} \leq t \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 0 & \text{if } x < -\frac{\beta}{2} \\ \frac{1}{2\beta}\left(x+\frac{\beta}{2}\right)^2 & \text{if } \lvert x\rvert \leq \frac{\beta}{2} \\ x & \text{if } x > \frac{\beta}{2} \end{cases}$	$\frac{1}{8}$	0

Table 6.1: Examples of smoothing functions.

6.4 Assessment on a simple problem

We consider a subset of the European gas network in Subsection 5.4, with only 4 agents, all producers. We assume that the ith player has L_i production units, and so the decision variables can be written as $z^i = (z^i_{l_1}, z^i_{l_2}, \ldots, z^i_{l_{L_i}})$ and $q^i_k = (q^i_{k,l_1}, q^i_{k,l_2}, \ldots, q^i_{k,l_{L_i}})$, for each uncertain realization ω_k, for $k = 1, \ldots, K$. Each production unit needs to take decisions over T periods of time. For example, the tth entry of the "here-and-now" variable $z^i_l = (z^i_{l1}, z^i_{l2}, \ldots, z^i_{lT}) \in \mathbb{R}^T$ represents the unit capacity expansion at the end of that period, which requires an investment cost of $\Gamma^i_l z^i_{lt}(1+r)^{-\sum_{s=1}^{t} y_s}$, where the constant Γ^i_l represents the unitary expansion cost, y_t represents the time that takes the period t and the factor $(1+r)^{-\sum_{s=1}^{t} y_s}$ pull back the cost to its present value. In this way, the "investment cost" $I^i(z^i, z^{-i})$ is given by

$$I^i(z^i, z^{-i}) = \sum_{t=1}^{T} \sum_{l=1}^{L_i} \Gamma^i_l z^i_{lt}(1+r)^{-\sum_{s=1}^{t} y_s},$$

which in fact is independent of the other players' decision variables. For each uncertainty realization k, the "wait-and-see" variable $q^i_{k,l}$ refers to the quantity of product delivered to a set of J markets where the unit l operates. So we write

$$q^i_{k,l} = (q^i_{k,lj_1}, q^i_{k,lj_2}, \ldots, q^i_{k,lj_J}),$$

where the tth entry of

$$q^i_{k,lj} = (q^i_{k,lj1}, q^i_{k,lj2}, \ldots, q^i_{k,ljT})$$

represents the quantity delivered to market j at period t. The random operational cost is given by

$$c^i_\omega(q^i_k, q^{-i}_k) = \sum_{t=1}^{T} \sum_{l=1}^{L_i} f_t \left\{ G^i_l \left(\sum_{j=1}^{J} q^i_{k,ljt} \right) - \sum_{j=1}^{J} [P^t_{k,j}(q^i_{k,jt}, q^{-i}_{k,jt}) - c^i_{ljt}] q^i_{k,ljt} \right\},$$

where the function

$$G^i_l(x) = a^i_l x - b^i_l \log(B^i_l - x)$$

represents the production cost of the unit l of player i. The constant c^i_{ljt} represents the unitary cost for shipping each item from unit l to market j at period t. The inverse-demand function is now nonlinear:

$$P^t_{k,j}(q_{k,jt}) = p_{0_{k,jt}} \left(\frac{D_{jt} + \sum_{i=1}^{N} \sum_{l=1}^{L_i} q^i_{k,ljt}}{q_{0_{k,jt}}} \right)^{\frac{1}{e_{jt}}}.$$

Here e_{jt} represents the price elasticity of demand in the market j at time t. The random quantities $p_{0_{k,jt}}$ and $q_{0_{k,jt}}$ are the base price and demand, respectively. They are expressed as

$$p_{0_{k,jt}} = p_{0_{jt}} \frac{p_t(\omega_k)}{r_t} \quad \text{and} \quad q_{0_{k,jt}}(\omega) = q_{0_{jt}} \left(\frac{p_t(\omega_k)}{r_t} \right)^{\eta_t},$$

where random variable $p_t(\omega)$ is uniformly distributed in some interval $[\pi_{Lt}, \pi_{Ut}]$. In our initial testing, we sample this vector to obtain a finite number of scenarios.

Finally, we describe feasible sets of each player, denoted by X^i. The points $(z^i, q^i) \in X^i$ have nonnegative entries and satisfy the following two constraints for $k = 1, \ldots, K$, $l = 1, 2, \ldots, L_i$ and $t = 1, 2, \ldots, T$:

$$R_l^i - \sum_{s=1}^{t} y_s \sum_{j=1}^{J} q_{k,ljs}^i \geq 0 \quad \text{and} \quad K_l^i + \sum_{s=1}^{t-1} z_l^i - \sum_{j=1}^{J} y_t q_{k,ljt}^i \geq 0\,.$$

In these relations, R_l^i is the initial amount of gas available for the player i at production unit l (so the constraint means that along the time steps we cannot sell more product than the initial available amount). The rightmost set of constraints relates production to the effective installed capacity.

Numerical results

We coded the smoothing approximation procedure in Matlab (R2012a), using PATH [18, 28] to solve the variational problems. The runs were performed on a PC operating under Ubuntu 12.04-64 bit with a processor Intel Atom 1.80GHz $\times$ 4 and 2GB of memory.

In this preliminary implementation we considered only 10 scenarios for uncertainty in a subregion of the European gas network from Chapter 5. The producers are:

1. Russia,
2. the Netherlands,
3. Norway,
4. Algeria.

These players act on the following markets:

1. Belgium-Luxembourg,
2. Germany,
3. France,
4. the Netherlands,
5. Italy,
6. UK.

For this game without traders or outsourcing players, we use the data given in [33] with $T = 1$.

To determine the influence of the smoothing parameter we took $\beta = 10^{-j}$ for $j = 1, 2, \ldots, 5$ for all the instances, using as smoothing function the third one in Table 6.1, which was the one with better numerical behavior in our tests. The instances were the risk neutral model ($\kappa = 0$ in (5.32)) and full risk-averse variants with $\kappa = 1$ and $\varepsilon \in \{0.025, 0.05, 0.075, 0.1\}$ for the four players.

For the (very simple) test-case considered here, the different smoothing parameters gave the same primal ("here-and-now" and "wait-and-see") and dual variables, except for one scenario in two instances, where the difference was smaller than 5%, and only in the "wait-and-see" part. There was no expansion ($\bar{z} = 0$) and for the price calculations we averaged over the different smoothing parameters the "wait-and-see" components of a solution, $\bar{q}$. Equilibrium prices are obtained by evaluating the inverse-demand function at $\bar{q}$.

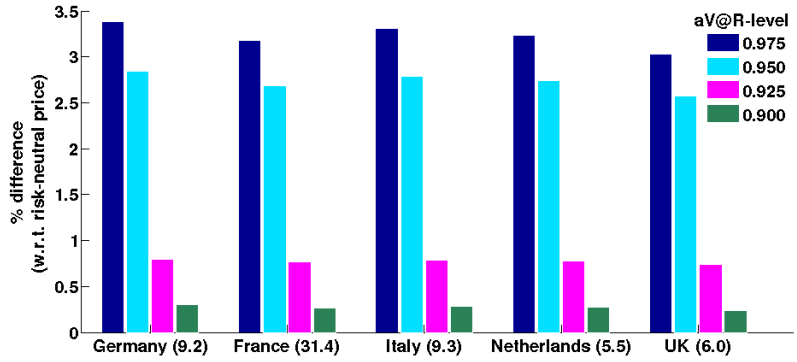

Figure 6.1: Variation in Equilibrium Prices for different values of risk aversion: $1 - \varepsilon \in \{0.975, 0.95, 0.925, 0.9\}$

Figure 6.1 reports the equilibrium prices for the five markets with actual production: there was no activity in Belgium-Luxembourg, whose optimal primal variables were always null. In the figure, the abscissa gives the name of the market and the risk-neutral equilibrium price. The different bars show the percentage difference in the equilibrium prices relative to the risk-neutral one. We observe a natural phenomenon: as producers become more averse to risk, the equilibrium prices increase. Since we employed the same parameters of risk aversion for all the producers, the perception of risk is the same and so is its impact in the different markets.

Chapter 7

Conclusions and Perspectives

This work was oriented into two main directions: decomposition strategies for variational problems and equilibrium models for energy markets.

On the first issue, we presented two decomposition algorithms for solving VI problems that make it possible to take advantage of special structures present in their feasible sets. The first algorithm follows the Dantzig-Wolfe paradigm, which was developed into a broad family of decomposition methods that can be applied to maximal monotone operators (possibly set-valued) or single-valued continuous operators (possibly nonmonotone). The approach allows for various kinds of approximations of the problem data and its derivatives in the single-valued case, as well as inexact solution of subproblems. The resulting algorithmic patterns are shown to be convergent under reasonable assumptions on the variational problem. These characteristics represent a significant improvement with respect to previous works, that required too strong assumptions and, hence, restricted considerably the algorithm's applicability. By contrast, the decomposition scheme introduced in this work is highly versatile and makes it possible to exploit structural properties of the problem even if the operator therein is not separable and maybe not monotone, as is common when solving GNEPs.

Our numerical results show that, even with an extremely simplified modeling of an electrical power market, large instances become intractable with a direct solution method, and can only be solved by decomposition. In terms of accuracy, speed, and scalability, among all the considered variants and for our battery of tests, the decomposition method using a combination of Newton and Jacobi approximations appears to be the best one.

The second algorithm follows the Benders decomposition paradigm. It also represents an improvement with respect to previous work; in fact naturally so, since it is based on our more general Dantzig-Wolfe method. The possibility to choose approximations in Dantzig-Wolfe algorithm leads to subproblems in the

Benders method that are not only simpler VIs but also, in some cases, reduce to simple linear or quadratic minimization problems. We have not yet made numerical tests for the different variants of this decomposition technique. Nevertheless, the good results obtained with the Dantzig-Wolfe algorithm provide us with a good expectation for the related Benders approach. The numerical assessment of this part is left for our future work.

The second part of this dissertation was devoted to modeling equilibria in energy markets. The contribution in this part begins by considering models based on GNEPs, that can be solved by means of VI reformulations, instead of the widely used mixed complementarity models. We showed that these two approaches are equivalent in the deterministic or risk neutral setting. One advantage of modeling the market as a GNEP is to ensure, easily and under mild assumptions, the existence of solutions. Moreover, the GNEP approach also allowed us to extend the deterministic model to the stochastic one in a meaningful way. Our numerical results on a real-life case, the European network of natural gas, show the type of information the GNEP model provides in terms of strategic behavior of the players.

We intend to study in more details topics related to stochastic variational equilibria, to understand better the connections with MCP formulations in the risk averse case. Moving from two stages to a multistage setting can also be of interest. This item is left for future research. In particular, its numerical validation will require to deepen further the study on approximation techniques for risk averse GNEPs. This item is a final contribution of this work. We put in place a smoothing technique which combined with simulation (defining larger and larger sample sets) can be used to find variational equilibria when the agents in the market hedge against risk. An interesting line of future research is to determine under which conditions the smoothed function becomes a coherent risk measure.

Finally, an important by-product of the GNEP formulation and its solution via the associated VI, is that the procedure reveals decomposable structures that can then be exploited within our decomposition approaches to reduce solution times. This is especially important for stochastic games, where the problem size is commonly large. The numerical tests in this case are still at a preliminary stage, and the need of decomposition methods did not appear for the size of the test-study. However, some interesting results were obtained from our model, for a subregion of the European gas network. The use of decomposition algorithms on larger stochastic models, possibly multistage, is another subject of future research.

Bibliography

[1] H. Attouch and M. Théra. A general duality principle for the sum of two operators. *Journal of Convex Analysis*, 3:1–24, 1996.

[2] A. Auslender and M. Teboulle. Lagrangian duality and related multiplier methods for variational inequality problems. *SIAM Journal on Optimization*, 10:1097–1115, 2000.

[3] H. Ben Amor, J. Desrosiers, and A. Frangioni. On the Choice of Explicit Stabilizing Terms in Column Generation. *Discrete Applied Mathematics*, 157:1167–1184, 2009.

[4] J.F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.

[5] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. Numerical Optimization: Theoretical and Practical Aspects. Springer-Verlag, Berlin, Germany, 2006 (2nd edition).

[6] O. Briant, C. Lemaréchal, Ph. Meurdesoif, S. Michel, N. Perrot, and F. Vanderbeck. Comparison of bundle and classical column generation. *Mathematical Programming*, 113:299–344, 2008.

[7] R. S. Burachik and A. N. Iusem. Set-valued Mappings and Enlargements of Monotone Operators. Springer Optimization and Its Applications, 8. Springer, New York, (2008).

[8] C. Chen and O. L. Mangasarian. A class of smoothing functions for nonlinear and mixed complementarity problems. *Computational Optimization and Applications*, 5:97–138, 1996.

[9] G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Mathematical Programming*, 64:81–101, 1994.

[10] X. Chen X and G. Lin. CVaR-based formulation and approximation method for stochastic variational inequalities *Numerical Algebra, Control and Optimization (NACO)*, 1(1):35–48, 2011.

[11] X. Chen, R.J.-B. Wets and Y. Zhang. Stochastic variational inequalities: residual minimization smoothing sampling average approximations. *SIAM Journal on Optimization*, 22(2):649–673, 2012.

[12] W. Chung and J.D. Fuller. Subproblem approximation in Dantzig-Wolfe decomposition of variational inequality models with an application to a multicommodity economic equilibrium model. *Operations Research*, 58:1318–1327, 2010.

[13] A.J. Conejo, M. Carrión and J. M. Morales. Decision Making Under Uncertainty in Electricity Markets. International Series in Operations Research & Management Science. Springer, New York (2010).

[14] R.W. Cottle, J.-S. Pang, and R.E. Stone. The Linear Complementarity Problem. Academic Press, Inc., Boston, MA, 1992.

[15] T.G. Crainic, A. Frangioni, and B. Gendron. Bundle-based Relaxation Methods for Multi-commodity Capacitated Fixed Charge Network Design Problems. *Discrete Applied Mathematics*, 112:73–99, 2001.

[16] G.B. Dantzig and P. Wolfe. The decomposition algorithm for linear programs. *Econometric*, 29:767–778, 1961.

[17] D. Dentcheva, A. Ruszczyński and A. Shapiro. Lectures on Stochastic Programming. SIAM, Philadelphia (2009)

[18] S.P. Dirkse and M.C. Ferris. The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software*, 5:123–156, 1995.

[19] J. Eckstein and D.P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[20] J. Eckstein. Some saddle-function splitting methods for convex programming. *Optimization Methods and Software*, 4:75–83, 1994.

[21] R. Egging, S. A. Gabriel, F. Holtz and J. Zhuang. A complementarity model for the European natural gas market. *Energy Policy*, 36:2385–2414, 2008.

[22] A. Ehrenmann and Y. Smeers. Capacity expansion in non-regulated electricity markets. In: HICSS, 2006.

[23] A. Ehrenmann and Y. Smeers. Generation capacity expansion in risky environment: A stochastic equilibrium analysis. *Operations Research* 59(6):1332–1346, 2011.

[24] J. Escobar and A. Jofré. Monopolistic competition in electricity networks with resistance losses. *Economic Theory*, 44:101–121, 2010.

[25] F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. *Annals of Operations Research*, 175:177–211, 2010.

[26] F. Facchinei, A. Fischer and V. Piccialli. On generalized Nash games and variational inequalities. *Operations Research Letters*, 35:159–164, 2007.

[27] F. Facchinei and J.S. Pang. Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer–Verlag, New York, NY, 2003.

[28] M.C. Ferris and T.S. Munson. Interfaces to PATH 3.0: Design, implementation and usage. *Computational Optimization and Applications*, 12:207–227, 1999.

[29] J.D. Fuller and W. Chung. Dantzig–Wolfe Decomposition of Variational inequalities. *Computational Economics*, 25:303–326, 2005.

[30] J.D. Fuller and W. Chung. Benders decomposition for a class of variational inequalities. *European Journal of Operational Research*, 185:76–91, 2008.

[31] S. Gabriel and J. Fuller. A Benders decomposition method for solving stochastic complementarity problems with an application in energy. *Computational Economics* 35(4):301–329, 2010.

[32] S.A. Gabriel, J. Zhuang, and R. Egging. Solving stochastic complementarity problems in energy market modeling using scenario reduction. *European Journal of Operational Research* 197(3):1028 – 1040, 2009.

[33] G. Gürkan, Y. Ozgë and S.M. Robinson. Sample path solution of stochastic variational inequalities. *Math. Programming* 83(Ser. A):313–333, 1999.

[34] P.T. Harker and J.S. Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Math. Programming* 48(2,(Ser. B)):161–220, 1990.

[35] B. He, L.Z. Liao, D. Han and H. Yang. A new inexact alternating directions method for monotone variational inequalities. *Mathematical Programming*, 92:103–118, 2002.

[36] J.B. Hiriart-Urruty and C. Lemaréchal. Convex Analysis and Minimization Algorithms. No. 305-306 in Grund. der math. Wiss. Springer-Verlag (1993). (two volumes)

[37] B. Hobbs, C. Metzler and J.S. Pang. Strategic gaming analysis for electric power systems: an MPEC approach. *Power Systems, IEEE Transactions on* 15(2):638–645, 2000.

[38] B. F. Hobbs and J. S. Pang. Nash-Cournot Equilibria in Electric Power Markets with Piecewise Linear Demand Functions and Joint Constraints *Operations Research*, 55:113–127, 2007.

[39] X. Hu and D. Ralph. Using epecs to model bilevel games in restructured electricity markets with locational prices. *Operations Research* 55(5):809–827, 2007.

[40] J. Eckstein and M. C. Ferris. Smooth methods of multipliers for complementarity problems. *Mathematical Programming*, 86:65–90, 1999.

[41] K.L. Jones, I.J. Lustig, J.M. Farwolden, and W.B. Powell. Multicommodity Network Flows: The Impact of Formulation on Decomposition. *Mathematical Programming*, 62:95–117, 1993.

[42] N.H. Josephy. Newton's method for generalized equations. *Technical Summary Report no. 1965. Mathematics Research Center, University of Wisconsin, Madison*, 1979.

[43] A. Kannan and V.M. Zavala. A Game-Theoretical dynamic model for electricity markets. Argonne National aboratory, Technical Report ANL/MCS-P1792-1010, 2011.

[44] A. Kannan, U.V. Shanbhag, and H.M. Kim. Addressing supply-side risk in uncertain power markets: stochastic generalized Nash models, scalable algorithms and error analysis. *Optimization Methods and Software*, `www.tandfonline.com/doi/abs/10.1080/10556788.2012.676756`.

[45] A. Kannan, U.V. Shanbhag, and H.M. Kim. Strategic Behavior in Power Markets under Uncertainty. *Energy Systems*, 2: 115–141, 2011.

[46] I.V. Konnov. Combined Relaxation Methods for Variational Inequalities (Lecture Notes in Economics and Mathematical Systems, Vol.495). Springer-Verlag, Berlin-Heiderberg-New York, 2001.

[47] A.A. Kulkarni and U.V. Shanbhag. On the variational equilibrium as a refinement of the generalized Nash equilibrium. *Automatica*, 48: 45-55, 2012.

[48] A.A. Kulkarni and U.V. Shanbhag. Revisiting generalized Nash games and variational inequalities. *Journal of Optimization Theory and Applications*, 154(1):175–186, 2012.

[49] J.P. Luna, C. Sagastizábal, and M. Solodov. A class of Dantzig–Wolfe type decomposition methods for variational inequality problems. *Mathematical Programming*, 2012. DOI 10.1007/s10107-012-0599-7. Online First.

[50] J.P. Luna, C. Sagastizábal, and M. Solodov. Complementarity, game-theoretical, and variational-inequality models for energy markets. 2013, submitted to *Risk Management for Energy Production and Trading*, Editors: R.M. Kovacevic, G.Ch. Pflug, and M.T. Vespucci. Springer, International Series in Operations Research and Management Science.

[51] J.P. Luna, C. Sagastizábal, and M. Solodov. A class of Benders type decomposition methods for variational inequalities, with an application to generation capacity expansion problems. 2013, in preparation.

[52] J.P. Luna, C. Sagastizábal, and M. Solodov. An approximation scheme for generalized Nash equilibrium problems with risk aversion. 2013, in preparation.

[53] P.A. Lotito, L.A. Parente and M.V. Solodov. A class of variable metric decomposition methods for monotone variational inclusions. *Journal of Convex Analysis*, 16:857–880, 2009.

[54] Z.-Q. Luo and P. Tseng. Error bound and convergence analysis of matrix splitting algorithm for the affine variational inequality problems. *SIAM Journal on Optimization*, 2:43–54, 1992.

[55] B. Martinet. Régularisation d'inéquations variationelles par approximations successives. *Revue Française d'Informatique et de Recherche Opérationelle*, 4:154–159, 1970.

[56] U. Mosco. Dual variational inequalities. *Journal of Mathematical Analysis and Applications* , 40:202–206, 1972.

[57] L.A. Parente, P.A. Lotito and M.V. Solodov. A class of inexact variable metric proximal point algorithms. *SIAM Journal on Optimization*, 19:240–260, 2008.

[58] T. Pennanen. A splitting method for composite mappings. *Numerical Functional Analysis and Optimization*, 23:875–890, 2002.

[59] A. Philpott and V. de Matos. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. Tech. rep., Optimization Online (2010). URL `http://www.optimization-online.org/DB_HTML/2010/12/2861.html`

[60] G.C. Pflug and W. Römisch. Modeling, Measuring and Managing Risk. *World Scientific.* 2007.

[61] D. Ralph and Y. Smeers. EPECs as models for electricity markets. In: Power Systems Conference and Exposition, 2006. PSCE '06. 2006 IEEE PES, pp. 74 –80 (2006). DOI 10.1109/PSCE.2006.296252

[62] S.M. Robinson. A reduction method for variational inequalities. *Math. Programming* 80(2,Ser. A):161–169, 1998.

[63] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877–898, 1976.

[64] R.T. Rockafellar. Local boundedness of nonlinear, monotone operators. *Michigan Mathematical Journal*, 16:397–407, 1969.

[65] R.T. Rockafellar. On the maximality of sums of nonlinear monotone operators. *Transactions of the Amererican Mathematical Society*, 149:75–88, 1970.

[66] R. Rockafellar and S. Uryasev. Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance* 26(7):1443–1471, 2002.

[67] R.T. Rockafellar, S. Uryasev and M. Zabarankin. Optimality conditions in portfolio analysis with general deviation measures. *Mathematical Programming* 108:515–540, 2006.

[68] R.T. Rockafellar and R.J.-B. Wets. Variational analysis. Springer–Verlag, Berlin, 1998.

[69] R.T. Rockafellar and R.J.-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research.* 16(1):119–147, 1991.

[70] C.A. Sagastizábal. Divide to conquer: decomposition methods for energy optimization. *Math. Program.* 134:187–222, 2012.

[71] C.A. Sagastizábal and M.V. Solodov. Parallel variable distribution for constrained optimization. *Computational Optimization and Applications*, 22:111–131, 2002.

[72] U.V. Shanbhag, G. Infanger, and P.W. Glynn. A complementarity framework for forward contracting under uncertainty. *Operations Research*, 59: 810–834, 2011.

[73] A. Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research* 209(1):63–72, 2011.

[74] A. Shapiro and W. Tekaya. Report for technical cooperation between Georgia Institute of Technology and ONS - Operador Nacional do Sistema Elétrico (2011). Report 2: Risk Averse Approach

[75] M.V. Solodov. A class of globally convergent algorithms for pseudomonotone variational inequalities. In *Complementarity: Applications, Algorithms and Extensions*, M C. Ferris, O.L. Mangasarian and J.-S. Pang (editors), Applied Optimization 50, Kluwer Academic Publishers, 2001, Chapter 14, pp. 297–315.

[76] M.V. Solodov. A class of decomposition methods for convex optimization and monotone variational inclusions via the hybrid inexact proximal point framework. *Optimization Methods and Software*, 19:557–575, 2004.

[77] M.V. Solodov. Constraint qualifications. In *Wiley Encyclopedia of Operations Research and Management Science*, James J. Cochran, et al. (editors), John Wiley & Sons, Inc., 2010.

[78] M.V. Solodov and B.F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numerical Functional Analysis and Optimization*, 22:1013–1035, 2001.

[79] M.V. Solodov and P. Tseng. Modified projection-type methods for monotone variational inequalities. *SIAM Journal on Control and Optimization*, 34:1814–1830, 1996.

[80] J.E. Spingarn. Applications of the method of partial inverses to convex programming. *Mathematical Programming*, 32:199–223, 1985.

[81] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM Journal on Control and Optimization*, 29:119–138, 1991.

[82] P. Tseng. Alternating projection-proximal methods for convex programming and variational inequalities. *SIAM Journal on Optimization*, 7:951–965, 1997.

[83] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38:431–446, 2000.

[84] J. Yao, S.S. Oren and I. Adler. Two-settlement electricity markets with price caps and Cournot generation firms. *European Journal of Operational Research* 1279–1296, 2007.

[85] H. Yin, U. Shanbhag and P. Mehta. Nash equilibrium problems with scaled congestion costs and shared constraints. *Automatic Control, IEEE Transactions on* 56(7):1702 –1708, 2011.

[86] C. Zalinescu. Convex analysis in general vector spaces. World Scientific Publishing Co. (2002)