ALGORITHMS FOR MATHEMATICAL PROGRAMS WITH EQUILIBRIUM CONSTRAINTS WITH APPLICATIONS TO DEREGULATED ELECTRICITY MARKETS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MANAGEMENT SCIENCE AND ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Heiko Pieper
June 2001

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Walter Murray
(Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Hung-Po Chao

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard W. Cottle

Approved for the University Committee on Graduate Studies:

Abstract

Mathematical programs with equilibrium constraints (MPECs) are optimization problems with some constraints defined in terms of complementarity systems. Important applications of these problems arise in engineering design problems of mechanical structures, economic models, and option pricing. We have developed a new algorithm for MPECs, which we apply to solve novel economic models of deregulated electricity markets.

It can be shown that constraint qualifications typically assumed to prove convergence of standard algorithms fail to hold for MPECs. As a result, applying standard algorithms is problematic. To circumvent these problems, various reformulations of MPECs have been proposed. One of these approaches involves the use of smoothing functions with favorable properties to substitute for the complementarity constraints. We investigate a new sequential quadratic programming algorithm for equilibrium-constrained optimization (ECOPT) based on such a smooth reformulation. The algorithm employs a specialized termination criterion as well as update rules for the Lagrangian Hessian. Numerical tests on standard test problems show its performance is superior to that of state of the art nonlinear optimization algorithms as well as some other algorithms specifically designed to solve MPEC problems.

We also present a new mathematical model of electricity forward markets. The lack of working forward markets in electricity has been identified as one of the main obstacles to current deregulation efforts. Our new model incorporates a Cournot equilibrium for the spot market and considers actions by producers in the forward market. The mathematical model is an instance of an MPEC. Using ECOPT to solve the producer's problem, one can find Nash equilibria in the forward market.

The application of the model to a six-node network with two competing producers reveals a fundamental relationship between transmission capacity and forward markets. We also demonstrate how to apply the model to gain a better understanding of transmission investment decisions in deregulated electricity markets.

Acknowledgements

I would like to thank my advisor, Professor Walter Murray. He was instrumental in my coming to Stanford and a constant source of enthusiasm, advice, guidance, and ideas. Working with him has been a joyful and enlightening experience. On many occasions our meetings lead to discussions about European soccer and skiing conditions in Tahoe.

Several other people have influenced the work in this thesis. I thank Professor Hung-Po Chao for his guidance and assistance in the development of the electricity model. His insight was most helpful in interpreting the results and focussing on the important issues. Many thanks are also due to Professor Richard W. Cottle for his advice and support over the years at Stanford and for serving on my oral and reading committee. His detailed reading of the draft of this dissertation and many suggestions improved the presentation notably. I thank Professor Michael A. Saunders and Professor Robert Wilson for serving on my orals committee and Christian Kanzow for introducing me to complementarity.

Many friends have made my time at Stanford most enjoyable. I benefited from frequent discussions with my office mates Michael P. Friedlander and Angel-Victor de Miguel. I wish that my friendship with them and many others will last well beyond the boundaries of the beautiful Stanford Campus. I would also like to thank all the faculty and staff of the Department of Management Science and Engineering for their support over the years. Financial support for this research was provided by a first-year fellowship from the School of Engineering and a research assistantship through EPRI.

Finally, I would like to thank my family for all the love and support throughout my life. My parents and my sister always encouraged me to pursue my dreams and enabled me to achieve my goals. I dedicate this dissertation to Omi. My grandmother was very proud to see me come here, and I wish she could be with us today.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem statement

The Mathematical Program with Equilibrium Constraints (MPEC) is given by

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{subject to} \quad & g(x,y) \geq 0 \\ & y \geq 0, \;\; F(x,y) \geq 0, \;\; y^T F(x,y) = 0, \end{aligned} \tag{1.1}$$

where $f : \mathbb{R}^{n+m} \to \mathbb{R}$, $g : \mathbb{R}^{n+m} \to \mathbb{R}^l$, and $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ are twice continuously differentiable functions. More specifically, (1.1) is an MPEC with nonlinear complementarity constraints, since the *lower-level constraints* take the form of a parametric nonlinear complementarity problem

$$y \geq 0, \;\; F(x,y) \geq 0, \;\; y^T F(x,y) = 0.$$

The formulation (1.1) extends easily to more general formulations of MPECs since the *upper-level constraints* $g(x,y) \geq 0$ are joint constraints in all variables.

Originating from engineering problems, the *first-level variables* x are sometimes called *design variables* while the second-level variables y are referred to as *state variables*. The history of the MPEC can be traced back to the economic notion of a

Stackelberg game [Sta52]. Many application in the fields of economics, engineering, and finance can be formulated as MPECs, see the monograph [LPR96] for an overview. The term "equilibrium constraints" in MPEC refers to the complementarity constraints, or more general variational inequality constraints, which represent certain equilibrium conditions in engineering and economic models.

Mathematically the MPEC is a very challenging problem. First of all, the solution set to the lower-level complementarity constraints is nonconvex so that the best one can hope for is to find local solutions of (1.1). Second, as we will see later, certain constraint qualifications needed to prove convergence for standard nonlinear optimization algorithms fail to hold at any feasible point. Specialized algorithms are needed to address this particular problem. Indeed, at the present time, no fast and reliable solvers for MPECs are available.

The MPEC can be seen as a generalization of a bilevel program where the second-level optimization problem is written in terms of its optimality conditions. In recent years, many interesting applications in engineering and economics have been formulated as MPECs. For example, the first application of MPECs to deregulated electricity markets was suggested in [HMP00].

In this thesis, we address mathematical as well as modeling aspects of MPECs. First, we investigate the theoretical and practical properties of a new algorithm for the solutions of MPECs. The numerical tests show that the new approach advances the current state of algorithms for MPECs and performs better than state-of-the-art algorithms for standard nonlinear optimization. Second, we present a new model for forward markets in deregulated electricity environments. The model addresses the important strategic interactions between forward and spot markets in electricity, which have not been studied in this detail until now. In particular the inclusion of an accurate network representation of the electricity grid allows the study of complicated strategic interactions between market players in the spot and forward market. As a byproduct of our analysis, an intrinsic relationship between transmission capacity and the development of forward markets is revealed.

This chapter begins by describing two applications; an optimal design problem from engineering and the Stackelberg game in economics. We review the current state of deregulated electricity markets and give a preview of later chapters.

1.2 Applications

1.2.1 Engineering: Optimal design of mechanical structures

The following structural optimization problem, whose objective is to minimize the weight of a truss subject to meeting certain load-bearing specifications, is described as one of the source problems in the monograph [LPR96].

Suppose that we have a mechanical structure, say a truss, with m solid steel bars or elements whose volumes are specified through the components of a vector $t \in \mathbb{R}^m$, e.g. each t_i may be the area of the circular cross section of a bar of fixed length. The stiffness equation relates the vector nodal forces F to the vector of nodal displacement u via

$$F := K(t)u,$$

where $K(t)$ is the symmetric stiffness matrix that is positive definite for $t > 0$, such as

$$K(t) := \sum_{i=1}^{m} t_i K_i$$

and each K_i is a symmetric positive definite matrix.

A point that may come into frictionless contact with a rigid obstacle, e.g. the ground, is considered. The kinematic conditions that nodes of the structure cannot penetrate the obstacle are expressed by

$$Cu \geq g, \tag{1.2}$$

where C is a kinematic transformation matrix and g is a vector of initial distances between nodes and the rigid obstacle. In terms of the matrix C, we can decompose the nodal forces F as the sum of the load (external) forces f and forces due to the unilateral constraints:

$$F = f + C^T p,$$

where p is the vector of contact forces, work conjugate to the vector Cu of contact displacements. Given the previous stiffness equation, this last equation amounts to

$$K(t)u - f - C^T p = 0. \tag{1.3}$$

Adhesionless contact requires that

$$p \leq 0; \tag{1.4}$$

and ruling out forces without contact gives the complementarity condition:

$$p^T(Cu - g) = 0. \tag{1.5}$$

For each fixed t, the conditions (1.2)–(1.5) define a mixed linear complementarity problem in the variables u and p. A simple form of the minimum weight design problem is

$$\begin{array}{rl} \min\limits_{t,u,p} & w(t) \\ \text{subject to} & \underline{t} \leq t \leq \overline{t} \\ & \underline{u} \leq u \leq \overline{u} \\ \text{and} & (1.2)\text{–}(1.5), \end{array}$$

where the goal is to minimize the cost $w(t)$ associated with volume of steel used over all elements, given lower and upper bounds $\underline{t}_i$ and $\overline{t}_i$ on the "volume" t_i of each element, and lower and upper bounds $\underline{u}_j$ and $\overline{u}_j$ on each displacement u_j. Clearly, this problem is an instance of the MPEC with mixed linear complementarity constraints; t is the upper-level variable and (u, p) is the lower-level variable. By including the equality constraints (1.3) into the first-level constraints, it can be cast as an MPEC of the form (1.1).

1.2.2 Economics: Generalized Stackelberg games

Nash games

The Stackelberg game is an extension of the renowned Nash game [Nas51]. In the Nash game, there are M players each of whom has a strategy set $Y_i \subseteq \mathbb{R}^{m_i}$. The objective of player i is to minimize its economic cost $\theta_i(y_i, y_{\neq i}^{\text{given}})$ by selecting a strategy $y_i \in Y_i$ given that the other players have chosen their strategies $y_{\neq i}^{\text{given}}$, where $y_{\neq i}^{\text{given}}$ denotes the vector $(y_j^{\text{given}} : j \neq i)$. In other words, each player observes the actions of the remaining players and then reacts optimally, assuming the other players' strategies remain unchanged. A strategy combination $y^* \in \prod_{j=1}^{m} Y_j$ is called a *Nash equilibrium* if no player has an incentive to deviate from his strategy y_i^* in the sense that

$$y_i^* \in \operatorname{argmin}\{\theta_i(y_i, y_{\neq i}^*) : y_i \in Y_i\}, \quad \forall i.$$

It should be noted that players in a Nash game are in some sense homogeneous since each of them has access to the same information regarding the other players' strategies and the strategy chosen is only dependent on this information.

As an example, consider the special case where the functions $\theta_i(y_i, y_{\neq i}^{\text{given}})$ are convex and continuously differentiable and $Y_i = \mathbb{R}_+^{m_i}$, i.e., each player i solves the optimization problem

$$\begin{aligned} \min_{y_i} \quad & \theta_i(y_i, y_{\neq i}^{\text{given}}) \\ \text{subject to} \quad & y_i \geq 0, \end{aligned} \tag{1.6}$$

the Nash equilibrium can be cast as a nonlinear complementarity problem in the following way: Let $p := \sum_{i=1}^{M} m_i$, $y \in \mathbb{R}^p$, and $F(y) := (F_i(y))_{i=1}^{M}$ with

$$F_i(y) := \nabla_{y_i} \theta_i(y), \quad i = 1, \ldots, M,$$

then the first-order optimality conditions of all players optimization problems (1.6) can be solved simultaneously by solving the nonlinear complementarity problem

$$y \geq 0, F(y) \geq 0, y^T F(y) = 0.$$

If the strategy spaces Y_i are given by general convex sets, the Nash equilibrium problem takes the more general form of a variational inequality problem, see [LPR96].

Stackelberg games

In contrast to the Nash game, the Stackelberg game has a distinctive player (called the leader) who can *anticipate* the (re)actions of the remaining players (called followers) and use this knowledge in selecting his optimal strategy (see [Ras94]). Specifically, the leader chooses a strategy from the strategy set $X \in \mathbb{R}^n$, while each follower i has, corresponding to each of the leader's strategies $x \in X$, a closed and convex strategy set $Y_i(x) \subseteq \mathbb{R}^{m_i}$ and a cost function $\theta(x, \cdot) : \prod_{j=1}^{M} \mathbb{R}_{m_j} \to \mathbb{R}$, where M is the number of followers in the Stackelberg game. Note that the follower's strategy is dependent on the particular strategy x of the leader and this follower's cost function is dependent on both the leader's and the follower's strategies. We assume that for any fixed but arbitrary $x^{\text{given}} \in X$ and $y_{\neq i}^{\text{given}} := (y_j^{\text{given}} : j \neq i)$, the function

$$\theta_i(x^{\text{given}}, y_i, y_{\neq i}^{\text{given}})$$

is convex and continuous differentiable in the variable $y_i \in Y_i(x^{\text{given}})$.

Collectively, the followers behave according to the Nash noncooperative principle described before. That is to say, they will choose, for each $x \in X$, the joint response vector

$$y^{\text{opt}} := (y_i^{\text{opt}})_{i=1}^{M} \in \prod_{i=1}^{M} Y_i(x)$$

such that for each $i = 1, \ldots, M$

$$y_i^{\text{opt}} \in \operatorname{argmin}\{\theta_i(x, y_i, y_{\neq i}^{\text{opt}}) : y_i \in Y_i(x)\}.$$

For simplicity, we will again assume that $Y_i(x) = \mathbb{R}_+^{m_i}$, for all $x \in X$, so that each follower i solves the parametric optimization problem

$$\begin{array}{ll}\min\limits_{y_i} & \theta_i(x^{\text{given}}, y_i, y_{\neq i}^{\text{given}}) \\ \text{subject to} & y_i \geq 0.\end{array}$$

The joint response vector $y \in \mathbb{R}^p$ is then the solution to the parametric nonlinear complementarity problem

$$y \geq 0, F(x,y) \geq 0, y^T F(x,y) = 0,$$

where $F(x,y) := (F_i(x,y))_{i=1}^M$ with

$$F_i(x,y) := \nabla_{y_i}\theta_i(x,y), \quad i = 1,\ldots,M.$$

Let $f : \mathbb{R}^{n+p} \to \mathbb{R}$ be the leader's cost function which depends on both his own and the follower's strategies. Assume that the feasible set X is described by inequalities of the form $g(x) \geq 0$. The Stackelberg game problem is to determine the solution of the following MPEC in the variables $(x,y) \in \mathbb{R}^{n+p}$:

$$\begin{array}{ll}\min\limits_{x,y} & f(x,y) \\ \text{subject to} & g(x) \geq 0 \\ & y \geq 0, F(x,y) \geq 0, y^T F(x,y) = 0.\end{array} \tag{1.7}$$

Notice that (1.7) is an instance of (1.1) with first-level constraints only involving first-level variables.

1.3 Deregulated electricity markets

1.3.1 History

The United States has recently begun the process of deregulating the electric power industry. In 1998, California, Massachusetts and Rhode Island were the first states to be deregulated. Many other US states and international countries have followed since then, or plan to do so in the near future. This, together with the deregulation of the electricity industry in England and Wales and many other countries in the late 1980's and beginning 1990's, has prompted a great deal of research into modeling power markets and trying to predict how prices will react to the new market structure.

Previously, the market was controlled by the government. In such a regulated market, there is no competition between companies. Each company has a given set of consumers, and the consumer has no choice about which company to use. Prices were set to allow companies to recover their cost and earn a reasonable profit. Thus, the companies' profits were restricted. In a regulated market, a company's goal is to minimize cost, whereas in a deregulated market, their goal is to maximize profit. These different goals will very likely lead to different outcomes. After complete deregulation, the companies are allowed to sell to any consumer at any agreed upon price. The hope is that the resulting competition between companies will prevent the prices from inflating over the fair market value (marginal cost including a reasonable profit).

Early research of the problem focused on simple duopoly models. One key reason for this was that at the beginning of the 1990's, two large power companies in England and Wales, National Power and PowerGen, controlled 79% of the market. PowerGen began to dominate the market, until regulators forced the company to agree to a price cap and downsizing by 15% [Gre96]. During this time, there was a drop in prices for consumers; however, during the same period as deregulation, the price of natural gas also dropped, and it has not been determined which change had the stronger effect on consumer prices

Recently, the focus of research activity has been the development in the US markets, where more players have to be considered and a number of different market designs exist simultaneaously. Most notably, the current energy crisis in California points out that more attention has to paid to details such as market rules and physical properties of electricity networks. The forward market model developed in this thesis is one contribution to this work.

1.3.2 Physical properties of electricity networks

There are two reasons why electricity markets differ significantly from other commodity markets. First, electricity is a non-storable good and all power flows over the same set of power lines. When an appliance or a light is switched on, electricity has to be provided instantaneously to avoid a drop in voltage in the overall network. This leads to a natural interconnection between everyone connected to the network. If there is not enough generation capacity available to satisfy the demand, the network collapses and power is lost for all consumers. To avoid this, the operator initiates blackouts to certain consumers in order to keep the whole network running.

Second, electricity has very special physical properties [SCTB88], which increases the need for a supervising authority coordinating all parties connected to the network. Unlike other commodities distributed over a network (gas, traffic, etc.), if a flow is generated at one node and extracted at another, the flow will be dispersed over all paths between the two nodes. This flow can be determined using Kirchhoff's voltage and current laws. To illustrate Kirchhoff's laws, consider the three-node network shown in Figure 1.1. Assume that a flow of 1 MW is injected at node 1 and

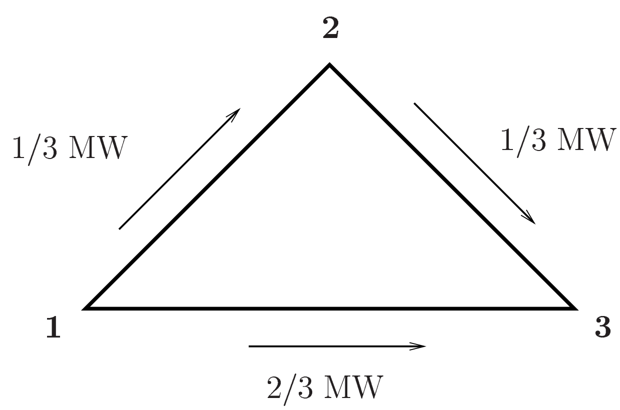

Figure 1.1: Power flow in a three-node network

transmitted to node 3. Ignoring losses in the network, Kirchhoff's law dictates that only 2/3 of this will flow directly on the line connecting node 1 and 3, while the other 1/3 flows on the lines between node 1 and 2 and the line connecting node 2 and 3. This special property of electricity flow is sometimes called "loop-flow phenomenon".

Another consequence of the physical laws is that flows in opposite directions offset each other. So, if in the above example, an additional 1 MW would be transmitted from node 2 to node 3, 2/3 of this would flow on the line from 2 to 3, while the other 1/3 takes the path from 2 to 1 to 3. This offsets the 1/3 MW transmission flowing from 1 to 2 resulting from the first MW transmitted from node 1 to 3. The full impact on the network resulting from 1 MW transmitted from 1 to 3 and 1 MW transmitted from 2 to 3 is depicted in Figure 1.2.

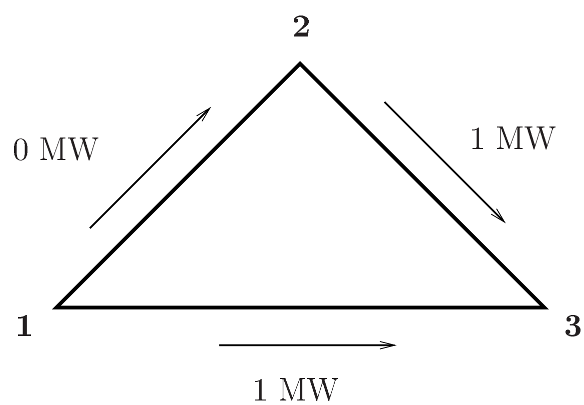

Figure 1.2: Offsetting power flows in the three-node network

Notice that no actual flow exists on the line from node 1 to node 2. Several researchers have studied the economic impact of the loop-flow phenomenon on mostly small (2 to 6 node) networks, see [BBS98, CP96, Sto99a]. Their findings show that limited competition on electricity networks can lead to counter-intuitive outcomes.

1.3.3 Forward markets in electricity

One of the main problems that emerged during the recent California energy crises was the strong reliance on spot markets. At the start of deregulation in California, the largest utility companies agreed to a rate freeze of retail electricity prices. This way, small customers were protected from the volatility of electricity prices in the initial phases of deregulation. The utilities on the other side were forced to buy the electricity needed for their customers in the volatile spot market. This market rule was put in place to enable active trading in the spot market and to avoid possible market manipulations. Long-term contracts between the utilities and generation companies were forbidden in the initial design. As a result, no active forward market for these contracts developed. The expectation was that energy prices would decline and utilities would be able to earn a profit and finance their sunk cost. It turned out that this was a fatal mistake.

Instead of decreasing, electricity prices soared during the last year and utilities were forced to buy energy at a much higher rate than the regulated retail prices they were able to charge their customers. This has led to immense debt and the bankruptcy filing of the largest utility in the state. The unavailability of long-term contracts has been blamed as one of the reasons for this development [FER00]. As a temporary and costly solution, recently the state of California has stepped in and negotiated some long-term contracts while also buying electricity in the spot market for cash-strapped utilities and their customers. The details of these long-term contracts are not publicly known at this point, but the average price of electricity in these contracts is much higher than the average price in previous years.

It is generally accepted that forward contracts have a benefit for the buyer. Locking in rates in a price-volatile market enables better planning and security for the buyer. The seller on the other side usually also benefits from long-term contracts. Under uncertainty of demand, accurate forecasting becomes crucial and forward contracts reduce the risk of over or under capacities in certain periods. In deregulated electricity markets, it is not so clear if this argument still holds for the seller. Generators having the choice between the forward and the spot market choose the profit maximizing combination in a deregulated environment. Where to sell depends mainly on the price paid in these two markets, especially if generation capacity is sparse as in the California market. The forward market model developed in this thesis addresses many of these interesting questions.

1.4 Main contributions

1.4.1 Theoretical and practical investigation of ECOPT

Developing better algorithms for MPECs is the focus in the first part of this dissertation. Our new algorithm ECOPT, which stands for Equilibrium Constraint Optimization, is a sequential quadratic programming (SQP) based algorithm that uses a smoothing function to reformulate the second-level complementarity condition of the MPEC. The key improvement over similar methods proposed by other researchers is the explicit use of second derivatives of the smoothing function.

In Chapter 3 we analyze the theoretical properties of ECOPT. We describe how to handle infeasible QP subproblems, describe the line-search procedure and prove the convergence properties of ECOPT. The implementation of ECOPT and extensive numerical comparison are presented in Chapter 4. We show how to use exact second derivatives of the smoothing function to update the approximation of the Hessian of the Lagrangian. In general, this update will not be positive definite using exact second derivatives. We propose and motivate a special procedure to ensure positive definiteness.

The numerical tests comparing ECOPT with state of the art nonlinear programming methods and other algorithms specifically designed to solve MPEC problems demonstrate the improvements made through the use of the smoothing function in connection with exact second derivatives.

1.4.2 Forward market model for electricity markets

The second part of this dissertation presents a new mathematical model for electricity forward markets. In Chapter 5 we motivate the development with the analysis of a small example. Considering only two firms and one transmission line, we derive the spot-market and forward-market equilibria analytically. The results show an intrinsic relationship between transmission capacity and the development of forward markets.

Building on the small scale case, we develop a general model for electricity forward markets in Chapter 6. The general model consists of the spot-market equilibrium and the firm's forward market problem. The detailed mathematical model of the spot market results in a parametric linear complementarity problem. The spot-market model serves as constraints in the firm's forward market problem. The objective function of each firm in the forward market is a function of forward sales and spot-market variables. The complete forward market model for a given firm is an instance of an MPEC.

We also show how to apply MPEC algorithms to find forward market equilibria. Using ECOPT to solve the firm's forward market problem, we demonstrate how to find equilibria in a six-node electricity network. The study of different levels of transmission capacity extends the results found for the small scale case. We end with a discussion of the application our results to the transmission expansion problem in deregulated electricity markets.

Chapter 2

Mathematical Properties of MPECs

In this chapter we will summarize some of the important mathematical properties of MPECs needed to develop algorithms that find local solutions of MPECs. The problem of finding global solutions is much harder and beyond the scope of this thesis. Even the problem of finding local solution to the MPEC is challenging as we will see in Section 2.1. Specialized constraint qualifications and optimality conditions are needed and will be discussed in Section 2.2

2.1 Why are MPECs difficult?

Recall that we want to solve MPECs of the following form:

$$\begin{array}{rl} \min\limits_{x,y} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & y \geq 0, \;\; F(x,y) \geq 0, \;\; y^T F(x,y) = 0, \end{array} \tag{2.1}$$

where $f : \mathbb{R}^{n+m} \to \mathbb{R}$, $g : \mathbb{R}^{n+m} \to \mathbb{R}^l$ and $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ are twice continuously differentiable functions. More general formulations are possible, see [LPR96], but most cases can be cast in the form (2.1) by including constraints into $g(x,y) \geq 0$.

The lower-level problem in (2.1) is cast as a *parametric nonlinear complementarity problem* in the first-level variables x, which is to find a solution of the following system of equations and inequalities:

$$y_i \geq 0, \;\; F_i(x,y) \geq 0 \;\text{ and }\; y_i F_i(x,y) = 0, \quad \forall i \in I := \{1, \ldots, m\}.$$

Let $(\bar{x}, \bar{y})$ be a stationary point of problem (2.1). Associated with $(\bar{x}, \bar{y})$ we define the following index sets which play an important role for the disussion of the first-order optimality conditions of MPECs in Section 2.2:

$$\begin{aligned}
\bar{\mathcal{I}} &:= \mathcal{I}(\bar{x}, \bar{y}) = \{l \,|\, g(\bar{x}, \bar{y}) = 0\}, \\
\bar{\alpha} &:= \alpha(\bar{x}, \bar{y}) = \{i \,|\, \bar{y}_i > 0 = F_i(\bar{x}, \bar{y})\}, \\
\bar{\beta} &:= \beta(\bar{x}, \bar{y}) = \{i \,|\, \bar{y}_i = 0 = F_i(\bar{x}, \bar{y})\}, \\
\bar{\gamma} &:= \gamma(\bar{x}, \bar{y}) = \{i \,|\, \bar{y}_i = 0 < F_i(\bar{x}, \bar{y})\}.
\end{aligned}$$

In particular, we say that $(\bar{x}, \bar{y})$ is *nondegenerate* if $\bar{\beta} = \emptyset$, i.e., $(\bar{x}, \bar{y})$ satisfies the *strict complementarity* condition $\bar{y}_i + F_i(\bar{x}, \bar{y}) > 0$ for all $i = 1, \ldots, m$.

Ill-conditioned feasible set

As a first approach to solve the smooth problem (2.1), one might be tempted to use a standard nonlinear programming algorithm. The behavior of these algorithms near the solution depends critically on the structure of the feasible set. Under certain conditions, called *constraint qualifications*, convergence proofs for standard nonlinear programming algorithms hold. Unfortunately, the feasible set of the MPEC, see Figure 2.1, is ill-posed, since the following two constraint qualifications which are commonly assumed to prove convergence of standard nonlinear programming algorithms do not hold at any feasible point of the complementarity constraints.

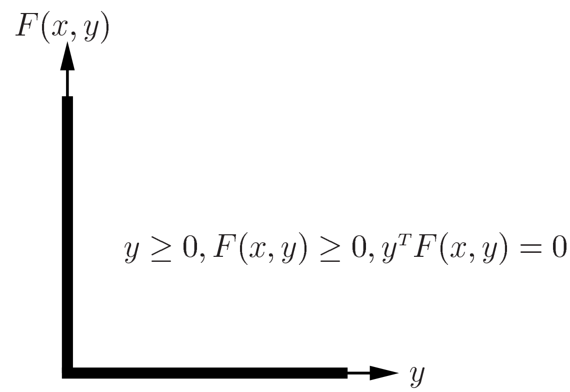

Figure 2.1: Feasible set of the complementarity constraint

Definition 2.1.1. The *Mangasarian-Fromovitz Constraint Qualification (MFCQ)* holds at a point $\bar{x}$ for the constraints

$$g(x) \geq 0,$$
$$h(x) = 0,$$

where $g : \mathbb{R}^n \to \mathbb{R}^l$, $h : \mathbb{R}^n \to \mathbb{R}^p$, if and only if the gradients $\nabla h_j(\bar{x}), j = 1, \ldots, p$, are linearly independent and there exists a $z \in \mathbb{R}^n$ such that

$$\nabla g_k(\bar{x})z > 0, \quad k \in \mathcal{I}(\bar{x})$$
$$\nabla h_j(\bar{x})z = 0, \quad j = 1, \ldots, p$$

and $\mathcal{I}(\bar{x}) := \{k : g_k(\bar{x}) = 0\}$.

Definition 2.1.2. The *Linear Independence Constraint Qualification (LICQ)* holds at a point $\bar{x}$ for the constraints

$$g(x) \geq 0,$$
$$h(x) = 0,$$

where $g : \mathbb{R}^n \to \mathbb{R}^l$, $h : \mathbb{R}^n \to \mathbb{R}^p$, if and only if the gradients $\nabla h_j(\bar{x}), j = 1, \ldots, p$, and $\nabla g_k(\bar{x}), k \in \mathcal{I}(\bar{x})$, are linearly independent.

Lemma 2.1.3. *An optimization problem in* $(x, y) \in \mathbb{R}^{n+m}$ *including the following complementarity system*

$$F(x,y) \geq 0, \tag{2.2}$$

$$y \geq 0, \tag{2.3}$$

$$y^T F(x,y) = 0, \tag{2.4}$$

as part of the constraints does not satisfy the MFCQ at any feasible point.

Proof. Assume, MFCQ is satisfied at a feasible point (x, y); that is, there exists a $(\zeta, \tau) \in \mathbb{R}^{n+m}$ such that

$$F(x,y)^T\zeta + y^T(\nabla_y F(x,y)^T\zeta + \nabla_x F(x,y)^T\tau) = 0, \tag{2.5}$$

$$\nabla_y F(x,y)^T\zeta + \nabla_x F(x,y)^T\tau > 0, \tag{2.6}$$

$$\zeta > 0, \tag{2.7}$$

and

$$\begin{pmatrix} F(x,y) + \nabla_y F(x,y)^T y \\ \nabla_x F(x,y)^T y \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.8}$$

(There is only one equality constraint, so the independence of the gradients of the equality constraints reduces to the nonvanishing of the gradient.)

From (2.2) and (2.7) we get

$$F(x,y)^T\zeta \geq 0$$

and (2.3) and (2.6) give

$$y^T(\nabla_y F(x,y)^T\zeta + \nabla_x F(x,y)^T\tau) \geq 0.$$

Together with (2.5) these two inequalities imply

$$F(x,y)^T\zeta = 0, \tag{2.9}$$

and

$$y^T(\nabla_y F(x,y)^T \zeta + \nabla_x F(x,y)^T \tau) = 0. \tag{2.10}$$

Considering (2.2), (2.7) and (2.9) we get

$$F(x,y) = 0,$$

and from (2.3), (2.6) and (2.10) it follows that

$$y = 0.$$

The last two equalities lead to a contradiction of (2.8). Therefore, MFCQ is not satisfied at any feasible point of an optimization problem including a parametric complementarity system as part of the constraints. □

As a consequence, the commonly assumed LICQ also fails.

Corollary 2.1.4. *An optimization problem in* $(x,y) \in \mathbb{R}^{n+m}$ *including the complementarity system* (2.2), (2.3), (2.4) *as part of the constraints does not satisfy the LICQ at any feasible point.*

The convergence analysis of many algorithms for standard NLP relies, aside some other assumptions, critically on the MFCQ at the limit point. Without MFCQ, convergence proofs do not hold any more and numerical problems are to be expected. In particular, the limit point is not guaranteed to be a stationary point or local minimum any more. From our experience, see Chapter 4, and some other reports [JR99, LF00], it is not clear how much this matters in practice. Other factors, such as bad scaling or inappropriate reformulations, could be equally important reasons for failures in earlier tests.

2.2 Constraint qualifications and optimality conditions

In general, the first-order conditions for the MPEC (2.1) are very complicated due to the combinatorial nature of the constraint system. The first approach to derive first-order optimality conditions will be the partitioning of the feasible set based on the complementarity conditions. Under some constraint qualification, the combinatorial problem reduces to checking stationarity of one particular nonlinear program.

2.2.1 B-stationarity

The thorough treatment of stationarity concepts for MPECs in [LPR96] is based on B-stationarity and piecewise partitioning of the tangent cone of the constraints. Let S be an arbitrary subset of $\mathbb{R}^n$. We call $\bar{u} \in S$ a *B-stationary point* of the minimization problem

$$\begin{array}{rl} \min\limits_{u} & f(u) \\ \text{subject to} & u \in S, \end{array}$$

if it satisfies

$$\nabla f(\bar{u})^T d \geq 0, \quad \forall d \in \mathcal{T}_S(\bar{u}), \tag{2.11}$$

where $\mathcal{T}_S(\bar{u})$ denotes the tangent cone of S at $\bar{u}$ that consists of all vectors d for which there exists a sequence $\{u^\nu\} \in S$ converging to $\bar{u}$ and a sequence of τ_ν of positive numbers converging to zero such that $d = \lim_{\nu\to\infty}(u^\nu - \bar{u})/\tau_\nu$. It is easy to see that every local minimum of (2.1) is a B-stationary point.

In general, condition (2.11) is difficult to deal with due to the possibly complicated structure of the tangent cone $\mathcal{T}_S(\bar{u})$. This difficulty can be overcome through the use of the linearized cone $\mathcal{T}_{lin}(\bar{u})$ and a primal-dual characterization of B-stationarity. We will specify this approach for the MPEC (2.1).

2.2.2 Piecewise stationarity

Recall the problem statement

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{subject to} \quad & g(x,y) \geq 0, \\ & y \geq 0, \;\; F(x,y) \geq 0, \;\; y^T F(x,y) = 0, \end{aligned}$$

and the index sets $\bar{\alpha}, \bar{\beta}$, and $\bar{\gamma}$ associated to a stationary point $(\bar{x}, \bar{y})$.

We define the following MPEC-Lagrangian for $(x, y, \xi, \pi) \in \mathbb{R}^{n \times m \times l \times m}$:

$$\mathcal{L}^{MPEC} := f(x,y) - \xi^T g(x,y) - \pi^T F(x,y).$$

The vectors ξ and π can be thought of as the MPEC multipliers of the constraints $g(x,y) \geq 0$ and $F(x,y) \geq 0$. Notice that there are no multipliers for the constraints $y \geq 0$ and the complementarity constraints $y^T F(x,y) = 0$ in the function $\mathcal{L}^{MPEC}$. The fact that no multiplier is needed for the latter complementarity constraint is a special feature of the MPEC that distinguishes it from standard nonlinear programming problems. Instead, the complementarity condition $y^T F(x,y) = 0$ will be decomposed into a disjunction of finitely many systems of equalities and inequalities depending on the subset of degenerate indices $\bar{\beta}$.

For each subset β_1 of the degenerate index set $\bar{\beta}$, consider the following nonlinear program in the variables (x, y):

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{subject to} \quad & g(x,y) \geq 0 \\ & y_i = 0, && i \in \beta_1, \\ & F_i(x,y) \geq 0, && i \in \beta_1, \\ & y_i \geq 0, && i \in \bar{\beta} \backslash \beta_1, \\ & F_i(x,y) = 0, && i \in \bar{\beta} \backslash \beta_1. \end{aligned} \tag{2.12}$$

For each index set β_1 the KKT conditions of (2.12) are

$$
\begin{aligned}
&\nabla_x f(\bar{x},\bar{y}) - g'_x(\bar{x},\bar{y})^T\xi - F'_x(\bar{x},\bar{y})^T\eta = 0,\\
&\nabla_y f(\bar{x},\bar{y}) - g'_y(\bar{x},\bar{y})^T\xi - F'_y(\bar{x},\bar{y})^T\eta - \pi = 0,\\
&0 \le g(\bar{x},\bar{y}) \perp \xi \ge 0,\\
&\bar{y}_i = 0, \quad 0 \le F_i(\bar{x},\bar{y}) \perp \eta_i \ge 0, \quad i \in \beta_1,\\
&F_i(\bar{x},\bar{y}) = 0, \quad 0 \le \bar{y}_i \perp \pi_i \ge 0, \quad i \in \bar{\beta}\backslash\beta_1.
\end{aligned}
\tag{2.13}
$$

A feasible point $(\bar{x},\bar{y})$ is called *piecewise stationary* if for each $\beta_1 \subseteq \beta$, there exist multipliers $\bar{\xi}$, $\bar{\eta}$, $\bar{\pi}$ such that the KKT conditions (2.13) are satisfied.

In order to derive the primal-dual characterization of stationarity, the structure of the linearized cone $\mathcal{T}_{lin}(\bar{x},\bar{y})$ has been studied in [LPR96]. The following constraint qualification for MPECS is similar to the Abadie constraint qualification [Aba67] in NLP:

Definition 2.2.1. We say the *full constraint qualification (full CQ)* holds for the MPEC at $(\bar{x},\bar{y})$ if

$$\mathcal{T}_{lin}(\bar{x},\bar{y}) = \mathcal{T}_S(\bar{x},\bar{y}).$$

Shown in [LPR96], the following theorem uses the structure of the linearized cone and a theorem of the alternative.

Theorem 2.2.2. *Let $f : \mathbb{R}^{n+m} \to \mathbb{R}$, $g : \mathbb{R}^{n+m} \to \mathbb{R}^l$ and $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ be continuously differentiable functions. If the full CQ holds for the MPEC* (2.1) *at the feasible vector $(\bar{x},\bar{y})$, then $(\bar{x},\bar{y})$ is a B-stationary point of* (2.1) *if and only if it is piecewise stationary.*

The complexity of the necessary conditions of the MPEC (2.1) is set by the amount of the degeneracy in the lower-level NCP(F). Indeed, the number of sets β_1 in Theorem 2.2.2 is $2^{|\bar{\beta}|}$. It would be desirable to overcome the combinatorial nature of the first-order conditions. This is possible by invoking some stronger conditions.

Relaxed NLP

One rather strong condition is nondegeneracy of the lower-level NCP(F). Since $\bar{\beta} = \emptyset$ in this case, it follows directly from Theorem 2.2.2 that we would only have to deal with one KKT system for optimality. Since degeneracy can be expected in many practical cases, this is not very promising.

Another approach builds on the LICQ and the relaxed NLP.

Definition 2.2.3. We say the *Linear Independence Constraint Qualification (LICQ)* holds for the MPEC at $(\bar{x}, \bar{y})$ if the vectors

$$\begin{aligned} \{(0, e^i) \in \mathbb{R}^n \times \mathbb{R}^m : i \in \bar{\alpha} \cup \bar{\beta}\} \quad &\cup \quad \{\nabla F_i(\bar{x}, \bar{y}) : i \in \bar{\beta} \cup \bar{\gamma}\} \\ &\cup \quad \{\nabla g_l(\bar{x}, \bar{y}) : l \in \bar{\mathcal{I}}\} \end{aligned}$$

are linearly independent.

The *relaxed NLP* corresponding to the MPEC (2.1) is defined by

$$\begin{array}{rll} \min\limits_{x,y} & f(x, y) & \\ \text{subject to} & g(x, y) \geq 0 & \\ & F_i(x, y) = 0, & i \in \bar{\alpha} \\ & y_i \geq 0, & i \in \bar{\beta} \\ & F_i(x, y) \geq 0, & i \in \bar{\beta} \\ & y_i = 0, & i \in \bar{\gamma}. \end{array} \tag{2.14}$$

The KKT conditions for the relaxed NLP are

$$\begin{aligned} &\nabla_x f(\bar{x}, \bar{y}) - g'_x(\bar{x}, \bar{y})^T \xi - F'_x(\bar{x}, \bar{y})^T \eta = 0, \\ &\nabla_y f(\bar{x}, \bar{y}) - g'_y(\bar{x}, \bar{y})^T \xi - F'_y(\bar{x}, \bar{y})^T \eta - \pi = 0, \\ &0 \leq g(\bar{x}, \bar{y}) \perp \xi \geq 0, \\ &F_i(\bar{x}, \bar{y}) = 0, \quad i \in \bar{\alpha} \\ &0 \leq \bar{y}_i \perp \pi_i \geq 0, \quad i \in \bar{\beta}, \\ &0 \leq F_i(\bar{x}, \bar{y}) \perp \eta_i \geq 0, \quad i \in \bar{\beta}, \\ &\bar{y}_i = 0, \qquad i \in \bar{\gamma}. \end{aligned} \tag{2.15}$$

Under some constraint qualification, any B-stationary point of the relaxed problem satisfies the KKT conditions (2.15) along with some Lagrange multipliers ξ, η and π. Furthermore, if the LICQ holds at $(\bar{x}, \bar{y})$, then any KKT point of (2.14) is also a B-stationary point for (2.1); see [LPR96].

Proposition 2.2.4. *If a feasible point* $(\bar{x}, \bar{y})$ *of problem* (2.1) *satisfies the KKT conditions* (2.15) *for the relaxed problem* (2.14) *and if the LICQ holds at* $(\bar{x}, \bar{y})$*, then* $(\bar{x}, \bar{y})$ *is a B-stationary point of the MPEC* (2.1)

Applications of this result have been used by some authors in their work. For example, the KKT conditions of the relaxed NLP can be used as a stopping criterion within an algorithm [JR99, LPR98] and as a means to prove convergence of a smoothing method [FP99] to the solution of the MPEC (2.1).

2.2.3 Weaker stationarity concepts

In general, algorithms for MPECs can not be expected to converge to B-stationary points in all cases. Convergence can sometimes only be proven to points satisfying weaker stationarity concepts. Following [SS00], a feasible point $(\bar{x}, \bar{y})$ of the MPEC (2.1) is called *weakly stationary* if there exists MPEC multipliers $\bar{\xi} \geq 0$, $\bar{\eta}$, and $\bar{\pi}$ satisfying

$$\begin{pmatrix} \nabla_x f(\bar{x}, \bar{y}) \\ \nabla_y f(\bar{x}, \bar{y}) \end{pmatrix} - \sum_{i \in \bar{\mathcal{I}}} \bar{\xi}_i \begin{pmatrix} \nabla g_x(\bar{x}, \bar{y}) \\ \nabla g_y(\bar{x}, \bar{y}) \end{pmatrix} + \sum_{j \in \bar{\alpha} \cup \bar{\beta}} \bar{\eta}_j \begin{pmatrix} \nabla F_x(\bar{x}, \bar{y}) \\ \nabla F_y(\bar{x}, \bar{y}) \end{pmatrix} + \sum_{k \in \bar{\beta} \cup \bar{\gamma}} \bar{\pi}_k \begin{pmatrix} 0 \\ e \end{pmatrix} = 0.$$

Notice that there are no sign constraints on the multipliers corresponding to the complementarity terms. Imposing additional sign constraints yields stronger stationarity concepts. The strongest of these concepts is

Strong Stationarity: $\bar{\eta}_m \bar{\pi}_m \geq 0$ for all $m \in \bar{\beta}$.

Notice that strong stationarity is equivalent to the KKT conditions (2.15) of the relaxed NLP. Therefore, as argued earlier, strong stationarity implies B-stationarity and is equivalent to it in the presence of LICQ.

2.2.4 Nonsmooth optimization

Some algorithms for MPECs are based on reformulations involving nonsmooth functions. Since the convergence analysis of our algorithm is based on such a nonsmooth reformulation, we will give a brief description of stationarity for nonsmooth programming. The nonsmooth program is given by:

$$\begin{aligned} \min_{u} \quad & f(u) \\ \text{subject to} \quad & g(u) \geq 0 \\ & h(u) = 0, \end{aligned} \tag{2.16}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^l$, $h : \mathbb{R}^n \to \mathbb{R}^m$ are locally Lipschitz. As an extension to nonsmooth programs, one can define generalized KKT conditions for (2.16).

Definition 2.2.5. The point $\bar{u} \in \mathbb{R}^n$ is said to be a *generalized stationary point* of (2.16) if there exists a KKT multiplier vector $(\lambda_g, \lambda_h) \in \mathbb{R}^{l+m}$ such that the following *generalized Karush-Kuhn-Tucker (GKKT)* conditions hold:

$$\begin{aligned} 0 \in \partial f(\bar{u}) - \partial g(\bar{u})^T \lambda_g + \partial h(\bar{u})^T \lambda_h, \\ 0 \leq g(\bar{u}) \perp \lambda_g \geq 0, \\ h(\bar{u}) = 0, \end{aligned}$$

where ∂ denotes the Clarke generalized gradient for a scalar function and the Clarke generalized Jacobian for a vector-valued function [Cla90].

If f, g, and h happen to be smooth at $\bar{u}$, then the GKKT conditions reduce to the usual KKT conditions for smooth nonlinear programming problems.

Chapter 3

MPEC Algorithms

In this chapter we consider algorithms to find local solutions of MPECs. In Section 3.1, we present several existing algorithms for MPECs. Other algorithms for solving MPECs and the related topic of bilevel optimization can be found in the monographs [LPR96, OKZ98] and the review article [VC94].

The description of our algorithm ECOPT for equilibrium-constrained optimization together with convergence results is given in Section 3.3. The implementation details of ECOPT and extensive numerical results comparing ECOPT with some algorithms from Section 3.1 as well as standard nonlinear optimization packages are presented in Chapter 4.

3.1 Existing algorithms for MPECs

3.1.1 Piecewise sequential quadratic programming (PSQP)

The basic idea behind PSQP is to solve for the piecewise stationarity conditions given by (2.12), which was first suggested in a different context in [KS86]. We will describe the local algorithm presented in [LPR98] to solve NCP-constrained MPECs with affine first-level constraints:

$$\begin{aligned}\min_{x,y} \quad & f(x,y) \\ \text{subject to} \quad & Gx + Hy + a \leq 0 \\ & y \geq 0, \;\; F(x,y) \geq 0, \;\; y^T F(x,y) = 0,\end{aligned}$$

where $f : \mathbb{R}^{n+m} \to \mathbb{R}$, $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ are twice continuously differentiable, $G \in \mathbb{R}^{l\times n}$, $H \in \mathbb{R}^{l\times m}$, and $a \in \mathbb{R}^l$. Other version of PSQP are discussed in [LPR96, Ral96]. Implementation issues, in particular a globalization via a line search, different stopping criteria, and numerical results for PSQP are contained in [JR99].

The main computational work in PSQP is to solve the follwing QP in the variable $dz := (dx, dy)$

$$\begin{aligned}\min_{dz\in\mathbb{R}^{n+m}} \quad & \nabla f(z^k)^T dz + \frac{1}{2}(dz)^T(\nabla^2 f(z^k) + \sum_i \pi_i^k \nabla^2 F_i(z^k))dz \\ \text{subject to} \quad & G(x^k + dx) + H(y^k + dy) + a \leq 0 \\ & y_i^k + dy_i \geq 0, \;\text{ for } i \in \mathcal{I}_1 \\ & F_i(x^k, y^k) + \nabla F_i(x^k, y^k)d = 0, \;\text{ for } i \in \mathcal{I}_1 \\ & y_i^k + dy_i = 0, \;\text{ for } i \in \mathcal{I}_2 \\ & F_i(x^k, y^k) + \nabla F_i(x^k, y^k)d \geq 0, \;\text{ for } i \in \mathcal{I}_2.\end{aligned} \tag{3.1}$$

The Piecewise Sequential Programming Algorithm (PSQP)

Step 0. (Initialization) Choose a vector $z^0 = (x^0, y^0) \in Z$, and $(\xi^0, \pi^0) \in \mathbb{R}^{l+m}$, where ξ^0 and π^0 are Lagrange multipliers corresponding to the constraints

$$Gx + Hy + a \leq 0$$

and

$$F(x,y) \geq 0.$$

Z denotes the first-level feasible set $\{(x,y)\,|\,Gx + Hy + a \leq 0\}$. Set $k := 0$.

Step 1. (Main Computation) Let $\mathcal{I}_1 \cup \mathcal{I}_2$ be an arbitrary partition of $\{1, \ldots, m\}$ satisfying $\mathcal{I}_1 \supseteq \{i \,|\, y_i^k > F_i(x^k, y^k)\}$, and $\mathcal{I}_2 \supseteq \{i \,|\, y_i^k < F_i(x^k, y^k)\}$. Solve the QP (3.1).

Let $(dz^k, \xi^{k+1}, \pi^{k+1})$ be a KKT tuple of (3.1), where ξ^{k+1} and π^{k+1} are the multipliers corresponding to the respective constraints $z^k + d^k \in Z$ and the constraints on $F(z^k) + \nabla F(z^k)d^k$.

Step 2. (Termination check and update) Set $z^{k+1} := z^k + d^k$. Terminate if a prescribed termination criterion is satisfied. Otherwise set $k := k + 1$ and return to Step 1.

Notice that in the above algorithm the multipliers ξ are not used. These multipliers would be used if the first-level feasible set would be given by general nonlinear twice continuously differentiable functions. One possible termination criterion in Step 2 of PSQP is based on piecewise stationarity. In this version, the algorithm stops if the KKT conditions (2.13) are satisfied for all subsets β_1 of the degenerate set β. This and other termination criteria are discussed in [JR99].

The local convergence properties of PSQP are summarized in the following result from [LPR98].

Theorem 3.1.1. *Suppose f and F are twice continuously differentiable and $\bar{z} = (\bar{x}, \bar{y})$ is stationary point with second-order sufficient conditions holding for all the nonlinear programs*

$$
\begin{array}{rl}
\min\limits_{x,y} & f(x, y) \\
\text{subject to} & Gx + Hy + a \leq 0 \\
& F_i(x, y) = 0, \quad y_i \geq 0, \quad \text{for } i \in \mathcal{I}_1 \\
& F_i(x, y) \geq 0, \quad y_i = 0, \quad \text{for } i \in \mathcal{I}_2,
\end{array}
$$

where $\mathcal{I}_1 \cup \mathcal{I}_2$ partition the set $\{1, \ldots, m\}$ such that $\mathcal{I}_1 \supseteq \{i \,|\, \bar{y}_i > F_i(\bar{x}, \bar{y})\}$, and $\mathcal{I}_2 \supseteq \{i \,|\, \bar{y}_i < F_i(\bar{x}, \bar{y})\}$. Suppose $(\bar{\xi}, \bar{\pi})$ is the unique KKT multiplier at $\bar{z}$ for all *these nonlinear programs. Then $\bar{z}$ is a strict local minimizer for the MPEC. Moreover, for any (z^0, ξ^0, π^0) near $(\bar{z}, \bar{\xi}, \bar{\pi})$, PSQP is well defined and produces a sequence $\{(z^k, \xi^k, \pi^k)\}$ that converges Q-superlinearly to $(\bar{z}, \bar{\xi}, \bar{\pi})$. If, in addition, $\nabla^2 f$ and $\nabla^2 F$ are Lipschitz near $\bar{z}$, then the rate of convergence is Q-quadratic.*

3.1.2 Penalty Interior-Point Algorithm (PIPA)

The algorithm PIPA is the most extensively treated algorithm in the monograph [LPR96]. Numerical tests and discussions of implementation issues also appear in the publications [JR99, DF99]. The version of PIPA we will describe solves MPECs with mixed complementarity constraints of the following form:

$$\begin{aligned} \min_{x,y,w,z} \quad & f(x, y, w, z) \\ \text{subject to} \quad & Gx \leq a \\ & F(x, y, w, z) = 0 \\ & y \geq 0, \;\; w \geq 0, \;\; y^T w = 0, \end{aligned} \tag{3.2}$$

where $x \in \mathbb{R}^n$, $y, w \in \mathbb{R}^m$, $z \in \mathbb{R}^l$, $a \in \mathbb{R}^p$, $G \in \mathbb{R}^{p \times n}$, and $f : \mathbb{R}^{n+2m+l} \to \mathbb{R}$, $F : \mathbb{R}^{n+2m+l} \to \mathbb{R}^{m+l}$ are once respectively twice continuously differentiable functions. Notice that the first-level constraints $Gx \leq a$ are only on x. Versions of PIPA for joint first-level constraints have been considered, but their global convergence properties are open [JR99, DF99] or they are cast for the special case of second-level LCP constraints [FP98]. Second-level NCP constraints are given in (3.2) by a suitable definition of $F(x, y, w)$.

The idea of PIPA is to replace the complementarity condition

$$y \geq 0, \;\; w \geq 0, \;\; y^T w = 0$$

by

$$y \geq 0, \;\; w \geq 0, \;\; y \circ w = \mu e,$$

where $y \circ w := (y_i w_i)$ denotes the Hadamard product of y and w, and to trace the interior path

$$\{(x(\mu), y(\mu), w(\mu), w(\mu)) \in \mathbb{R}^{n+2m+l} \mid \mu > 0\}$$

of the parameterized problem

$$\begin{aligned} \min_{x,y,w,z} \quad & f(x, y, w, z) \\ \text{subject to} \quad & Gx \le a \\ & F(x, y, w, z) = 0 \\ & y \ge 0, \;\; w \ge 0, \;\; y \circ w = \mu e, \end{aligned} \tag{3.3}$$

when the positive parameter μ is decreased to zero.

Similar to other interior point methods, problem (3.3) is not solved to full accuracy for each μ; instead only one SQP step is calculated, and then the parameter μ is updated. The solution of the following quadratic program is used as a search direction

$$\begin{aligned} \min_{(dx,dy,dw,dz)\in\mathbb{R}^{n+2m+l}} \quad & \nabla f(x^k, y^k, w^k, z^k)^T \begin{pmatrix} dx \\ dy \\ dw \\ dz \end{pmatrix} + \frac{1}{2}(dx)^T B_k dx \\ \text{subject to} \quad & G(x^k + dx) \le a \\ & |dx| \le \sqrt{c(\|F(x^k, y^k, w^k, z^k)\| + (w^k)^T y^k)}\, e \\ & \nabla F(x^k, y^k, w^k, z^k)^T \begin{pmatrix} dx \\ dy \\ dw \\ dz \end{pmatrix} = -F(x^k, y^k, w^k, z^k) \\ & w^k \circ dy + y^k \circ dw = -w^k \circ y^k + \sigma_k \mu_k e, \end{aligned} \tag{3.4}$$

where $\sigma_k \in (0, 1)$ is a given scalar which balances between a pure Newton search direction ($\sigma_k = 0$) and a purely centralized direction ($\sigma_k = 1$), $c > 0$ is a scalar controlling the step size and $\mu_k := \frac{(w^k)^T y^k}{m}$. The matrix $B_k \in \mathbb{R}^{n\times n}$ is symmetric positive definite. Under certain conditions, the solution of (3.4) is unique.

Definition 3.1.2. A partitioned matrix $Q = [\ A\ \ B\ \ C\ ]$ is said to have the *mixed P_0 property* if C has full column rank and the implication

$$\left.\begin{array}{c} Ar + Bs + Ct = 0 \\ (r,s) \neq 0 \end{array}\right\} \Rightarrow r_i s_i \geq 0 \text{ for some } i \text{ with } |r_i| + |s_i| > 0$$

holds.

Proposition 3.1.3 ([LPR96] Lemma 6.1.7). *Let $(x^k, y^k, w^k, z^k) \in X \times \mathbb{R}^{2m}_{++} \times \mathbb{R}^l$ and $B_k \in \mathbb{R}^{nt \times n}$ be a symmetric positive definite matrix. Suppose that the partitioned matrix*

$$[\ \nabla_y F(x,y,w,z)\ \ \nabla F_w(x,y,w,z)\ \ \nabla F_z(x,y,w,z)\ ]$$

has the mixed P_0 property. Then the QP (3.4) *has a unique solution.*

The Penalty Interior-Point Algorithm (PIPA)

Step 0. (Initialization) Choose a vector $(x^0, y^0, w^0, z^0) \in X \times \mathbb{R}^{2m}_{++} \times \mathbb{R}^l$, where $X \subseteq \mathbb{R}^n$ denotes the feasible set of $Gx \leq a$. Let $c > 0$, $\bar{\sigma}, \rho_1, \gamma, \gamma', \epsilon \in (0,1)$ and $\alpha_{-1} > 1$. Let $\bar{\rho} \in (0,1)$ and $\sigma_0 \in [0,1)$ satisfy the conditions

$$\bar{\rho}\frac{(y^k)^T w^k}{m} \leq \min_{1 \leq i \leq m} y_i^k w_i^k \quad \text{and} \quad \sigma_0 \leq \min\{\bar{\sigma}, \bar{\rho}\}.$$

Let $B_0 \in \mathbb{R}^{n \times n}$ be a given symmetric positive semidefinite matrix. Set $k := 0$.

Step 1. (Direction generation and penalty update) Solve the QP (3.4) to determine the unique solution (dx^k, dy^k, dw^k, dz^k). Update the penalty parameter $\alpha_k = \alpha_{k-1}^{\ell}$ where ℓ is the smallest nonnegative integer such that

$$\nabla f(x^k, y^k, w^k, z^k) \begin{pmatrix} dx^k \\ dy^k \\ dw^k \\ dz^k \end{pmatrix} + \alpha_{k-1}^{\ell} \nabla \phi(x^k, y^k, w^k, z^k) \begin{pmatrix} dx^k \\ dy^k \\ dw^k \\ dz^k \end{pmatrix}$$
$$< -\phi(x^k, y^k, w^k, z^k),$$

where $\phi(x,y,w,z) := F(x,y,w,z)^T F(x,y,w,z) + w^T y$.

Step 2. (Step size determination) Define the linear function

$$g_k(\tau) := (1-\bar{\rho})\sigma_k\mu_k + \tau\left(\min_{1\le i\le m} dy_i^k dw_i^k - \bar{\rho}\frac{(dy^k)^T dw^k}{m}\right).$$

Let $\bar{\tau}$ be the (unique) root of the function $g_k(\tau)$ for $\tau \in (0,1]$ if this root exists, otherwise let $\bar{\tau}_k := 1$. If $\bar{\tau}_k = 1$, redefine $\bar{\tau}_k$ to be $1-\epsilon$.

Let $\tau_k := \bar{\tau}_k\rho_1^\ell$, where ℓ is the smallest nonnegative integer such that

$$\phi(x^k+\tau_k dx^k, y^k+\tau_k dy^k, w^k+\tau_k dw^k, z^k+\tau_k dz^k) \le \phi(x^k, y^k, w^k, z^k)$$

and

$$\begin{aligned} &P_{\alpha_k}((x^k,y^k,w^k,z^k)+\tau_k(dx^k,dy^k,dw^k,dz^k)) - P_{\alpha_k}(x^k,y^k,w^k,z^k) \\ &\le \gamma'\tau_k\nabla P_{\alpha_k}(x^k,y^k,w^k,z^k)\begin{pmatrix} dx^k \\ dy^k \\ dw^k \\ dz^k \end{pmatrix}. \end{aligned}$$

Note that this implies by the penalty update rule in Step 1 that we have computed a descent direction for the penalty function

$$P_\alpha(x,y,w,z) := f(x,y,w,z) + \alpha\phi(x,y,w,z).$$

Step 3. (Termination check) Test the new iterate

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ w^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} x^k \\ y^k \\ w^k \\ z^k \end{pmatrix} + \tau_k\begin{pmatrix} dx^k \\ dy^k \\ dw^k \\ dz^k \end{pmatrix}$$

for termination. If a prescribed stopping rule is not fulfilled, choose a symmetric positive definite matrix B_{k+1} and a scalar $\sigma_{k+1} \in (0, \sigma_k]$. Increase k to $k+1$ and return to Step 1.

We will now summarize the main convergence properties of PIPA shown in [LPR96].

Proposition 3.1.4. *If the partitioned matrix*

$$[\ \nabla_y F(x,y,w,z)\ \ \nabla F_w(x,y,w,z)\ \ \nabla F_z(x,y,w,z)\ ]$$

has the mixed P_0 property for all vectors $(x,y,w,z) \in X \times \mathbb{R}^{2m}_{++} \times \mathbb{R}^l$ and the level set

$$\{(x,y,w,z) \in X \times \mathbb{R}^{2m}_{++} \times \mathbb{R}^l \,|\, \phi(x,y,w,z,) \leq \phi(x^0,y^0,w^0,z^0)\}$$

is bounded, then the penalty interior point algorithm generates a well-defined, bounded sequence. In addition, the sequence $\{dx^k\}$ is also bounded.

Before we state the convergence properties of PIPA, we will introduce the following two conditions:

(SC) (Strict Complementarity) $y^* + w^* > 0$. This is equivalent to the set of degenerate indices β being empty.

(NS) (Non-singularity) The submatrix

$$\begin{pmatrix} \nabla_y F(x^*,y^*,w^*,z^*) & \nabla F_w(x^*,y^*,w^*,z^*) & \nabla F_z(x^*,y^*,w^*,z^*) \\ \begin{matrix} w_1^* & 0 & \cdots & 0 \\ 0 & w_2^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_m^* \end{matrix} & \begin{matrix} y_1^* & 0 & \cdots & 0 \\ 0 & y_2^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & y_m^* \end{matrix} & 0 \end{pmatrix}$$

corresponding to the second-level constraints in the QP subproblem (3.4) is nonsingular.

We state two different convergence theorems. The first considers the case where the penalty parameter tends to infinity and the second where it remains constant after a finite number of iterations.

Theorem 3.1.5. *Suppose there exists a scalar* $c_2 > 0$ *such that*

$$0 \leq x^T B_k x \leq c_2 \|x\|^2, \forall x \in \mathbb{R}^n, \forall k \in \mathbb{N}.$$

If $\lim_{k\to\infty} \alpha_k = \infty$ *then*

$$\lim_{k(\in\mathcal{K})\to\infty} dx^k = 0,$$

where $\mathcal{K} := \{k \,|\, \alpha_{k-1} < \alpha_k\}$. *Furthermore, if* (x^*, y^*, w^*, z^*) *is any limit point of the subsequence* $\{(x^k, y^k, w^k, z^k) \,|\, k \in \mathcal{K}\}$ *satisfying the assumptions (SC) and (NS), then* (x^*, y^*, w^*, z^*) *is a stationary point.*

Theorem 3.1.6. *Suppose there exists a scalar* $c_2 > 0$ *such that*

$$0 \leq x^T B_k x \leq c_2 \|x\|^2, \forall x \in \mathbb{R}^n, \forall k \in \mathbb{N},$$

and both sequences $\{\alpha_k\}$ *and* $\{\sigma_k^{-1}\}$ *are bounded, then every accumulation point* (x^*, y^*, w^*, z^*) *of the generated sequence that satisfies the assumptions (SC) and (NS) is stationary.*

3.1.3 Implicit function based approaches

A different concept of solving MPECs is to use an implicit function approach [OKZ98]. If the state variables y are uniquely determined by the design variables x, i.e., for any fixed x there exists an implicit function $y(x)$, the problem can be reformulated as

$$\begin{aligned} \min_{x} \quad & f(x, y(x)) \\ \text{subject to} \quad & x \in X_{ad}, \end{aligned} \tag{3.5}$$

where X_{ad} denotes the feasible of the first-level constraints. Notice that the problem only depends on the design variables.

A problem with this approach is that even if the implicit function $y(x)$ exists, one cannot expect it to be everywhere differentiable, although, under reasonable assumptions, it will be Lipschitz continuous. Using the formulation (3.5), some authors suggest the use of a bundle technique [OZ95, OKZ98], while others [PHR91] compute a descent direction of $f(x, y(x))$ and determine the next iterate using an Armijo line search.

The major work in the bundle method presented in [OZ95] is the computation of the sub-gradient of the implicit function. The approach presented in [PHR91] is more conceptual, the algorithm presented by the authors involves the solution of mathematical programs with mixed complementarity constraints and the line search requires the repeated evaluation of $y(x)$, involving the solution of a variational inequality.

We will not go into greater detail of the implicit function approach and refer the interested reader to the above mentioned publications. Numerical tests of the bundle algorithm can be found in [OZ95, OKZ98, DF99].

3.2 Sequential Quadratic Programming (SQP)

Before presenting our SQP algorithm for MPECs in Section 3.3, we give a short overview of SQP methods.

3.2.1 Background

The original SQP method due to Wilson [Wil63] generates a sequence of search directions d^k, each of which is the solution to a QP subproblem that is a local approximation of the convex nonlinear optimization problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & g(x) \geq 0, \end{aligned} \tag{3.6}$$

where f and g are assumed to be twice continuously differentiable functions, with f being a convex and g a concave function. In the original version, the SQP algorithm starts from a point x^0, and at iteration k, takes a full step $x^{k+1} = x^k + d^k$.

The QP subproblem corresponding to (3.6) has the form

$$\begin{aligned} \min_{d} \quad & \nabla f(x^k)^T d + \frac{1}{2} d^T H^k d \\ \text{subject to} \quad & \nabla g(x^k) d + g(x^k) \geq 0, \end{aligned} \tag{3.7}$$

where $H^k = \nabla^2_{xx} \mathcal{L}(x^k, \lambda)$ is the Hessian of the Lagrangian

$$\mathcal{L}(x, \lambda) := f(x) - \lambda^T c(x),$$

at the current point x^k, with λ being the latest estimate for the Lagrange multipliers, and $\nabla g(x)$ is the Jacobian of the constraints. Notice that all subproblems (3.7) are feasible if g is assumed to be concave. It is often also assumed that H^k is positive definite, in which case the QP (3.7) has a unique solution.

The linear constraints of (3.7) are a first-order approximation to the nonlinear constraints at the current iterate x^k. The quadratic objective of the QP models the curvature of the Lagrangian (and not only of the objective function f). In the neighborhood of a solution and under certain assumptions, the quadratic model is a very good approximation to the original problem (3.6). In other words, $d^k = 0$ is the solution to (3.7) at a KKT point x^* and λ^* of (3.6).

3.2.2 Practical SQP methods

In the original form, Wilson's SQP method requires exact second derivatives and is not guaranteed to converge, for the same reasons Newton's method for unconstrained optimization may fail to converge. A practical SQP method has to address these two problems first.

Even if exact second derivatives of f and g are available, they might be difficult to evaluate, and the Hessian update H^k is not guaranteed to be positive definite. Considering both of these issues, Murray [Mur69] suggested to replace the Hessian of the Lagrangian by a quasi-Newton approximation B^k. Under certain assumptions, the sequence $\{B^k\}$ will generate a good approximation to the true Lagrangian and fast local convergence can be achieved without using exact second derivatives.

To ensure global convergence as well as fast local convergence, the local SQP method is often combined with a line search of the form $x^{k+1} = x^k + t_k d^k$. The step length t_k is determined so that an appropriate merit function is reduced. Merit functions incorporate information about the objective function and possible constraint violations to decide whether progress towards a solution is made.

For an overview of SQP methods see the reports [GMW81] and [Pow83]. Recent developments in large-scale SQP methods are contained in [GMS97] and [Mur97].

3.3 A smooth SQP method (ECOPT)

3.3.1 Overview

The idea of approximating the MPEC by smooth nonlinear programs was first suggested by Facchinei, Jiang and Qi in [FJQ99]. Smoothing techniques are commonly applied to complementarity problems, so an extension to MPECs seems promising. Using a so-called *smoothing function*, a complementarity system with $y, s \in \mathbb{R}^m$

$$y \geq 0, s \geq 0, y^T s = 0 \tag{3.8}$$

is approximated by a smooth system of equations

$$\Phi_\mu(y, s) = 0,$$

where Φ_μ is continuously differentiable and approximates (3.8) for small $\mu > 0$. We will discuss the approximation used in this work in Section 3.3.2.

After the introduction of slack variables s, the MPEC can be written as

$$\begin{array}{rl} \min\limits_{x,y,s} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & F(x,y) - s = 0 \\ & y \geq 0, \quad s \geq 0, \quad y^T s = 0. \end{array}$$

Using the smoothing function Φ_μ, with $\mu > 0$, we can approximate the MPEC by

$$\begin{aligned} \min_{x,y,s} \quad & f(x,y) \\ \text{subject to} \quad & g(x,y) \geq 0 \\ & F(x,y) - s = 0 \\ & \Phi_\mu(y,s) = 0. \end{aligned} \tag{3.9}$$

The smoothing approach in [FJQ99] is only conceptual since it assumes the solution of nonlinear programs similar to (3.9) in every step. Consequently, several authors suggested related methods which only calculate one step towards a solution of (3.9) and then update the parameters.

In this section we describe one such method. First, we introduce the particular smoothing function we use in our algorithm in Section 3.3.2. Next, we discuss the QP subproblem used to determine a search direction and its properties in Section 3.3.3. We globalize the local SQP method by the Penalty function from Section 3.3.4. Finally, we analyze the overall algorithm and its convergence properties in Sections 3.3.5 and 3.3.6, respectively. A discussion of other smoothing methods to solve MPECs and their relationship to our algorithm in Section 3.4 ends this chapter.

3.3.2 Smooth approximation

The Fischer-Burmeister function

Our algorithm is based on the Fischer-Burmeister function [Fis92]

$$\varphi(a,b) := a + b - \sqrt{a^2 + b^2}.$$

The function φ is widely used in algorithms for complementarity problems. Its main characteristic is given in the following result.

Proposition 3.3.1. *The Fischer-Burmeister function has the following property:*

$$\varphi(a,b) = 0 \quad \Longleftrightarrow \quad a \geq 0, b \geq 0, ab = 0. \tag{3.10}$$

Proof: First, let $(a, b) \in \mathbb{R}^2$ with $a \geq 0$, $b \geq 0$, and $ab = 0$. If $a = 0$, then $\varphi(a, b) = \varphi(0, b) = b - \sqrt{b^2} = b - |b| = b - b = 0$. If $a > 0$, then $b = 0$ and it follows again that $\varphi(a, b) = 0$.

Next, let a and b such that $\varphi(a, b) = 0$. Then

$$0 \leq \sqrt{a^2 + b^2} = a + b. \tag{3.11}$$

Squaring both sides of (3.11), it is easy to see that $ab = 0$. Therefore, it also follows from (3.11) that $a \geq 0$ and $b \geq 0$. □

As we can see from Proposition 3.3.1, one can replace a complementarity system $a \geq 0$, $b \geq 0$, $ab = 0$ by the nonlinear equation $\varphi(a, b) = 0$. Unfortunately, the function φ is nonsmooth at the origin so that derivatives at so called degenerate points $a = b = 0$ do not exist. Nevertheless, the so called generalized Jacobian [Cla90]

$$\partial\varphi(0, 0) := \{r \in \mathbb{R}^2 \,|\, r = \lim_{k\to\infty} \nabla\varphi(a^k, b^k) \text{ with } (a^k, b^k) \to (0, 0) \text{ and } \nabla\varphi(a^k, b^k) \text{ exist}\}$$

exists, and is contained in the ball

$$\partial_C\varphi(0, 0) := \{(p, q) : (1 - p)^2 + (1 - q)^2 \leq 1\}, \tag{3.12}$$

where ∂_C denotes the C-subdifferential. Some algorithms for nonlinear and mixed complementarity problems make use of this generalized derivative, see [LFK96, LFK00, FKM99] and references therein. In the context of optimization problems, it is more difficult to handle nonsmooth constraints. A more promising approach is to use a smooth approximation of φ. We will describe this approximation next.

The smooth Fischer-Burmeister function

A smooth approximation of the Fischer-Burmeister function φ [Kan96] is given by

$$\varphi_\mu(a, b) := a + b - \sqrt{a^2 + b^2 + 2\mu}, \quad \mu > 0.$$

The smooth function φ_μ has the following property:

Proposition 3.3.2.

$$\varphi_\mu(a,b) = 0 \quad \Longleftrightarrow \quad a > 0, b > 0, ab = \mu. \tag{3.13}$$

The feasible set of $\varphi_\mu = 0$ is depicted in Figure 3.1. The smooth function φ_μ

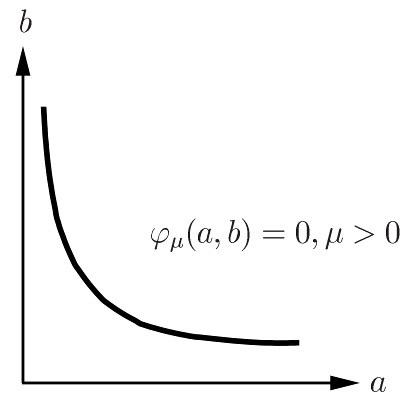

Figure 3.1: Feasible set of $\varphi_\mu(a,b) = 0, \mu > 0$

approximates the Fischer-Burmeister function for small μ in the following sense:

Proposition 3.3.3. *The function φ_μ satisfies the inequality*

$$|\varphi_\mu(a,b) - \varphi(a,b)| \leq \sqrt{2}\sqrt{\mu}$$

for all $(a,b) \in \mathbb{R}^2$ and all $\mu > 0$.

For $\mu > 0$, the first derivatives of the smoothing function $\varphi_\mu(a,b)$ are given by

$$\begin{aligned}\frac{\partial}{\partial a}\varphi_\mu(a,b) &= 1 - \frac{a}{\sqrt{a^2+b^2+2\mu}},\\ \frac{\partial}{\partial b}\varphi_\mu(a,b) &= 1 - \frac{b}{\sqrt{a^2+b^2+2\mu}},\end{aligned}$$

and the second derivatives are

$$\begin{aligned} \frac{\partial^2}{\partial a^2}\varphi_\mu(a,b) &= -\frac{b^2+2\mu}{(a^2+b^2+2\mu)^{3/2}}, \\ \frac{\partial^2}{\partial b^2}\varphi_\mu(a,b) &= -\frac{a^2+2\mu}{(a^2+b^2+2\mu)^{3/2}}, \\ \frac{\partial^2}{\partial a \partial b}\varphi_\mu(a,b) &= \frac{ab}{(a^2+b^2+2\mu)^{3/2}}. \end{aligned} \tag{3.14}$$

The smoothing function φ_μ also approximates the generalized derivative (3.12) of φ, see [KP99].

Proposition 3.3.4. *Let* $\{(a^k, b^k)\} \subseteq \mathbb{R}^2$ *and* $\{\mu_k\} \subseteq \mathbb{R}$ *be two sequences with* $\{(a^k, b^k)\} \to (\bar{a}, \bar{b}) = (0, 0)$ *and* $\{\mu_k\} \downarrow 0$. *Then*

$$\lim_{k\to\infty} \operatorname{dist}[\nabla\varphi_{\mu_k}(a^k, b^k), \partial_C\varphi(\bar{a}, \bar{b})] = 0,$$

where $\operatorname{dist}[y, \mathcal{S}] := \min\{\|y - y'\| \,|\, y' \in \mathcal{S}\}$ *for* $y \in \mathbb{R}^n$ *and* $\mathcal{S} \subset \mathbb{R}^n$. *If* $(\bar{a}, \bar{b}) \neq (0, 0)$ *then*

$$\lim_{k\to\infty} \nabla\varphi_{\mu_k}(a^k, b^k) = \nabla\varphi(\bar{a}, \bar{b}).$$

Using the smoothing function φ_μ, the MPEC can be approximated by a smooth optimization problems.

The smooth reformulation

Using the smoothing function φ_μ, $\mu \geq 0$, we can approximate the MPEC

$$\begin{aligned} \min_{x,y,s} \quad & f(x,y) \\ \text{subject to} \quad & g(x,y) \geq 0 \\ & F(x,y) - s = 0 \\ & y \geq 0, \quad s \geq 0, \quad y^T s = 0. \end{aligned} \tag{3.15}$$

by

$$\begin{aligned} \min_{x,y,s} \quad & f(x,y) \\ \text{subject to} \quad & g(x,y) \geq 0 \\ & F(x,y) - s = 0 \\ & \Phi_\mu(y,s) = 0, \end{aligned} \tag{3.16}$$

where Φ_μ denotes the mapping

$$\Phi_\mu(y,s) := \begin{pmatrix} \varphi_\mu(y_1, s_1) \\ \vdots \\ \varphi_\mu(y_m, s_m) \end{pmatrix}.$$

For $\mu = 0$, problem (3.16) is equivalent to (3.15) in the sense that feasible points, local and global solutions coincide. This is a direct consequence of Proposition 3.3.1. For $\mu > 0$, the solutions of (3.16) are good approximations of (3.15) due to Proposition 3.3.3.

Also, constraint qualifications holding for the MPEC (3.15) carry over to an approaching sequence of feasible points of the smooth reformulation for small perturbations $\mu > 0$. Here we discuss the LICQ. Some other constraint qualifications are discussed in [JR00].

Proposition 3.3.5. *For each $\mu > 0$, let (x^μ, y^μ, s^μ) be a feasible point of* (3.16)*. Suppose that $(x^\mu, y^\mu, s^\mu) \to (\bar{x}, \bar{y}, \bar{s})$. Then $(\bar{x}, \bar{y}, \bar{s})$ is a feasible point of* (3.15)*. Moreover, if the MPEC-LICQ holds at $(\bar{x}, \bar{y}, \bar{s})$ for* (3.15)*, then for all $\mu > 0$ small enough the NLP-LICQ holds for the feasible point (x^μ, y^μ, s^μ) of* (3.16)*.*

Proof: The first part follows from (3.10), (3.13) and the continuity of φ. The proof of the second part uses the continuity of g and the fact that the distance between $\nabla \Phi_{\mu,i}$ and the generalized gradient of Φ_i goes to zero as $\mu \to 0$. For the details, see [FP99]. □

Let (x^μ, y^μ, s^μ) be a local minimizer of the smoothed problem. Then, since the smoothed problem is an ordinary nonlinear programming problem, under some constraint qualification, there exist Lagrange multipliers $\pi^\mu \in \mathbb{R}^m$, $\xi^\mu \in \mathbb{R}^l$, and $\eta^\mu \in \mathbb{R}^m$ such that the vector $(x, y, s, \pi, \xi, \eta) = (x^\mu, y^\mu, s^\mu, \pi^\mu, \xi^\mu, \eta^\mu)$ satisfies the following KKT conditions for the smooth problem:

$$\begin{aligned}
&\nabla_x f(x,y) - g'_x(x,y)^T\xi - F'_x(x,y)^T\eta = 0,\\
&\nabla_y f(x,y) - g'_y(x,y)^T\xi - F'_y(x,y)^T\eta - A^T\pi = 0,\\
&\eta - B^T\pi = 0,\\
&0 \le g(x,y) \perp \xi \ge 0,\\
&F(x,y) - s = 0,\\
&\Phi_\mu(y,s) = 0,
\end{aligned} \tag{3.17}$$

where $[A, B] = \Phi'_\mu(y, s)$. Moreover, $(x, y, s, \pi, \xi, \eta) = (x^\mu, y^\mu, s^\mu, \pi^\mu, \xi^\mu, \eta^\mu)$ satisfies the inequalities

$$d^T \nabla^2 \mathcal{L}^\mu(x, y, s, \pi, \xi, \eta)\, d \ge 0, \quad \text{for all } d \in \mathcal{T}^\mu(x, y, s),$$

where

$$\mathcal{L}^\mu(x,y,s,\pi,\xi,\eta) := \\ f(x,y) - \sum_{i=1}^m \pi_i\, \Phi_{\mu,i}(y,s) - \sum_{j=1}^l \xi_j\, g_j(x,y) - \sum_{k=1}^m \eta_k\,(F_k(x,y) - s_k), \tag{3.18}$$

and

$$\mathcal{T}^\mu(x,y,s) := \\ \{d \in \mathbb{R}^{n+2m} \mid \nabla\Phi_\mu(y,s)^T d = 0, \nabla g_{\mathcal{I}_\mu}(x,y,s)^T d = 0, (\nabla F(x,y) - I)^T d = 0\},$$

with $\mathcal{I}_\mu := \{l \mid g_l(x^\mu, y^\mu) = 0\}$.

A different smooth reformulation

In the initial stages of this research, we also considered the following reformulation using Φ_μ in an inequality:

$$\begin{array}{rl} \min\limits_{x,y,s} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & F(x,y) - s = 0 \\ & \Phi_\mu(y,s) \leq 0, \\ & y \geq 0, s \geq 0. \end{array}$$

Notice that we also have to add nonnegativity constraints on y and s in this case. The obvious advantage of this reformulation is that y and s stay feasible, which is important since the equilibrium function F is sometimes only defined on the positive orthant. Another advantage is the knowledge of the sign of the multipliers corresponding to the inequality $\Phi_\mu(y,s) \leq 0$. Unfortunately, since Φ_μ is concave and not convex for all $\mu \geq 0$, this does not help in our situation. The Hessian update using exact second derivatives of Φ_μ and the multipliers corresponding to the inequality $\Phi_\mu(y,s) \leq 0$ will produce an indefinite update. The final reason why we did not pursue this formulation any further is that tests with an initial implementation of this approach produced infeasible QP subproblems during the SQP iterations. Although there are ways to deal with indefinite Hessian updates and infeasible QP subproblems, avoiding them in the first place seems to be more promising. As we will see next, the feasibility of the QP subproblems for the equality formulation (3.16) is given under mild conditions.

3.3.3 QP subproblems and solvability

As a general approach to find a solution to the MPEC (3.15), one can solve a sequence of the smooth nonlinear programs (3.16) for a decreasing sequence of μ's. An efficient way to do this is to given by SQP methods.

QP subproblem

For any given (x, y, s), $\mu > 0$, we will try to find a suitable search direction $d = (dx, dy, ds)$ by solving the following QP subproblem:

$$\begin{aligned}\min_{d=(dx,dy,ds)\in\mathbb{R}^{n+2m}} \quad & \nabla f(x,y)^T(dx,dy) + \frac{1}{2}d^T W d \\ \text{subject to} \quad & g'(x,y)(dx,dy) + g(x,y) \geq 0 \\ & F'(x,y)(dx,dy) - ds + (F(x,y) - s) = 0 \\ & Ady + Bds + \Phi_\mu(y,s) = 0,\end{aligned} \tag{3.19}$$

where the matrix $W \in \mathbb{R}^{(n+2m)\times(n+2m)}$ is symmetric positive definite and $[A, B] = \Phi'_\mu(y, s)$.

If the QP (3.19) has a solution d, then its KKT conditions have the following form:

$$\begin{aligned}&\begin{pmatrix}\nabla_x f(x,y)\\ \nabla_y f(x,y)\\ 0\end{pmatrix} + Wd - \begin{pmatrix}g'_x(x,y)^T\\ g'_y(x,y)^T\\ 0\end{pmatrix}\lambda_g - \begin{pmatrix}F'_x(x,y)^T\\ F'_y(x,y)^T\\ -I\end{pmatrix}\lambda_F - \begin{pmatrix}0\\ A^T\\ B^T\end{pmatrix}\lambda_\Phi = 0,\\ &0 \leq g'(x,y)(dx,dy) + g(x,y) \perp \lambda_g \geq 0,\\ &F'(x,y)(dx,dy) - ds + (F(x,y) - s) = 0,\\ &Ady + Bds + \Phi_\mu(y,s) = 0,\end{aligned} \tag{3.20}$$

where $\lambda = (\lambda_g, \lambda_F, \lambda_\Phi) \in \mathbb{R}^{(l+2m)}$ is the vector of Lagrange multipliers for (3.19).

Elastic QP subproblem

The general SQP approach fails if the QP subproblem (3.19) is infeasible or unbounded. The solution set will be bounded, i.e., unique, if W is positive definite, but the possibility of infeasibility must be taken into account.

If a QP subproblem is infeasible, we introduce a vector of elastic variables $t \in \mathbb{R}^l$

into the original problem in the following way:

$$\begin{aligned} \min_{x,y,s,t} \quad & f(x,y) + \rho e^T t \\ \text{subject to} \quad & g(x,y) + t \geq 0, \quad t \geq 0 \\ & F(x,y) - s = 0 \\ & \Phi_\mu(y,s) = 0. \end{aligned} \tag{3.21}$$

By introducing elastic variables into the QP subproblem, we allow some of the constraints to be violated and penalize this violation for being nonzero. Notice that we introduce elastic variables only to the inequality constraints. This has been suggested in [JR00], but our approach is different in that we only introduce elastic variables if necessary.

To find a search direction for the elastic problem (3.21), we solve the following elastic QP:

$$\begin{aligned} \min_{d=(dx,dy,ds,dt)\in\mathbb{R}^{n+2m+l}} \quad & \nabla f(x,y)^T(dx,dy) + \frac{1}{2}d^T W d + \rho e^T dt \\ \text{subject to} \quad & g'(x,y)(dx,dy) + dt + g(x,y) + t \geq 0, \quad dt \geq -t \\ & F'(x,y)(dx,dy) - ds + (F(x,y) - s) = 0 \\ & Ady + Bds + \Phi_\mu(y,s) = 0, \end{aligned} \tag{3.22}$$

where the matrix $W \in \mathbb{R}^{(n+2m+l)\times(n+2m+l)}$ is symmetric positive definite, $[A, B] = \Phi'_\mu(y,s)$, and $\rho > 0$ is the penalty parameter for the elastic variables t.

The SQP method proposed in [JR00] always solves an elastic QP similar to (3.22), and consequently needs to adjust a penalty parameter in every iteration. In our approach, we switch to elastic mode only if necessary, which is similar to [GMS97, Bom99]. The elastic QP subproblem is feasible under mild conditions as we will see next.

Solvability

The following properties [MR73] plays an important role for the solvability of our QP subproblems.

Definition 3.3.6. F is said to be a P_0-function with respect to y if for each $x \in \mathbb{R}^n$, $F(x, \cdot)$ is a P_0-function; i.e., for any y, $\bar{y} \in \mathbb{R}^m$ with $y \neq \bar{y}$, there exists an index i such that $y_i \neq \bar{y}_i$, and

$$(y_i - \bar{y}_i)(F_i(x, y) - F_i(x, \bar{y})) \geq 0.$$

We will also use the following matrix properties [FP66].

Definition 3.3.7. M is said to be a P-matrix if all its principal minors are positive.

Definition 3.3.8. M is said to be a P_0-matrix if all its principal minors are nonnegative.

The condition of F being a P_0-function is considered mild in the field of complementarity problems. It is for example given when the second-level problem in the MPEC is given by the optimality conditions of a convex optimization problems or a monotone complementarity problem. If F is a P_0-function with respect to y, then $F'_y(x, y)$ is a P_0-matrix [MR73].

To show feasibility of the QP's (3.19) and (3.22), we first look at an important submatrix.

Proposition 3.3.9. *Suppose F is a P_0-function with respect to y, then the matrix*

$$U = \begin{pmatrix} F'_y(x, y) & -I \\ A & B \end{pmatrix}$$

is nonsingular for any (x, y, s) with $\mu > 0$, where $[A, B] = \Phi'_\mu(y, s)$.

Proof: Since $[A, B] = \Phi'_\mu(y, s)$, both A and B are diagonal matrices with nonzero diagonal elements. It turns out that nonsingularity of the matrix U is equivalent to nonsingularity of the matrix $A + BF'_y(x, y)$, or $B^{-1}A + F'_y(x, y)$. Note that $B^{-1}A$

is a diagonal positive definite matrix, and $F'_y(x, y)$ is a P_0-matrix. Therefore, the matrix $B^{-1}A + F'_y(x, y) \in P$, see Theorem 3.4.2 in [CPS92]. Since P-matrices are nonsingular, the result follows. □

As a consequence of Proposition 3.3.9, solvability of the QP subproblems (3.19) and (3.22) is given by the following result.

Proposition 3.3.10. *Suppose $F'_y(x, y)$ is a P_0-matrix and $\mu > 0$. Let U be as defined in Proposition 3.3.9. Then*

(i) the elastic QP (3.22) *is always feasible;*

(ii) the QP (3.19) *has a nonempty feasible set if and only if the following system is consistent with respect to dx:*

$$[g'_x(x, y) - g'_y(x, y)(U^{-1})_{yy} F'_x(x, y)]dx \\ - g'_y(x, y)[(U^{-1})_{yy}(F(x, y) - s) + (U^{-1})_{ys}\Phi_\mu(y, s)] + g(x, y) \geq 0;$$

(iii) if furthermore $g(x, y)$ does not depend on y, then (3.19) *has a nonempty feasible set if*

$$g'(x)dx + g(x) \geq 0$$

is consistent.

Proof: The proof follows directly from Proposition 3.3.9 by using the fact that the matrix U is invertible and eliminating the y and s variables. □

3.3.4 Penalty function

To globalize the local SQP method, the following ℓ_1 penalty function is used:

$$\Theta_{(\rho^g, \rho^{NCP}, \mu)}(x, y, s) \\ = f(x, y) + \rho^g \sum_{i=1}^{l} \max\{-g_i(x, y), 0\} + \rho^{NCP} \sum_{j=1}^{m} (|F_j(x, y) - w_j| + |\varphi_\mu(y_j, s_j)|),$$

where ρ^g and ρ^{NCP} are positive penalty parameters. Two penalty parameters are necessary due to the special use of elastic variables for only some of the constraints, see [JR00]. The relationship between the penalty function and the solutions of the QP's (3.19) and (3.22) is given in the following results.

Proposition 3.3.11. *Let $\mu > 0$.*

(i) $\Theta_{(\rho^g,\rho^{NCP},\mu)}$ *is directionally differentiable at* (x, y, s).

(ii) If d is a solution of the QP (3.19), and if $\min\{\rho^g = \rho, \rho^{NCP}\} \geq \max_{1\leq i\leq l+2m} |\lambda_i^k|$ *where λ is the KKT multiplier of the QP (3.19), then*

$$\begin{aligned}\Theta'_{(\rho^g,\rho^{NCP},\mu)}(x,y,s;d) \quad &\leq \quad \nabla f(x,y)^T(dx,dy) - (\lambda_g)^T g'(x,y)(dx,dy) \\ &\quad + (\lambda_F)^T(F'(x,y)(dx,dy) - ds) \\ &\quad + (\lambda_\Phi)^T \Phi'_\mu(y,s)(dy,ds),\end{aligned}$$

and

$$\Theta'_{(\rho^g,\rho^{NCP},\mu)}(x,y,s;d) \leq -d^T W d.$$

(iii) If d is a solution of the elastic QP (3.22), $\rho^g = \rho$, and $\rho^{NCP} \geq \max_{1\leq i\leq l+2m} |\lambda_i^k|$ *with λ its KKT multiplier, then*

$$\begin{aligned}\Theta'_{(\rho^g,\rho^{NCP},\mu)}(x,y,s;d) \quad &\leq \quad \nabla f(x,y)^T(dx,dy) - (\lambda_g)^T g'(x,y)(dx,dy) \\ &\quad + (\lambda_F)^T(F'(x,y)(dx,dy) - ds) \\ &\quad + (\lambda_\Phi)^T \Phi'_\mu(y,s)(dy,ds),\end{aligned}$$

and

$$\Theta'_{(\rho^g,\rho^{NCP},\mu)}(x,y,s;d) \leq -d^T W d.$$

Proof: Part (i) follows from the continuous differentiability of f, g F, and Φ_μ. The proofs for parts (ii) and (iii) use the KKT condition of the QP's (3.19) and (3.22), respectively. The details can be found in [JR00].

Proposition 3.3.12. *Let $\mu > 0$ and suppose W is symmetric positive definite.*

(i) If d is a solution of the QP (3.19) with $d \neq 0$, and if $\min\{\rho^g = \rho, \rho^{NCP}\} \geq \max_{1 \leq i \leq l+2m} |\lambda_i^k|$ where λ is the KKT multiplier of the QP (3.19), then d is a descent direction of the penalty function $\Theta_{(\rho^g, \rho^{NCP}, \mu)}$.

(iii) If d is a solution of the elastic QP (3.22), $\rho^g = \rho$, and $\rho^{NCP} \geq \max_{1 \leq i \leq l+2m} |\lambda_i^k|$ with λ its KKT multiplier, then d is a descent direction of the penalty function $\Theta_{(\rho^g, \rho^{NCP}, \mu)}$.

Proof: The result is a direct consequence of parts (ii) and (iii) in Proposition 3.3.11. Proposition 3.3.12 shows that the solutions of either QP subproblem generates a descent direction of the merit function $\Theta_{(\rho^g, \rho^{NCP}, \mu)}$ for sufficiently large penalty parameters when W is symmetric positive definite and $\mu > 0$. This will be important for the line search of the algorithm ECOPT.

3.3.5 Statement of the algorithm

We will now describe the overall algorithm ECOPT.

Smooth SQP method (ECOPT)

Step 0. **(Initialization)** Let $\rho_{-1} > 0$, $\delta_1 > 0$, $\delta_2 > 0$, $\beta_\mu \in (0, 1)$, $\beta_\epsilon \in (0, 1)$, $\sigma \in (0, 1)$, $\tau \in (0, 1)$. Choose $u^0 = (x^0, y^0, w^0) \in \mathbb{R}^{n+2m}$, and choose $\mu_0 > 0$, $\epsilon_0 > 0$, and a symmetric positive definite matrix $W_0 \in \mathbb{R}^{(n+2m)\times(n+2m)}$. Set $k := 0$.

Step 1a. **(Search direction (inelastic))**

If all QP subproblems for $k' < k$ have been feasible **then** find a solution to the QP (3.19) with $(x, y, w) = (x^k, y^k, w^k)$, $\mu = \mu_k$, $W = W_k$.

If a solution exists, let $d^k \in \mathbb{R}^{n+2m}$ be the solution of this QP and $\lambda_k = (\lambda_g, \lambda_F, \lambda_\Phi)$ be its corresponding KKT multiplier. Set $\xi := 0$, $\lambda_\xi := 0$ and go to Step 2.

else go to Step 1b.

else go to Step 1b.

Step 1b. (Search direction (elastic)) Find a solution to the QP (3.22) with $(x, y, w) = (x^k, y^k, w^k)$, $\mu = \mu_k$, $W \in \mathbb{R}^{n+2m+l}$ defined as

$$W := \begin{pmatrix} W_k & 0 \\ 0 & I_l \end{pmatrix},$$

where I_l denotes the identity matrix of dimension l. Let $(d^k, \xi^k) \in \mathbb{R}^{n+2m+l}$ be the solution of this QP and $\lambda_k = (\lambda_g, \lambda_F, \lambda_\Phi, \lambda_\xi)$ be its corresponding KKT multiplier. Go to Step 2.

Step 2. (Termination check) If a stopping rule is satisfied, STOP. Otherwise, go to Step 3.

Step 3. (Penalty update) Let

$$\tilde{\rho}_k = \begin{cases} \rho_{k-1} & \text{if } \rho_{k-1} \geq \max_{1 \leq i \leq l+2m+1} |\lambda_i^k|, \\ \delta_1 + \max_{1 \leq i \leq l+2m+1} |\lambda_i^k| & \text{otherwise.} \end{cases}$$

Define $\rho_k^g = \rho_{k-1}$ and $\rho_k^{NCP} = \tilde{\rho}_k$ and

$$\rho_k = \begin{cases} \tilde{\rho}_k & \text{if } \sum_{1 \leq i \leq l} \xi_i^k = 0, \\ \tilde{\rho}_k + \delta_2 & \text{otherwise.} \end{cases}$$

Step 4. (Line Search) Find the smallest m_k in $\{0, 1, 2, \ldots\}$ such that

$$\Theta_{(\rho_k^g, \rho_k^{NCP}, \mu_k)}(u^k + \tau^{m_k} d^k) - \Theta_{(\rho_k^g, \rho_k^{NCP}, \mu_k)}(u^k) \leq -\sigma \tau^{m_k} (d^k)^T W d^k.$$

Set $t_k := \tau^{m_k}$ and $u^{k+1} := u^k + t_k d^k$.

Step 5. (Update) Let

$$\mu_{k+1} = \begin{cases} \beta_\mu \mu_k & \text{if } \|d^k\| \leq \epsilon_k, \\ \mu_k & \text{otherwise,} \end{cases}$$

$$\epsilon_{k+1} = \begin{cases} \beta_\epsilon \epsilon_k & \text{if } \|d^k\| \leq \epsilon_k, \\ \epsilon_k & \text{otherwise.} \end{cases}$$

Choose a symmetric positive definite matrix $W_{k+1} \in \mathbb{R}^{(n+2m)\times(n+2m)}$. Set $k \leftarrow k+1$, and return to Step 1a .

From the results in Sections 3.3.3 and 3.3.4, we know that the above algorithm is well-defined if F is at least a P_0-function in y. In particular, the QP subproblems are feasible and their solutions d^k are descent directions for the merit function.

3.3.6 Convergence analysis

We will now summarize the convergence properties of ECOPT. Our algorithm is very similar to the explicit smooth SQP analyzed in [JR00], so most of the proofs are similar to those of that reference. To prove global convergence, the following two assumptions are made:

Assumption 1. There exist two positive numbers $\alpha < \beta$ such that each of the symmetric matrices W_k used in ECOPT satisfies the following condition for all vectors v of appropriate dimension:

$$\alpha\|v\|^2 \leq v^T W_k v \leq \beta\|v\|.$$

Assumption 2. For all large k, $\rho_k = \rho_*$.

While the first assumption is commonly assumed for SQP methods, the second is more restrictive. Besides other things, it implies that all QP subproblems are feasible in the limit so that no update of the penalty parameter takes place in Step 5. Assumption 2 holds under the following additional Assumptions:

Let H be the function representing the equality constraints of (3.16) or (3.21), i.e., $H(u) = (F(x,y) - s, \Phi_\mu(y,s))$ with $u = (x,y,s)$.

Assumption 3. $\{u^k\} = \{x^k, y^k, s^k\}$ is bounded.

Assumption 4. The generalized Jacobian $\partial H(u^*)$ has full row rank at any accumulation point u^* of $\{u^k\}$.

Assumption 5. For any accumulation point u^* of $\{u^k\}$ and any $V \in \partial H(u^*)$, there exists $d = (dx, dy, ds)$ such that $g'(x^*, y^*)(dx, dy) + g(x^*, y^*) > 0$ and $Vd + H(u^*) = 0$.

Theorem 3.3.13. *Assume that Assumption 1 holds and F is a P_0-function with respect to y. Let $\mu_0 > 0$ and $\{u^k\}$,$\{\mu_k\}$, and $\{\epsilon_k\}$ be the sequences generated by* ECOPT.

(i) If Assumption 2 holds and $\{u^k\}$ has a limit point, then

$$\lim_{k\to\infty} \mu_k = 0, \qquad \lim_{k\to\infty} \epsilon_k = 0.$$

(ii) Let $K = \{k\colon \|d^k\| \leq \epsilon_k\}$. If we assume that Assumption 2 holds and $\{u^k\}_{k\in K}$ has an accumulation point $u^* = (x^*, y^*, s^*)$, then u^* is a generalized stationary point of (3.15) *with $\mu = 0$. Furthermore, if (x^*, y^*) is lower-level nondegenerate, then (x^*, y^*) is a piecewise stationary point of the MPEC.*

(iii) If Assumptions 1, 3, 4, and 5 hold, then so does Assumption 2.

Proof: We will discuss the main points of the proof here, the details can be found in [JR00].

(i) It is clear that $\{\mu_k\}$ is bounded, so the sequence has an accumulation point. Suppose μ_* is such an accumulation point. If $\mu_* > 0$, then $\|d^k\| \leq \epsilon_k$ occurs only finitely many times. This means that after finitely many iterations, μ_k and ϵ_k remain unchanged. This implies that for some k_0 and all $k \geq k_0$, $\mu_k = \mu_{k_0} > 0$ and $\epsilon_k = \epsilon_{k_0} > 0$. In this case, ECOPT reduces to a regular SQP method for the smooth problem (3.16), where some care has to be taken if elastic mode is entered due to infeasible QP subproblems. In any case, it follows that some subsequence of $\{d^k\}$ approaches 0 as $k \to \infty$, which implies that $\|d^k\| \leq \epsilon_{k_0}$ will eventually happen, see [JR00]. This is a contradiction. Therefore, $\lim_{k\to\infty} \mu_k = 0$. By the update rule in Step 5, it is also true that $\lim_{k\to\infty} \epsilon_k = 0$.

(ii) By Assumption 2 and the update rule of the penalty parameter, the KKT multiplier sequence $\{\lambda^k\}_{k\in K}$ is bounded and the QP subproblem (3.19) is solvable (or, in elastic mode, $dt = t = 0$), since $\rho_k = \rho_*$ for all sufficiently large k. Note that for each $k \in K$, $\|d^k\| \leq \epsilon_k$. Hence $\lim_{k\to\infty, k\in K} d^k = 0$. By passing to the limit for $k \in K$, it follows from the KKT conditions of the QP subproblem (3.20) (or (3.22) with $dt = t = 0$) and Assumption 3 that u^* is a generalized stationary point of (3.15) with $\mu = 0$. The limit point (x^*, y^*) is a piecewise stationary point of the MPEC if (x^*, y^*) is lower-level nondegenerate, see Proposition 3.7 in [JR00].

(iii) The result has been proven in [JR00]. □

The assertion that the limit point u^* is a generalized stationary point of (3.15) with $\mu = 0$ is not the strongest result one could hope for. For example, it does not imply B-stationarity unless (x^*, y^*) is lower-level nondegenerate or some further conditions are satisfied.

To prove convergence to B-stationary points, one has to show that the multipliers $\lambda_{\Phi,i}$, for $i \in \bar{\beta}$, corresponding to lower-level degenerate points are nonnegative. In [FP99], Fukushima and Pang consider a smoothing algorithm very similar to ours and prove that under some further conditions, the limit point u^* will indeed be a B-stationary point. In essence, the main assumption in their proof is that the LICQ is satisfied and there exists a subsequence that satisfies second-order necessary conditions for the smooth approximation. These assumption are likely to be satisfied in practice so that convergence to B-stationary is given. Another technical assumption used by the authors is called “asymptotically weakly nondegenerate”. Although it is not clear when this condition is satisfied, we observed in our numerical tests that $\lambda_{\Phi,i}$, for $i \in \bar{\beta}$, is indeed nonnegative at the solution so that the algorithm converges to a B-stationary point.

Closely related to the work [FP99] of Fukushima and Pang is a recent paper by Scholtes [Sch01] who considers a regularization scheme for MPECs. His convergence results are similar to [FP99] but instead of asymptotic weak nondegeneracy he uses a condition called “upper level strict complementarity” to prove convergence to a

B-stationary point under LICQ and the existence of a subsequence satisfying second-order necessary conditions. In conclusion, we conjecture that a stronger convergence result than Theorem 3.3.13 is possible. This is also indicated by the numerical results and the fact that λ_Φ had the correct sign in our tests.

3.4 Relationship to other smoothing methods

As we mentioned earlier, the first smoothing method for MPECs is due to Facchinei, Jiang, and Qi. In contrast to ECOPT, the smoothing method presented in [FJQ99] uses a slightly different formulation of the MPEC; instead of nonlinear complementarity constraints, the authors consider variational inequalities in KKT form. Their method uses a smoothing function for the min-function and a black-box NLP solver to solve the resulting smooth problems. Convergence is proved under slightly stronger conditions than given for ECOPT in Theorem 3.3.13. Numerical results for the nonlinear test problems tested in Section 4.13 are also presented.

A smoothing method for MPECs with linear complementarity constraints has been suggested by Fukushima, Luo, and Pang in [FLP98]. Similar to our approach, the authors also use the Fischer-Burmeister function φ and its smoothing function φ_μ. Indeed, specializing ECOPT to MPECs with linear complementarity constraints would result essentially in the smoothing algorithm from [FLP98]. One notable difference is that our implementation of ECOPT uses exact second derivatives of the objective and smoothing function, while the algorithm tested in [FLP98] does not use this information. From our experience, this is in important feature that improves the performance significantly.

Finally, Jiang and Ralph [JR00] propose two smooth SQP methods for MPECs: an *implicit smooth SQP* and an *explicit smooth SQP* method. While the implicit method treats the smoothing parameter μ as a variable and updates it in every iteration, the explicit SQP method updates μ separately. Our algorithm ECOPT shares many common ideas with the explicit smooth algorithm analyzed by Jiang and Ralph. The main difference is the way we handle infeasible QP subproblems and the fact that the authors did not implement and test either of the two proposed algorithms. Important details for an implementation that are left open in [JR00] are the stopping criterion and especially an update strategy for the matrix W_k. Solutions of these issues and extensive numerical tests will be presented in the next chapter.

Chapter 4

Implementation and Numerical Comparison

4.1 Implementation details for ECOPT

Before presenting the numerical results, we give some of the details of the MATLAB implementation of ECOPT.

Termination criterion

The following termination criterion is motivated by the convergence analysis in Section 3.3.6. Recall that a feasible point $(\bar{x}, \bar{y})$ of the MPEC is called weakly stationary if there exist MPEC multipliers $\bar{\lambda}_i \geq 0$, $\bar{\mu}_j, \bar{\nu}_k$ satisfying

$$\begin{bmatrix} \nabla f_x(\bar{x}, \bar{y}) \\ \nabla f_y(\bar{x}, \bar{y}) \end{bmatrix} - \sum_{i \in \bar{\mathcal{I}}} \bar{\lambda}_i \begin{bmatrix} \nabla g_x(\bar{x}, \bar{y}) \\ \nabla g_y(\bar{x}, \bar{y}) \end{bmatrix} - \sum_{j \in \alpha \cup \beta} \bar{\mu}_j \begin{bmatrix} \nabla F_x(\bar{x}, \bar{y}) \\ \nabla F_y(\bar{x}, \bar{y}) \end{bmatrix} - \sum_{k \in \beta \cup \gamma} \bar{\nu}_k \begin{bmatrix} 0 \\ I \end{bmatrix} = 0.$$

We say that $(\bar{x}, \bar{y}, \bar{\lambda}, \bar{\mu}, \bar{\nu})$ is an ϵ-weakly stationary point of the MPEC if

$$\left\| \begin{bmatrix} \nabla f_x(\bar{x}, \bar{y}) \\ \nabla f_y(\bar{x}, \bar{y}) \end{bmatrix} - \sum_{i \in \bar{\mathcal{I}}} \bar{\lambda}_i \begin{bmatrix} \nabla g_x(\bar{x}, \bar{y}) \\ \nabla g_y(\bar{x}, \bar{y}) \end{bmatrix} - \sum_{j \in \alpha \cup \beta} \bar{\mu}_j \begin{bmatrix} \nabla F_x(\bar{x}, \bar{y}) \\ \nabla F_y(\bar{x}, \bar{y}) \end{bmatrix} - \sum_{k \in \beta \cup \gamma} \bar{\nu}_k \begin{bmatrix} 0 \\ I \end{bmatrix} \right\|$$
$$+ \|\min(g(x, y), \lambda)\| + \|\min(y, F(x, y))\| \leq \epsilon.$$

For $\epsilon = 0$, any ϵ-weakly stationary point is weakly stationary and vice versa.

In our implementation, we check this termination criterion by solving a linear programming problem if the feasibility gap, in particular complementarity between y and F, is below 1.e-6.

Forcing a stronger stationarity concept for termination would require sign constraints on the multipliers μ and ν; see [Sch01] for details. These could be easily incorporated into the LP, but in light of the convergence analysis, it is not guaranteed that our algorithm converges to such points.

QP subproblem solution

We implemented our algorithm ECOPT in MATLAB using the relatively new TOMLAB environment [Hol99a, Hol99b]. TOMLAB replaces routines from the Optimization Toolbox with its own routines and interfaces to FORTRAN packages such as the nonlinear optimization solvers SNOPT [GMS97], NPSOL [GMSW86] and MINOS [MS95] and the QP-solvers QPOPT and SQOPT.

In our implementation, we can choose to solve the QP subproblems by either the Optimization Toolbox QP-solver `quadprog`, QPOPT, SQOPT, or MINOS. In the tests reported later, we use QPOPT throughout, but MINOS or SQOPT would give very similar results. Our experience suggests that `quadprog` is less reliable and slower than either of the other three solvers we used. Notice that the use of a sparse QP-solver, such as SQOPT or MINOS, would allow the solution of large scale, sparse MPECs. We did not pursue this in the current work, but minor modifications would allow the solution of much larger, sparse problems with ECOPT.

Line search

The implemented line search is the same Armijo line search as described in the algorithm. We stop the algorithm if the step size becomes too small, i.e., $t_k < t_{\min}$.

Hessian approximation

One of the most critical details of any SQP method is the update of the matrix W_k. Ideally, W_k should be a good approximation to the Hessian matrix of the Lagrangian at the current point. Although theoretically, we only need W_k to be symmetric positive definite, in practice the particular choice of W_k is critical. Using the identity matrix, for example, turns out to be a poor choice.

In the following, we focus on the use of exact second derivatives of the smoothing function Φ in the update of W_k. In particular, we will first assume that g, F are affine and that exact second derivatives of the objective function f are available. These assumptions are valid for most of the problems in this work and in particular for the electricity model in Chapter 6. We will also discuss the extension to the general nonlinear case.

The Hessian matrix of the Lagrangian (3.18) is given by

$$\nabla^2 \mathcal{L}^\mu(x, y, s, \pi, \xi, \eta) =$$
$$\nabla^2 f(x, y) - \sum_{i=1}^{m} \pi_i \nabla^2 \Phi_{\mu,i}(y, s) - \sum_{j=1}^{l} \xi_j \nabla^2 g_j(x, y) - \sum_{k=1}^{m} \eta_k \nabla^2 F_k(x, y).$$

If g and F are affine, this reduces to

$$\nabla^2 \mathcal{L}^\mu(x, y, s, \pi) = \nabla^2 f(x, y) - \sum_{i=1}^{m} \pi_i \nabla^2 \Phi_{\mu,i}(y, s).$$

The second derivatives of Φ_μ are given by (3.14). It is easy to see that Φ_μ is a concave function in y and s for all $\mu \geq 0$, so that W_k is positive semidefinite if f is convex and the multipliers π_i are nonnegative for all i. Unfortunately, the sign of the multipliers is not clear since we use Φ_μ in an equality.

In our implementation, we approximate $\nabla^2 \mathcal{L}^\mu$ by a positive semidefinite matrix W_k in the following way: First, as multiplier estimates, we use the QP multipliers of the most recently solved QP subproblem. These are automatically given and don't need to be computed separately like least-squares multiplie, which involves the solution of an extra system of linear equations, for example. Next, we experimented with several different ways of ensuring positive semidefiniteness of W_k. Given that f is convex in all cases considered, we propose an easy update strategy that worked well for the problems tested:

We first use the positive QP multipliers $\lambda^k_{\Phi,i} > 0$ together with exact second derivatives of Φ_μ to form the update

$$\tilde{W}_k = \nabla^2 f(x^k, y^k) - \sum_{\lambda_{\Phi,i} > 0} \lambda_{\Phi,i} \nabla^2 \Phi_{\mu,i}(y^k, s^k).$$

For the QP multipliers $\lambda^k_{\Phi,i} < 0$, we penalize y and s symmetrically by a diagonal matrix $D \in \mathbb{R}^{n+2m}$ with elements

$$\begin{aligned} D_{i,i} &= 0, && \text{for } i = 1, \ldots, n, \\ D_{n+j,n+j} = D_{n+m+j,n+m+j} &= \gamma \frac{|y_j s_j|}{|y^T s|}, && \text{for } j = 1, \ldots, m. \end{aligned}$$

The motivation for this is given by the fact that a negative multiplier $\lambda^k_{\Phi,i}$ indicates that the optimal point of the current subproblem would violate the constraint $\varphi_\mu(y^k_i, s^k_i) \le 0$, see Figure 4.1. A positive penalization of y_i and s_i therefore pushes

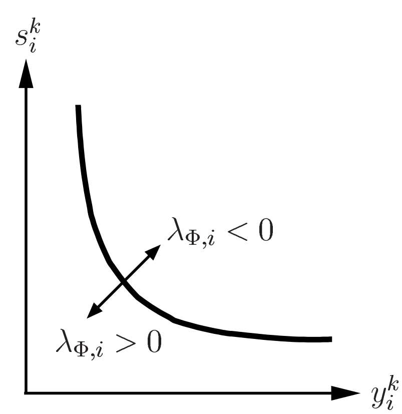

Figure 4.1: Lagrange multipliers of $\varphi_\mu(y^k_i, s^k_i) = 0$

y^{k+1}_i and s^{k+1}_i inside the constraint more likely. In our implementation we set $\gamma =$

10^{-4}. The matrix W_k used is then

$$W_k = \tilde{W}_k + D.$$

We can extend the above approach to the general nonlinear case by including second-order derivative information of g and F in the same disaggregated way as we did for Φ_μ. If exact second derivatives are not available, updates for g and F can be used separately. Keeping the update positive (semi-) definite is more difficult in this case, since the curvatures of g and F are possibly unknown. For general ways to deal with this, see [GMW81]. One could also work with indefinite updates and apply ideas introduced in [MP99] and [Gol99].

Parameter settings

We use the following parameter settings for all runs of ECOPT:

$$\begin{array}{cccccc} \rho_{-1} = 10^2, & \rho_1 = 10^2, & \rho_2 = 10^2, & \mu_0 = 10^{-4}, & \beta_\mu = 10^{-4}, & \mu_{\min} = 10^{-16}, \\ \epsilon_0 = 10, & \beta_c = 0.3, & \sigma = 0.5, & \tau = 0.5, & t_{\min} = 10^{-12}, & k_{\max} = 300. \end{array}$$

The main termination criterion is set to 10^{-6}.

4.2 Implementation details for NLP codes

To facilitate a direct comparison of the efficiency of ECOPT with that of state of the art nonlinear optimization solvers, we used the TOMLAB interface to the SQP algorithms NPSOL and SNOPT, as well as the sequential linearization algorithm MINOS. The MATLAB interface we wrote passes the following formulation to each of the solvers:

$$
\begin{aligned}
\min_{x,y,s} \quad & f(x,y) \\
\text{subject to} \quad & g(x,y) \geq 0 \\
& F(x,y) - s = 0 \\
& y \geq 0, \quad s \geq 0, \quad y^T s = 0.
\end{aligned}
$$

Notice the inclusion of the slack variables s which make the complementarity constraints "less nonlinear" and should improve the numerical performance. Of the three solvers, only MINOS has been previously applied to MPEC problems similar to the ones considered here [JR99]. The new interface to the solvers through TOMLAB seems to be more stable than the MEX interface to MINOS that has been used in [JR99].

After some tests, the only changes we made to the standard parameters set by TOMLAB were to increase the accuracy of the subproblem solutions for both NPSOL and SNOPT to 1.e-7 from 1.e-6, the default set by TOMLAB. This change enabled a faster convergence in some of the test problems reported later. The major convergence tolerance is set to the default 1.e-6.

4.3 Quadratic problems (QPECs)

The first type of problem we test the algorithms on are *quadratic problems with equilibrium constraints* (QPECs for short). A QPEC is a quadratic MPEC, that is an optimization problem with quadratic objective function, linear first-level constraints and second-level constraints that are given by parametric affine variational inequalities or linear complementarity problems. To do this, we use the MATLAB program QPECgen by Jiang and Ralph [JR99]. QPECgen is a random problem generator which allows the control over important properties of the problem, like dimension, convexity and, in particular, degeneracy in the first- and second level of the constraints. It has been used in [JR99] to compare the piecewise sequential quadratic programming algorithm (PSQP) and the penalty interior point algorithm (PIPA) with standard NLP codes MINOS and the constrained programming solver from the Matlab Optimization Toolbox.

In the following, we describe and test two types of problems generated by QPECgen: quadratic problems with affine variational inequality (AVI) constraints and quadratic problems with LCP constraints.

4.3.1 QPECs with AVI constraints

Using the first-order necessary conditions of the second-level AVI, the quadratic problem with AVI constraints in $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $\lambda \in \mathbb{R}^p$ can be formulated as [JR99]:

$$
\begin{aligned}
\min_{x,y,\lambda} \quad & \frac{1}{2}[x,y]P\begin{bmatrix} x \\ y \end{bmatrix} + c^T x + d^T y \\
\text{subject to} \quad & Gx + Hy + a \leq 0 \\
& Nx + My + q + E^T\lambda = 0 \\
& 0 \leq \lambda \perp Dx + Ey + b \leq 0,
\end{aligned} \tag{4.1}
$$

where $P \in \mathbb{R}^{(n+m)\times(n+m)}$, $c \in \mathbb{R}^n$, $d \in \mathbb{R}^m$, $G \in \mathbb{R}^{l\times n}$, $H \in \mathbb{R}^{l\times m}$, $a \in \mathbb{R}^l$, $N \in \mathbb{R}^{m\times n}$, $M \in \mathbb{R}^{m\times m}$, $q \in \mathbb{R}^m$, $E \in \mathbb{R}^{p\times m}$, $D \in \mathbb{R}^{p\times n}$, and $b \in \mathbb{R}^p$.

Notice, that the first-level equality constraints are joint constraints in all variables x, y, and λ.

Characteristics of AVI-constrained problems

The following summarizes the problem characteristics of the first set of test problems. We use the same parameters and random seed as the problems tested in [JR99], so a direct comparison of results will be possible. The starting points are generated randomly in the same way as has been done in [JR99].

The test set consists of a total of 16 small to medium size problems in four groups. The dimensions of these problem ranges from $n + m = 28$ to 70 variables (x, y) and $l + m + 2p = 40$ to 106 constraints excluding the complementarity constraints.

qpec_type	cond_P(scale_P)	convex_f	symm_M	mono_M
100	100(100)	1	1	1
cond_M(scale_M)	second_deg	first_deg	mix_deg	tol_deg
200(200)	0	2	0	1.e-6
implicit	rand_seed			
0	0			

Table 4.1: Parameters for AVI-QPECs

Group 1, Problems 1–4. The parameters of the first group are summarized in Table 4.1. For a detailed explanation of the terminology we refer the reader to [JR99]. The problems in the first group have the nice property that the objective function is strictly convex, the Hessian of the objective function is well-conditioned, the second-level problem is symmetric, strongly monotone and well-conditioned with respect to the second-level variables y, and the second-level degeneracy does not exist at the generated solution.

Group 2, Problems 5–8. The second group has the same parameters as Group 1 except for `second_deg` = `4` and `mix_deg` = `2`.

Group 3, Problems 9–12. The same parameters as in Group 2 are used except that `mono_M` = `0` and `symm_M` = `0`, i.e., the lower-level matrix M is not necessarily monotone or symmetric. The solution set to the parametric second-level problem is less well behaved and algorithms can be expected to have more problems detecting the generated solutions.

Group 4, Problems 13–16. The same parameters as in Group 2 are used except that `second_deg` = `8`. In theory, higher degeneracy affects the performance of some algorithms, so this parameter setting is to test how much this matters in practice.

P	Problem number
(m, n, l, p)	Dimension of the problem
`deg`	No. of degenerate lower-level indices at the generated solution (x_{gen}, y_{gen})
Maj	No. of major iterations (QPs solved for ECOPT, NPSOL, SNOPT/ linearly-constrained NLPs solved by MINOS)
Min	No. of minor iterations (QP iterations by ECOPT and SNOPT/ pivots performed by MINOS) Minor iterations are not passed on by the TOMLAB interface for NPSOL at this time.
F-ev.	No. of function evaluations for θ, F, and g
f	The objective function value at the found solution
f_{gen}	The objective function at the generated solution (x_{gen}, y_{gen})
Norm	The infinity norm of the difference vector between the found solution and the generated solution (x_{gen}, y_{gen})

Table 4.2: Notation for numerical results

P	(m, n, l, p)	`deg`	Maj	Min	F-ev.	f	f_{gen}	Norm
1	$(20, 8, 4, 8)$	0	8	15	9	−65.0099	−65.0099	6.4e−07
2	$(30, 12, 8, 12)$	0	5	21	6	−118.886	−118.886	4.7e−07
3	$(40, 16, 12, 16)$	0	7	27	8	−74.1346	−74.1346	1.1e−07
4	$(50, 20, 16, 20)$	0	7	34	8	−133.573	−133.567	2.7e−02
5	$(20, 8, 4, 8)$	4	16	35	28	−71.4787	−71.4787	1.0e−05
6	$(30, 12, 8, 12)$	4	11	32	17	−115.8	−115.8	8.9e−09
7	$(40, 16, 12, 16)$	4	13	40	17	−50.5499	−50.5499	1.1e−06
8	$(50, 20, 16, 20)$	4	18	54	54	−73.2094	−73.1951	3.2e−02
9	$(20, 8, 4, 8)$	4	8	15	11	−80.161	−83.9831	1.7e−01
10	$(30, 12, 8, 12)$	4	12	32	24	−179.695	−179.695	8.9e−05
11	$(40, 16, 12, 16)$	4	15	44	18	−71.7339	−71.7339	2.8e−06
12	$(50, 20, 16, 20)$	4	7	34	8	−96.5076	−110.073	3.7e−01
13	$(20, 8, 4, 8)$	8	22	49	42	−87.0353	−87.0353	2.3e−05
14	$(30, 12, 8, 12)$	8	16	43	20	−128.801	−128.801	1.5e−06
15	$(40, 16, 12, 16)$	8	24	70	48	−86.3115	−86.3114	2.3e−06
16	$(50, 20, 16, 20)$	8	16	70	29	−47.3336	−47.3336	4.8e−06

Table 4.3: Numerical results for ECOPT on AVI-QPECs

P	(m,n,l,p)	`deg`	Maj	Min	F-ev.	f	f_{gen}	Norm
1	$(20,8,4,8)$	0	47	–	56	−65.0099	−65.0099	1.5e−09
2	$(30,12,8,12)$	0	21	–	35	−118.887	−118.886	4.3e−03
3	$(40,16,12,16)$	0	30	–	38	−74.1346	−74.1346	1.4e−07
4	$(50,20,16,20)$	0	49	–	57	−133.564	−133.567	3.0e−02
5	$(20,8,4,8)$	4	24	–	33	−71.4787	−71.4787	6.7e−07
6	$(30,12,8,12)$	4	25	–	36	−115.8	−115.8	6.2e−08
7	$(40,16,12,16)$	4	22	–	28	−50.5499	−50.5499	7.7e−08
8	$(50,20,16,20)$	4	52	–	61	−73.2094	−73.1951	3.2e−02
9	$(20,8,4,8)$	4	18	–	25	−82.7748	−83.9831	1.2e−01
10	$(30,12,8,12)$	4	38	–	49	−179.695	−179.695	1.0e−06
11	$(40,16,12,16)$	4	32	–	48	−71.7339	−71.7339	1.6e−07
12	$(50,20,16,20)$	4	34	–	52	−110.073	−110.073	8.5e−07
13	$(20,8,4,8)$	8	17	–	25	−87.0353	−87.0353	1.7e−08
14	$(30,12,8,12)$	8	18	–	26	−128.801	−128.801	3.8e−07
15	$(40,16,12,16)$	8	23	–	28	−86.2204	−86.3114	8.2e−02
16	$(50,20,16,20)$	8	78	–	96	−47.3336	−47.3336	3.3e−10

Table 4.4: Numerical results for NPSOL on AVI-QPECs

P	(m,n,l,p)	deg	Maj	Min	F-ev.	f	f_{gen}	Norm
1	$(20,8,4,8)$	0	98	341	107	−65.0099	−65.0099	3.8e−09
2	$(30,12,8,12)$	0	32	267	38	−118.887	−118.886	4.3e−03
3	$(40,16,12,16)$	0	39	283	44	−74.1346	−74.1346	2.4e−07
4	$(50,20,16,20)$	0	96	717	110	−133.45	−133.567	1.0e−01
5	$(20,8,4,8)$	4	42	223	48	−71.4787	−71.4787	7.4e−08
6	$(30,12,8,12)$	4	46	350	53	−115.8	−115.8	1.6e−08
7	$(40,16,12,16)$	4	48	434	57	−50.5499	−50.5499	1.6e−06
8	$(50,20,16,20)$	4	231	1123	259	−73.1951	−73.1951	1.5e−07
9	$(20,8,4,8)$	4	49	237	55	−83.9831	−83.9831	1.2e−13
10	$(30,12,8,12)$	4	46	323	51	−177.071	−179.695	1.9e−01
11	$(40,16,12,16)$	4	22	345	27	−29.0859	−71.7339	9.2e−01
12	$(50,20,16,20)$	4	30	415	35	120.23	−110.073	1.4e+00
13	$(20,8,4,8)$	8	42	254	48	−87.0348	−87.0353	6.4e−04
14	$(30,12,8,12)$	8	20	285	26	−128.801	−128.801	2.8e−08
15	$(40,16,12,16)$	8	37	354	42	−86.3114	−86.3114	1.5e−07
16	$(50,20,16,20)$	8	53	614	58	−47.3336	−47.3336	1.6e−07

Table 4.5: Numerical results for SNOPT on AVI-QPECs

P	(m,n,l,p)	`deg`	Maj	Min	F-ev.	f	f_{gen}	Norm
1	$(20,8,4,8)$	0	6	125	170	−65.0099	−65.0099	3.4e−12
2	$(30,12,8,12)$	0	21	166	249	−117.348	−118.886	1.5e−01
3	$(40,16,12,16)$	0	16	310	491	−74.0042	−74.1346	5.0e−02
4	$(50,20,16,20)$	0	11	363	498	−133.573	−133.567	2.7e−02
5	$(20,8,4,8)$	4	17	375	746	−71.4787	−71.4787	9.7e−12
6	$(30,12,8,12)$	4	19	239	343	−110.86	−115.8	2.5e−01
7	$(40,16,12,16)$	4	5	192	231	−50.5499	−50.5499	8.8e−11
8	$(50,20,16,20)$	4	9	288	357	−73.1951	−73.1951	7.8e−09
9	$(20,8,4,8)$	4	5	104	141	−82.8222	−83.9831	1.2e−01
10	$(30,12,8,12)$	4	15	205	243	2176	−179.695	5.5e+00
11	$(40,16,12,16)$	4	17	201	202	−9.17031	−71.7339	1.2e+00
12	$(50,20,16,20)$	4	4	189	162	−110.073	−110.073	1.4e−12
13	$(20,8,4,8)$	8	4	88	106	−87.0353	−87.0353	4.9e−08
14	$(30,12,8,12)$	8	29	367	562	−124.033	−128.801	2.2e−01
15	$(40,16,12,16)$	8	6	150	146	−86.3114	−86.3114	1.9e−09
16	$(50,20,16,20)$	8	6	269	314	−47.3336	−47.3336	1.1e−10

Table 4.6: Numerical results for MINOS on AVI-QPECs

Discussion of AVI-QPEC results

Numerical results for ECOPT, NPSOL, SNOPT and MINOS for the first four groups of problems are shown in Tables 4.3–4.6. The notation is explained in Table 4.2.

In the following, we make some observations about the numerical results.

- All algorithms tested here perform well on the small- to medium scale AVI-QPEC problems. In particular, all four algorithms terminate with local solutions.

- Nondegenerate problems, Group 1, were generally easier for ECOPT in terms of iterations and function evaluations. There is no evidence on the effect degeneracy has on the efficiency of the NLP codes NPSOL, SNOPT and MINOS. The number of iterations and the quality of the solutions varies equally between nondegenerate and degenerate problems.

- For nonmonotone problems, Group 3, all algorithms had to work harder to identify the generated solution, as expected. All algorithms found local solutions, but NPSOL and ECOPT seemed to get closer to the best solutions than SNOPT and MINOS.

- Although the two SQP methods NPSOL and SNOPT are very similar in nature, it seems that SNOPT has more difficulty in handling complementarity constraints than NPSOL. In particular, the number of minor iterations increases for SNOPT with problem size. This is not well understood at this point. One explanation for this behavior could be the difference in how infeasible subproblems are handled by the two methods. Also, noting that QP subproblems are degenerate due to the complementarity constraint, different ways of handling degeneracy by the QP solvers could be a second reason. Further investigation is needed to provide a complete answer.

To facilitate a better overall comparison of the four algorithms, and also to compare our results to some algorithms tested in [JR99], we present Table 4.7. In this table, four measurements are exhibited. They are "# reach (x_{gen}, y_{gen})" or the number of

Algorithm	# reach (x_{gen}, y_{gen})	# reach best solution	# success	# QPs
ECOPT	12	14	16	205
NPSOL	12	13	16	528
SNOPT	12	10	16	931
MINOS	8	9	16	205
PIPA	11	10	15	305
PSQP/B	11	11	15	145

Table 4.7: Comparison of algorithms for AVI-QPECs

times the (infinity-norm) distance between the final point and (x_{gen}, y_{gen}) is less than 10^{-2}, "# reach best solution" which is the number of times each method converges to the best found solution, "# success", the number of problems where the method indicated successful termination, and " # QPs", the total number of quadratic programs solved over all successfully terminated runs. In the table, we include the results for the penalty interior point method PIPA and the piecewise sequential quadratic method PSQP (version B) from [JR99] for comparison.

Table 4.7 shows that on small to medium scale AVI-constrained problems, ECOPT and NPSOL perform better than the other four algorithms. Both converge to the best found solution more often than other methods, while ECOPT is, in addition, more efficient in terms of the overall number of subproblems that needed to be solved. Although MINOS seems efficient on this set of problems, it does not converge to, or get near the best solution for almost half of the problems. The other three methods, SNOPT, PIPA and PSQP/B are similar in performance, but do not find the best solution as often, on some problems they actually fail.

4.3.2 QPECs with LCP constraints

The quadratic problem with LCP constraints in $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ is given by:

$$\begin{aligned} \min_{x,y} \quad & \frac{1}{2}[x,y]P \begin{bmatrix} x \\ y \end{bmatrix} + c^T x + d^T y \\ \text{subject to} \quad & Gx + Hy + a \leq 0 \\ & 0 \leq y \perp Nx + My + q \geq 0, \end{aligned} \tag{4.2}$$

where $P \in \mathbb{R}^{(n+m)\times(n+m)}$, $c \in \mathbb{R}^n$, $d \in \mathbb{R}^m$, $G \in \mathbb{R}^{l\times n}$, $H \in \mathbb{R}^{l\times m}$, $a \in \mathbb{R}^l$, $N \in \mathbb{R}^{m\times n}$, $M \in \mathbb{R}^{m\times m}$, and $q \in \mathbb{R}^m$.

A special case of this problem is given if $H = 0$, so that the first-level constraints only depend on the first-level variables. These problems are termed implicit programs and some algorithms rely and exploit this special feature. The problems we test the algorithms on now are of this type.

In our numerical tests for the NLP codes, we observed some problems finding initial feasible points for the larger dimensional problems considered now. It is not difficult to find such a point for implicit programs. We first find a feasible point x^0 satisfying the inequality constraints $Gx + a \leq 0$. For the given x^0, we next find a feasible y^0 by solving the lower-level LCP $0 \leq y \perp Nx^0 + My + q \geq 0$. We use this procedure to provide the NLP codes with a feasible starting point. For ECOPT we use the random starting points.

Characteristics of LCP constrained problems

We will now present numerical results for implicit programs with LCP constraints. The problems tested here include the ones considered in [JR99], but we added some more problems including larger dimension and a larger number of first-level constraints. In detail, we test 24 problems in the following four groups.

Group 5, Problems 17–22. The parameters for the first group of LCP constrained problems are the same as Group 1 in the AVI constrained test set (see Table 4.1) except for `implicit` = 1 and `qpec_type` = 300. We use the same randomly generated starting points as before.

Group 6, Problems 23–28. The second group has the same parameters as Group 1, but has four times as many first-level variables x and first-level constraints.

Group 7, Problems 29–34. The same parameters as in Group 1 are used except that `second_deg` = 4 and `mix_deg` = 2.

Group 8, Problems 35–40. The last group has the same parameters as Group 3 but again more first-level variables x and more first-level constraints.

Discussion of LCP-QPEC results

Numerical results for the four methods for the second set of problems are shown in Tables 4.8–4.11. We use the same notation as before, see Table 4.2.

Overall performance of the four methods is very similar on these larger dimensional LCP-constrained problems compared to the small- and medium AVI constrained problems. ECOPT and NPSOL converge to the generated solution (x_{gen}, y_{gen}) for all 24 test problems. This is an indication that the problems have few local minimizers. SNOPT uses increasingly more minor iterations for higher dimensional problems. This actually leads to a few failures, denoted by F1 in Table 4.10, for problems with 250 and more second-level variables y. MINOS fails on half of the problems in this second set due to reaching its minor iteration limit also. We indicate this by F2 in Table 4.11

Similar to before, we present a summary of results in Table 4.12. Using the same notation, we compare the overall performance of the four methods. We leave out PIPA and PSQP/B in this comparison, since they have not been tested on all the problems considered here in [JR99]. Recall that "# QP's" only counts the major iterations for successfully completed problems, so that the low number of QP's solved

P	(m, n, l)	`deg`	Maj	Min	F-ev.	f	f_{gen}	Norm
17	$(50, 8, 4)$	0	10	31	11	−142.829	−142.829	1.1e−10
18	$(100, 8, 4)$	0	11	26	12	−664.389	−664.389	1.4e−09
19	$(150, 8, 4)$	0	9	27	10	−535.743	−535.743	6.0e−08
20	$(200, 8, 4)$	0	12	29	13	−109.595	−109.595	2.5e−08
21	$(250, 8, 4)$	0	14	42	16	−148.279	−148.279	2.4e−10
22	$(300, 8, 4)$	0	12	36	13	−846.126	−846.126	4.1e−08
23	$(50, 32, 16)$	0	9	60	10	−572.108	−572.108	2.3e−11
24	$(100, 32, 16)$	0	11	50	12	−25.9279	−25.9279	2.7e−08
25	$(150, 32, 16)$	0	12	58	13	−183.805	−183.805	8.8e−10
26	$(200, 32, 16)$	0	14	61	15	−340.145	−340.145	6.7e−09
27	$(250, 32, 16)$	0	12	65	13	−367.696	−367.696	2.2e−08
28	$(300, 32, 16)$	0	10	47	11	−109.995	−109.995	7.2e−08
29	$(50, 8, 4)$	4	13	29	17	−41.8764	−41.8764	1.7e−06
30	$(100, 8, 4)$	4	10	29	16	−599.936	−599.936	9.1e−06
31	$(150, 8, 4)$	4	11	35	13	−536.444	−536.444	4.9e−07
32	$(200, 8, 4)$	4	14	25	20	−23.7817	−23.7817	6.3e−07
33	$(250, 8, 4)$	4	12	33	13	−331.619	−331.619	9.4e−07
34	$(300, 8, 4)$	4	11	26	12	−739.468	−739.468	2.6e−07
35	$(50, 32, 16)$	4	18	62	21	−564.136	−564.136	3.7e−09
36	$(100, 32, 16)$	4	11	48	12	−155.606	−155.606	1.2e−07
37	$(150, 32, 16)$	4	16	68	21	−69.3098	−69.3098	1.4e−07
38	$(200, 32, 16)$	4	14	62	19	−304.964	−304.964	1.5e−07
39	$(250, 32, 16)$	4	18	80	27	−169.129	−169.129	6.1e−06
40	$(300, 32, 16)$	4	12	56	13	29.2539	29.2538	1.0e−03

Table 4.8: Numerical results for ECOPT on LCP-QPECs

P	(m,n,l)	deg	Maj	Min	F-ev.	f	f_{gen}	Norm
17	$(50,8,4)$	0	19	–	26	−142.829	−142.829	4.7e−08
18	$(100,8,4)$	0	19	–	28	−664.389	−664.389	3.6e−08
19	$(150,8,4)$	0	19	–	23	−535.743	−535.743	2.0e−08
20	$(200,8,4)$	0	15	–	18	−109.595	−109.595	3.0e−07
21	$(250,8,4)$	0	16	–	20	−148.279	−148.279	5.9e−07
22	$(300,8,4)$	0	14	–	18	−846.126	−846.126	9.8e−08
23	$(50,32,16)$	0	42	–	76	−572.108	−572.108	1.6e−07
24	$(100,32,16)$	0	34	–	58	−25.9279	−25.9279	2.1e−07
25	$(150,32,16)$	0	28	–	38	−183.805	−183.805	2.8e−07
26	$(200,32,16)$	0	21	–	25	−340.165	−340.145	3.1e−02
27	$(250,32,16)$	0	17	–	20	−367.696	−367.696	2.1e−07
28	$(300,32,16)$	0	20	–	23	−109.995	−109.995	1.1e−06
29	$(50,8,4)$	4	19	–	27	−41.8764	−41.8764	1.5e−08
30	$(100,8,4)$	4	19	–	25	−599.936	−599.936	1.5e−06
31	$(150,8,4)$	4	29	–	34	−536.444	−536.444	2.4e−07
32	$(200,8,4)$	4	15	–	18	−23.7817	−23.7817	1.3e−07
33	$(250,8,4)$	4	16	–	19	−331.619	−331.619	1.4e−08
34	$(300,8,4)$	4	15	–	19	−739.468	−739.468	4.3e−08
35	$(50,32,16)$	4	44	–	80	−564.136	−564.136	1.4e−07
36	$(100,32,16)$	4	36	–	52	−155.606	−155.606	1.6e−07
37	$(150,32,16)$	4	31	–	44	−69.3098	−69.3098	5.6e−07
38	$(200,32,16)$	4	22	–	27	−304.964	−304.964	8.9e−07
39	$(250,32,16)$	4	18	–	22	−169.129	−169.129	3.3e−07
40	$(300,32,16)$	4	26	–	31	29.2538	29.2538	1.4e−07

Table 4.9: Numerical results for NPSOL on LCP-QPECs

P	(m,n,l)	`deg`	Maj	Min	F-ev.	f	f_{gen}	Norm
17	$(50,8,4)$	0	18	439	24	−142.829	−142.829	2.3e−07
18	$(100,8,4)$	0	24	883	29	−664.389	−664.389	4.2e−08
19	$(150,8,4)$	0	30	1257	36	−535.743	−535.743	5.4e−07
20	$(200,8,4)$	0	14	6797	18	−109.595	−109.595	6.2e−07
21	$(250,8,4)$	0	20	7637	24	−148.279	−148.279	4.3e−07
22	$(300,8,4)$	0	F1	F1	F1	F1	F1	F1
23	$(50,32,16)$	0	22	663	28	−572.108	−572.108	1.0e−07
24	$(100,32,16)$	0	32	836	38	−25.9279	−25.9279	1.3e−06
25	$(150,32,16)$	0	44	1285	54	−183.805	−183.805	1.5e−06
26	$(200,32,16)$	0	139	2993	159	−340.145	−340.145	3.1e−05
27	$(250,32,16)$	0	F1	F1	F1	F1	F1	F1
28	$(300,32,16)$	0	F1	F1	F1	F1	F1	F1
29	$(50,8,4)$	4	24	481	30	−41.8764	−41.8764	5.0e−07
30	$(100,8,4)$	4	21	1020	26	−599.936	−599.936	3.3e−07
31	$(150,8,4)$	4	20	1471	25	−536.444	−536.444	4.8e−07
32	$(200,8,4)$	4	15	7378	19	−23.7817	−23.7817	1.5e−07
33	$(250,8,4)$	4	18	9525	23	−331.619	−331.619	3.2e−07
34	$(300,8,4)$	4	F1	F1	F1	F1	F1	F1
35	$(50,32,16)$	4	26	615	32	−564.136	−564.136	1.5e−07
36	$(100,32,16)$	4	13	909	19	−155.603	−155.606	7.3e−03
37	$(150,32,16)$	4	31	1527	36	−69.3098	−69.3098	7.0e−07
38	$(200,32,16)$	4	26	4136	31	−304.964	−304.964	3.0e−06
39	$(250,32,16)$	4	F1	F1	F1	F1	F1	F1
40	$(300,32,16)$	4	F1	F1	F1	F1	F1	F1

Table 4.10: Numerical results for SNOPT on LCP-QPECs

P	(m, n, l)	`deg`	Maj	Min	F-ev.	f	f_{gen}	Norm
17	$(50, 8, 4)$	0	10	309	433	−142.829	−142.829	9.1e−12
18	$(100, 8, 4)$	0	12	369	465	−664.389	−664.389	1.4e−10
19	$(150, 8, 4)$	0	15	776	865	−535.743	−535.743	2.5e−08
20	$(200, 8, 4)$	0	F2	F2	F2	F2	F2	F2
21	$(250, 8, 4)$	0	F2	F2	F2	F2	F2	F2
22	$(300, 8, 4)$	0	F2	F2	F2	F2	F2	F2
23	$(50, 32, 16)$	0	19	376	587	−569.849	−572.108	2.1e−01
24	$(100, 32, 16)$	0	17	686	957	−25.9279	−25.9279	3.0e−08
25	$(150, 32, 16)$	0	19	915	1018	−183.805	−183.805	4.7e−08
26	$(200, 32, 16)$	0	F2	F2	F2	F2	F2	F2
27	$(250, 32, 16)$	0	F2	F2	F2	F2	F2	F2
28	$(300, 32, 16)$	0	F2	F2	F2	F2	F2	F2
29	$(50, 8, 4)$	4	12	519	680	−41.8764	−41.8764	1.9e−11
30	$(100, 8, 4)$	4	11	511	547	−599.936	−599.936	8.1e−10
31	$(150, 8, 4)$	4	16	818	891	−536.444	−536.444	2.9e−11
32	$(200, 8, 4)$	4	F2	F2	F2	F2	F2	F2
33	$(250, 8, 4)$	4	F2	F2	F2	F2	F2	F2
34	$(300, 8, 4)$	4	F2	F2	F2	F2	F2	F2
35	$(50, 32, 16)$	4	8	269	369	−564.136	−564.136	1.4e−09
36	$(100, 32, 16)$	4	26	803	1171	−154.524	−155.606	1.7e−01
37	$(150, 32, 16)$	4	22	917	1204	−69.3098	−69.3098	2.8e−07
38	$(200, 32, 16)$	4	F2	F2	F2	F2	F2	F2
39	$(250, 32, 16)$	4	F2	F2	F2	F2	F2	F2
40	$(300, 32, 16)$	4	F2	F2	F2	F2	F2	F2

Table 4.11: Numerical results for MINOS on LCP-QPECs

Algorithm	# reach (x_{gen}, y_{gen})	# reach best solution	# success	# QPs
ECOPT	24	24	24	296
NPSOL	24	24	24	554
SNOPT	18	18	18	541
MINOS	12	12	12	187

Table 4.12: Comparison of algorithms for LCP-QPECs

for MINOS, for example, only corresponds to half of the problems.

Table 4.12 shows the superiority of ECOPT and NPSOL over SNOPT and MINOS for this second set of problems. ECOPT is again more efficient in terms of major iterations compared to NPSOL. It should be pointed out that the two other methods PIPA and PSQP/B show good results also for the subset of problems tested in [JR99]. Both methods converge to the generated points for the problems tested but tend to need more major iterations (QP's solved).

4.4 Some nonlinear test problems

In addition to quadratic problems with linear complementarity constraints, we also tested some nonlinear MPECs. These problems have either a nonlinear objective function, nonlinear constraint functions, or both.

The problems are part of MPECLIB [DF99], a library of MPEC problems coded in GAMS with an interface to MATLAB. We used the MATLAB interface to access the test problem data. For the details of the problems we refer to [DF99] and the references therein. We use the same parameters and update rules for ECOPT. This means in particular that we do not use second-order information of the constraint functions g and the equilibrium function F in the update of the Lagrangian Hessian. Some of the problems in MPECLIB have equilibrium conditions which take the form of mixed complementarity problems rather than NCPs. To handle this type of constraint, we generalize the definition of Φ to take this into account. Details of this are given in [Bil95] and [Pie98].

In Table 4.13, we summarize the results for MPECLIB. In particular we show the number of major iterations (Maj), function evaluations (F-ev.) and the final value of the objective function (f) for each problem.

Problem		Maj	F-ev.	f
bard1		6	7	17.00
desilva		10	13	−1.00
fjq1	a	12	29	3.2077
fjq1	b	15	43	3.2077
fjq2	a	6	9	3.4494
fjq2	b	6	8	3.4494
fjq3	a	14	23	4.6043
fjq3	b	10	14	4.6043
fjq4	a	14	21	6.5927
fjq4	b	12	18	6.5927
gauvin		5	6	20.00
hq1		5	6	−3266.67
mss	3	13	22	−5.3491
qvi		2	3	0

Table 4.13: Numerical results for ECOPT on MPECLIB

ECOPT solves all problems successfully and reaches the best known local minimum for all cases. The number of iterations and function evaluations also compares favorably with the results reported for other algorithms in [FJQ99, DF99, Stö99b] and the talk [LF00] given at the SIAM Conference in Atlanta in September 2000.

Before summarizing our numerical experiments, we consider a small academic example which has been proposed by Scholtes in [Sch01] as a problem on which smoothing methods, like ECOPT, would not be able to converge to the optimal solution.

4.5 Scholtes's counter example

The following problem has been suggested by Scholtes in [Sch01]:

$$\begin{aligned} \min_{u,v,w} \quad & \frac{1}{2}[(u-1)^2 + (v-2)^2 + (w+1)^2] \\ \text{subject to} \quad & u, v, w \geq 0, \\ & uw = 0, \\ & vw = 0. \end{aligned}$$

Theoretically, the smoothing approach cannot identify the minimizer $(1, 2, 0)$ of this problem since it generates a sequence of points with $u, v, w > 0$ and $uw = \mu$, $vw = \mu$ for $\mu > 0$. It is argued in [Sch01] that a smoothing method can only converge to a point on either the nonnegative w-axis or the diagonal in the nonnegative orthant of the (u, v) plane. The difficulty of this problem lies in the fact that the MPEC-LICQ is violated at every feasible point in the (u, v) plane.

Although, theoretically ECOPT generates search directions towards points with $u, v, w > 0$ and $uw = \mu$, $vw = \mu$ for $\mu > 0$, our implementation of ECOPT does not enforce the nonnegativity constraints on the complementary variables (in this case all variables u, v, and w) and does not solve the subproblems for each value of $\mu > 0$ exactly.

The problem can be cast as an LCP constraint QPEC in the form (4.2), with $x := w$, $y := (u, v)$, and

$$P := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad N := \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$c := 1, \qquad M := 0,$$

$$d := \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \qquad q := \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Running the problem with exactly the same implementation as before, we can solve

Iteration k	u^k	v^k	w^k	f^k	μ_k
0	0.00000	0.00000	0.0e+00	3.00000	1.0e−06
1	0.00141	0.00141	−5.4e−20	2.99576	1.0e−10
2	0.00341	0.00341	−1.8e−20	2.98977	1.0e−14
3	0.00683	0.00683	−6.0e−36	2.97956	1.0e−16
4	0.99980	2.00000	−2.1e−10	0.50000	1.0e−16
5	1.00000	2.00000	1.1e−16	0.50000	1.0e−16

Table 4.14: Iterations for a proposed counter example

the problem and find the optimal solution $(u, v, w) = (1, 2, 0)$. We tried many starting points and always converged to the optimal solution. In Table 4.14, we show a typical sequence of points and perturbations μ_k generated by ECOPT for this problem. Starting from the origin, it can be seen that at first the sequence converges to a point on the diagonal of the positive orthant of the (u, v) plane, but eventually converges to the optimal solution. Notice that the w variable becomes slightly negative, in particular in the fourth iteration. This seems to enable the step towards the solutions.

4.6 Summary of numerical results

We have presented extensive numerical comparison of different algorithms to solve mathematical programs with equilibrium constraints. Although the three standard nonlinear programming methods, in particular NPSOL, performed quite respectably, the smoothing approach incorporated into ECOPT seems more efficient. Furthermore, from our experience with the electricity model presented in the next chapter, we feel that ECOPT avoids local minima more effectively than NPSOL or the other methods. Future research could investigate this point more systematically.

One of the main features of ECOPT is the use of exact second derivative information of the smoothing function. At this point, no second-order information of the other constraint functions is used. A disaggregated Hessian approximation as discussed in [Gol99] can be applied in a similar way to accomplish this. Also, since the smoothing function has the "wrong curvature" as discussed earlier, we use an ad hoc method to ensure positive (semi-) definiteness of the matrix given to the QP subproblems. Although this works quite well for the problems tested here, we feel that a more sophisticated way of doing this could improve the results further. Another option would be to use the unmodified indefinite Hessian approximation and work with the nonconvex QP subproblem, see [MP99, Gol99].

Chapter 5

Forward Markets in Electricity

The lack of working forward markets for electricity represents one of the main obstacles to current deregulation efforts in the electricity industry. Inelasticity and uncertainty of electricity demand in real time gives producers market power in high demand time periods that leads to price spikes, which force the introduction of price caps in most deregulated electricity markets.

Under rate-of-return regulation, price hedges for consumers were an inherent feature of electricity rates. In a deregulated environment, this will not be the case any more; the expectation of regulators is that independently operated forward markets will arise to provide this hedge for consumers in this situation. Although economic theory suggests the development of forward markets in deregulated markets to hedge risk, this has not been the case in a number of markets, such as California. Recently, the Federal Energy Regulation Commission (FERC) [FER00] has identified the lack of an active forward market as one of the reasons for high electricity prices in particular in the California market, where all trading was forced into the spot market. In its order [FER00], the FERC commission considers forcing producers of electricity to sell in the forward market.

In this chapter we argue that with deterministic demand, there are situations in which it is individually rational for firms to enter the forward market. We consider the simple case of two firms separated from the demand by a possibly congested transmission line. We find equilibria for unlimited and limited transmission capacity

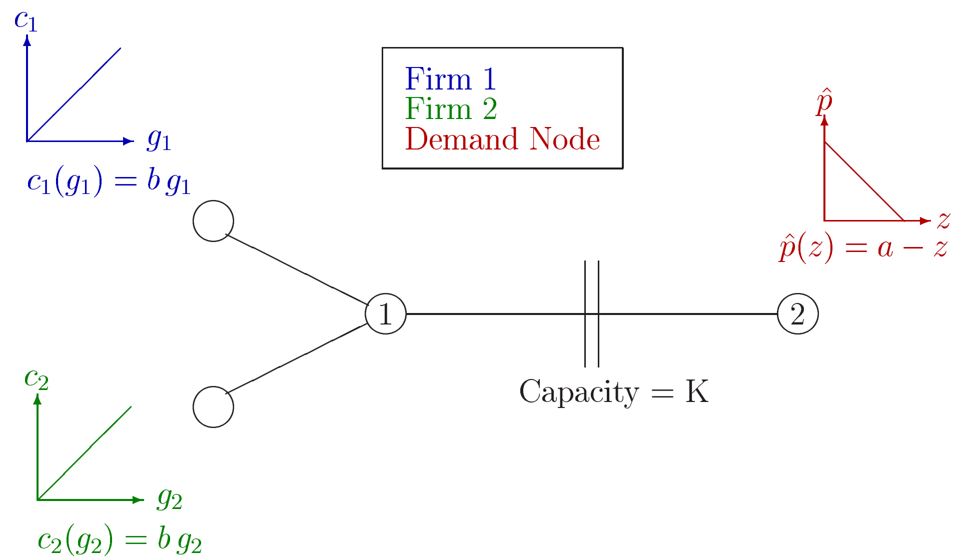

Figure 5.1: Two-node network with two firms and one demand node

separately in Sections 5.1 and 5.2. Producers will enter the forward market if the capacity of the line is above a certain threshold. Below that threshold, no firm has an incentive to enter the forward market and no forward sales will take place.

In Chapter 6 we expand the small example by presenting a general mathematical model for forward markets in electricity.

5.1 Forward markets with no transmission constraint

The following simple two-period model without transmission constraints was analyzed by Allaz [All92] and Allaz and Vila [AV93] to show a rationale for a forward market under certainty and perfect foresight. We will summarize the main results of [AV93] in condensed form in this section. The situation with no transmission constraints will serve as our base case for the more complicated models analyzed later.

In the simplest situation, see Figure 5.1, there are two firms at node 1 connected by a transmission line of capacity K to a demand at node 2. First, we will assume that $K = \infty$, so that the model is basically a one-node market and transmission prices can be neglected. It is assumed that in the first period (forward market), the two firms sell or buy forward contracts that call for delivery of electricity in the second period (spot market). We also assume that these forward contracts are *binding* and *observable* precommitments. In the second period, the firms play a Cournot game in quantity but payoff functions are modified by the position they took on the forward market.

Under perfect foresight, equilibrium requires the forward market to be efficient: The forward price as a function of the forward positions must be equal to the price that will result from the Cournot competition on the spot market given these positions. Therefore, no arbitrage is possible.

We will use the following notation:

g_1, g_2 : generation of firm 1 and 2

f_1, f_2 : firm's forward sales

$p, \hat{p}$: forward price p and spot-market price $\hat{p}$

u_1, u_2 : payoff functions of firm 1 and 2 in the second period

π_1, π_2 : overall profit functions of firm 1 and 2 for the two-stage game

c_1, c_2 : cost functions of firm 1 and 2

5.1.1 Spot-market equilibrium

Given the forward positions f_1 and f_2, the firms maximize the following payoff functions $u_1(f_1, f_2)$ and $u_2(f_1, f_2)$ in the second period:

$$[u_1(f_1, f_2)](g_1, g_2) = \hat{p}(g_1 + g_2)(g_1 - f_1) - c_1(g_1),$$
$$[u_2(f_1, f_2)](g_1, g_2) = \hat{p}(g_1 + g_2)(g_2 - f_2) - c_2(g_2).$$

Indeed, given that the first firm has already sold out f_1, it can only sell the quantity $(g_1 - f_1)$. If g_1 is less than f_1 then the firm must buy electricity from its competitor to meet its commitment of sales or, equivalently, it can buy back its forward position at the spot-market price.

To derive some analytical insight, we will assume that firms are symmetric, i.e., have the same cost structure, and both cost and demand functions to be linear and production capacity to be also unrestricted:

$$c_1(g_1) = b\,g_1, \quad c_2(g_2) = b\,g_2,$$
$$\hat{p}(z) = a - z,$$
$$0 < b < a.$$

The units of a and b are $\$/MWh$, and z in $\hat{p}(z)$ is premultiplied by a conversion factor of $\$/MWh^2$.

Given f_1 and f_2, the two competitors play the following game:

$$\max_{g_1,g_2} [u_1(f_1, f_2)](g_1, g_2) = (a - g_1 - g_2)\,(g_1 - f_1) - b\,g_1,$$
$$\max_{g_1,g_2} [u_2(f_1, f_2)](g_1, g_2) = (a - g_1 - g_2)\,(g_2 - f_2) - b\,g_2.$$

The first-order conditions for the players give:

$$g_1 = \frac{a - b + f_1 - g_2}{2}$$
$$g_2 = \frac{a - b + f_2 - g_1}{2}.$$

The key point is that the solution (reaction function) of player i is increasing in f_i; Indeed when a competitor has a short position (i.e., $f_i > 0$), then he is less "price sensitive" and therefore cares less about the price elasticity effect of an increase in generation. Note that the marginal revenue from selling a further MW on the spot market is $\hat{p}'(g_1+g_2)(x_i - f_i) + \hat{p}(g_1+g_2)$ and not $\hat{p}'(g_1+g_2)x_i + \hat{p}(g_1+g_2)$, as is typical in Cournot models, because the decrease in price necessary to sell this additional MW does not affect the revenue from forward sales.

The following proposition from [AV93] presents the Nash equilibrium quantities and price.

Proposition 5.1.1. *The Nash equilibrium in the spot market is unique with equilibrium quantities and price given by:*

$$g_1 = \frac{a-b+2f_1-f_2}{3}, \quad g_2 = \frac{a-b+2f_2-f_1}{3}, \quad \hat{p} = \frac{a+2b-f_1-f_2}{3}.$$

5.1.2 Emergence of forward markets

In the first period, when the orders f_1 and f_2 are submitted to the forward market, traders in this market know the spot price in the subsequent market will be $\hat{p}(g_1, g_2)$ as given in Proposition 5.1.1. Now suppose that in the forward market, prices are set by a Bertrand auction where several buyers (at least two) bid for the aggregated supply $f_1 + f_2$. The equilibrium outcome of this auction will generate a forward price $p(f_1, f_2)$ equal to the perfectly anticipated spot price $\hat{p}(g_1, g_2)$.

Suppose first that only one firm (firm 1) is allowed to trade forward ($f_2 = 0$). When making his trading decision, this producer knows that he will not make any arbitrage profit and that his payoff will be

$$\hat{p}[g_1(f_1,0) + g_2(f_1,0)]\, g_1(f_1,0) - c(g_1(f_1,0)). \tag{5.1}$$

Thus, firm 1 faces the following problem: Choose f_1 such that the Nash equilibrium outcome in the second period is optimal for him.

Proposition 5.1.2. *The equilibrium outcome in the forward market is the Stackelberg outcome of the Cournot duopoly game without a forward market when firm 1 is the leader:*

$$f_1 = \frac{a-b}{4}, \quad g_1 = \frac{a-b}{2}, \quad g_2 = \frac{a-b}{4}, \quad q = \frac{a+3b}{4}.$$

Proof: Substituting the results of Proposition 5.1.1 into (5.1) and maximizing with respect to f_1 gives the result. □

Thus, if one player has the opportunity to trade forward, he can improve his profit. There is a strategic incentive for trading forward. Total output goes up from $(2(a-b)/3)$ to $(3(a-b)/4)$. Therefore, taking the actions of the first trader as given, firm 2 will want to trade forward too in an attempt to reap similar profits. Hence we can see that both firms will sell forward part of their production. This will lead to the emergence of a forward market.

It can be argued (see [AV93]) that the opposite position in the forward market will be taken by speculators who do not make a profit, but will be able to get transaction costs for their provision of liquidity.

5.1.3 Forward-market equilibrium

We will now summarize the equilibrium in the forward market when both firms are allowed to participate.

Given the positions f_1 and f_2, the total profits of firm i are given by

$$\pi_i(f_1, f_2) = p(f_1, f_2)\, f_i + [u_i(f_1, f_2)](g_1(f_1, f_2), g_2(f_1, f_2)),$$

which can be rewritten as:

$$\pi_i(f_1, f_2) = [\hat{p}(f_1, f_2)\, g_i(f_1, f_2) - c(g_i(f_1, f_2))] + [p(f_1, f_2) - \hat{p}(f_1, f_2)]\, f_i.$$

Under perfect foresight there will be no arbitrage, so that $p(f_1, f_2) = \hat{p}(g_1, g_2)$. Hence:

$$\pi_i(f_1, f_2) = \hat{p}(g_1, g_2)\, g_i(f_1, f_2) - c(g_i(f_1, f_2)), \quad \text{for } i = 1, 2. \tag{5.2}$$

Proposition 5.1.3. *The only forward market equilibrium outcome is given by:*

$$g_1 = g_2 = \frac{2(a-b)}{5}, \quad f_1 = f_2 = \frac{a-b}{5}, \quad \hat{p} = b + \frac{a-b}{5}.$$

Hence, allowing forward trading decreases the firms' profits and increases social welfare.

Proof: After substituting the results of Proposition 5.1.1 into (5.2), it can be shown that both firms maximize (5.2) for the given values. □

It turns out that trading on the forward market represents a prisoner's dilemma for the two firms. When one of them succeeds in being the only producer to trade forward, he greatly benefits from doing so. However, when both firms trade forward, they both end up being worse off.

To extend the results of Allaz and Villa, we will consider the case of limited transmission capacity ($K < \infty$) in the next section. We will show that similar results hold, although the emergence of a forward market will not be given in any case.

5.2 Forward markets with transmission constraint

What happens in the presence of transmission constraints in the current model? To analyze this situation, we will assume that the transmission line of limited capacity K is owned by an independent grid operator who charges a wheeling fee w for transmitting power from node 1 to node 2. The transmission price is based on congestion.

Similar to the last section, we will first characterize the spot-market equilibrium given forward positions f_1 and f_2. Then we will examine the conditions under which firms participate in the forward market and summarize the equilibrium for the two-stage game.

5.2.1 Spot-market equilibrium

In the presence of a transmission constraints, the game consists of three players, the two firms plus the grid owner, and a market clearing condition.

More specifically, firm 1 solves

$$\max_{g_1} (a - g_1 - g_2)\,(g_1 - f_1) - (b + w)\,g_1, \tag{5.3}$$

and similarly for firm 2

$$\max_{g_2} (a - g_1 - g_2)\,(g_2 - f_2) - (b + w)\,g_2. \tag{5.4}$$

In this model, it is assumed that producers are transmission price-takers.

The grid owner is assumed to ration limited transmission interface capacity to maximize the value of the transmission services t (transmission from node 1 to node

2 in this case), as expressed by producers' willingness to pay. This behavior can be shown to be equivalent to having the grid choose values of t to maximize its revenue wt as if w were fixed, while respecting the interface constraint. It is also equivalent to a competitive market for transmission rights in which generators do not exercise market power [Sto99a].

The transmission owner therefore solves

$$\begin{aligned} \max_t \quad & w\,t \\ \text{subject to} \quad & t \leq K. \end{aligned} \tag{5.5}$$

The connection between the firms' and the transmission owner's problem is given by the market clearing condition:

$$g_1 + g_2 = t. \tag{5.6}$$

The transmission price w will be such that the market clearing condition is satisfied in equilibrium. This equilibrium is characterized by the optimality conditions of the players and the market clearing condition.

The first-order conditions of (5.3), (5.4) and (5.5) plus the market clearing condition (5.6) give rise to a mixed linear complementarity problem. These problems are in general harder to analyze than the simple model with no transmission constraints, but it is still possible to derive analytical expressions in this simple case.

We will split the analysis in two cases. First, if the transmission constraint is not binding, then $w = 0$ and we have the outcome of the last section. We see from there that in the second period, firms choose quantities

$$g_1 = \frac{a - b + 2f_1 - f_2}{3}, \quad \text{and} \quad g_2 = \frac{a - b + 2f_2 - f_1}{3}.$$

Therefore, the transmission line will be uncongested if

$$g_1 + g_2 = \frac{2(a-b) + f_1 + f_2}{3} \leq K.$$

In particular, if no forward sales are given ($f_1 = f_2 = 0$), the line will be congested if $\frac{2}{3}(a-b) \geq K$. So if $\frac{2(a-b)+f_1+f_2}{3} \leq K$, then without a forward market, the unconstrained Cournot equilibrium will be reached.

In the second case, if $\frac{2(a-b)+f_1+f_2}{3} > K$, the line will be congested. From the first-order conditions of the grid and the market clearing condition, it follows that

$$g_1 + g_2 = K. \tag{5.7}$$

The first-order conditions of the two firms are

$$(a - g_1 - g_2) - g_1 + f_1 - b - w = 0, \tag{5.8}$$

$$(a - g_1 - g_2) - g_2 + f_2 - b - w = 0. \tag{5.9}$$

Solving (5.7), (5.8) and (5.9) simultaneously yields:

$$g_1 = \frac{K + f_1 - f_2}{2}, \quad g_2 = \frac{K + f_2 - f_1}{2}, \quad \hat{p} = a - K,$$

$$\text{and} \quad w = \frac{2(a-b) + f_1 + f_2 - 3K}{2}.$$

Note that $w = 0$ if $K = \frac{2(a-b)+f_1+f_2}{3}$ and $w > 0$ for smaller transmission capacity K. We summarize the results in the following

Proposition 5.2.1. *The Nash Equilibrium in the spot market with a transmission line of capacity K is unique with equilibrium quantities and prices given by:*

(a) If $K \geq \frac{2(a-b)+f_1+f_2}{3}$, then

$$g_1 = \frac{a - b + 2f_1 - f_2}{3}, \quad g_2 = \frac{a - b + 2f_2 - f_1}{3}, \quad \hat{p} = \frac{a + 2b - f_1 - f_2}{3}, \quad w = 0.$$

(b) If $K < \frac{2(a-b)+f_1+f_2}{3}$*, then*

$$g_1 = \frac{K + f_1 - f_2}{2}, \quad g_2 = \frac{K + f_2 - f_1}{2}, \quad \hat{p} = a - K,$$

$$\text{and} \quad w = \frac{2(a-b) + f_1 + f_2 - 3K}{2}.$$

5.2.2 Emergence of forward markets

Similar to the case without transmission constraints, we will now analyze if and under what conditions a forward market will emerge. To do so, we will again first look at the case where only one firm is allowed to participate in the forward market. We will see that the outcome depends critically on the capacity K of the transmission line connecting the generators with the demand.

Suppose again that only firm 1 is allowed to make forward sales ($f_2 = 0$). Under perfect foresight, no arbitrage is possible, and firm 1's payoff function in the forward market will be:

$$\pi_1(f_1, 0) = \hat{p}(g_1(f_1, 0) + g_2(f_1, 0))\, g_1(f_1, 0) - c(g_1(f_1, 0)) - w\, g_1(f_1, 0).$$

The no arbitrage condition implies $p = \hat{p}$ so that firm 1 receives $\hat{p}(g_1(f_1, 0)+g_2(f_1, 0))$ for both forward and spot-market sales, and in addition the transmission price $w = w(g_1(f_1, 0), g_2(f_1, 0))$ will be known in the forward market also.

To find the optimal decision of firm 1 in the forward market, we will again consider two cases. First, we consider the case where the transmission line is uncongested.

This will happen exactly if $g_1(f_1, 0) + g_2(f_1, 0) \leq K$, where $g_1(f_1, 0)$ and $g_2(f_2, 0)$ are given by Proposition 5.2.1 as

$$g_1 = \frac{a - b + 2f_1}{3}, \quad g_2 = \frac{a - b - f_1}{3}, \quad \text{and } \hat{p} = \frac{a + 2b - f_1}{3}.$$

Therefore, the line is congested if $K \geq \frac{2(a-b)+f_1}{3}$, or $f_1 \leq 3K - 2(a - b)$. So if

$f_1 \leq 3K - 2(a-b)$, then $w = 0$ and the payoff function of firm 1 is given by:

$$\pi_1(f_1, 0) = \hat{p}(g_1(f_1, 0) + g_2(f_1, 0))\, g_1(f_1, 0) - c(g_1(f_1, 0)).$$

Substituting and simplifying gives the following payoff function for firm 1 if it is the only firm trading in the forward market:

$$\pi_1(f_1, 0) = \frac{1}{9}[(a-b)^2 + (a-b)f_1 - 2f_1^2], \quad \text{if } f_1 \leq 3K - 2(a-b).$$

Next, we consider the case where the line is congested in the spot market, i.e., $f_1 > 3K - 2(a-b)$. In this case, the payoff function is

$$\pi_1(f_1, 0) = \hat{p}(g_1(f_1, 0) + g_2(f_1, 0))\, g_1(f_1, 0) - c(g_1(f_1, 0)) - w g_1(f_1, 0),$$

where g_1, g_2, q and w are given by Proposition 5.2.1 as

$$g_1 = \frac{K + f_1}{2}, \quad g_2 = \frac{K - f_1}{2}, \quad \hat{p} = a - K, \quad w = \frac{2(a-b) + f_1 - 3K}{2}.$$

Substituting and simplifying gives the following payoff function:

$$\pi_1(f_1, 0) = \frac{1}{4}[K^2 - f_1^2] \quad \text{if } f_1 > 3K - 2(a-b).$$

Putting both cases together, the payoff function of firm 1 as the only firm in the forward market is given by:

$$\pi_1(f_1, 0) = \begin{cases} \frac{1}{9}[(a-b)^2 + (a-b)f_1 - 2f_1^2] & \text{if } f_1 \leq 3K - 2(a-b), \\ \frac{1}{4}[K^2 - f_1^2] & \text{otherwise.} \end{cases}$$

As an example, we will show the graph of a payoff function for $a = 100$, $b = 10$ and $K = 55, 65$ and 75 in Figure 5.2.

Given transmission capacity K, firm 1 chooses forward sales f_1 to maximize his payoff function π_1.

Proposition 5.2.2. *The equilibrium outcome with transmission capacity K and firm*

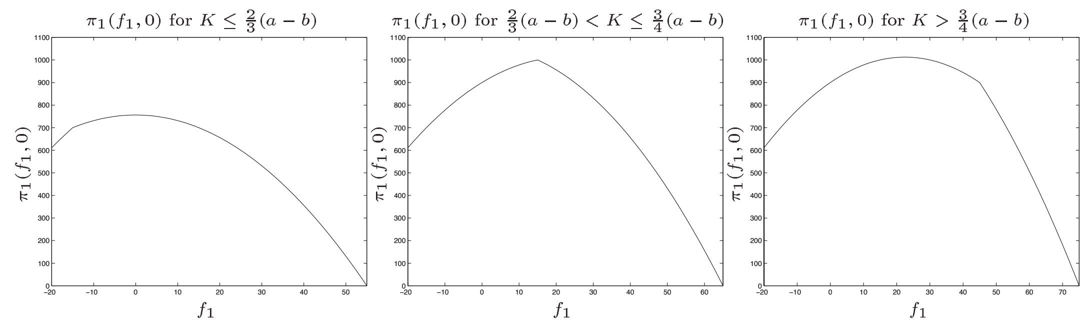

Figure 5.2: Payoff functions for low, medium, and high transmission capacity

1 as the only firm in the forward market is given by:

$$f_1 = \begin{cases} 0 & \textit{if } K \leq \frac{2}{3}(a-b), \\ 3K - 2(a-b) & \textit{if } \frac{2}{3}(a-b) < K \leq \frac{3}{4}(a-b), \\ \frac{1}{4}(a-b) & \textit{if } K > \frac{3}{4}(a-b). \end{cases}$$

Proof: If $k \leq \frac{2}{3}(a-b)$, then $3K - 2(a-b) \leq 0$ and the maximum over f_1 is found in the second part of the payoff function (see the first graph in Figure 5.2). In this case, $f_1 = 0$ will yield the maximum payoff for the two-stage game.

If $\frac{2}{3}(a-b) < K \leq \frac{3}{4}(a-b)$, then $3K - 2(a-b) > 0$ and the maximum over f_1 is found in the intersection of the two parts of the payoff function (see the second graph in Figure 5.2). In this case, $f_1 = 3K - 2(a-b)$ will yield the maximum payoff for the two-stage game.

In the third case, $K > \frac{3}{4}(a-b)$, the transmission limit does not come into play and the unconstrained result from Section 5.1 holds (see graph 3 of Figure 5.2). Therefore, $f_1 = \frac{1}{4}(a-b)$ in this case. □

The result of Proposition 5.2.2 has an important consequence for the emergence of forward markets under the assumed conditions. If the transmission line is already congested in a real-time market with no committed forward sales ($K \leq \frac{2}{3}(a-b)$), then no firm has an incentive to sell in the forward market to increase profits. In Section 6.4, we will see an example of a six-node network where a forward market develops, even though one transmission line is congested in the real-time market with no committed forward sales. Nevertheless, the small example indicates that expanding transmission capacity could support the development of a forward market in deregulated electricity markets.

If, on the other side, the transmission line is uncongested in a real-time market with no committed forward sales ($K > \frac{2}{3}(a-b)$), then the firms have an incentive of selling in the forward market. Similar to before, we will now analyze what kind of equilibria are possible if all firms are allowed to participate in the forward market under the presence of transmission constraints.

5.2.3 Forward-market equilibrium

Throughout the analysis, we will assume that the transmission capacity is large enough to allow the development of a forward market, i.e., $K > \frac{2}{3}(a-b)$. In this case, both players have an incentive to sell in the forward market.

If the transmission capacity is large enough to allow for the unconstrained forward equilibrium from Proposition 5.1.3, i.e., $K > \frac{4}{5}(a-b)$, this will obviously be the equilibrium.

Assume therefore that $\frac{2}{3}(a-b) < K < \frac{4}{5}(a-b)$. To find the equilibria in this case, we derive the payoff function of firm 1, given forward sales f_2 of firm 2. As in the previous case, the transmission line is uncongested if $g_1(f_1, f_2)+g_2(f_1, f_2) \leq K$; this is equivalent to $K \geq \frac{2(a-b)+f_1+f_2}{3}$, or $f_1 \leq 3K-2(a-b)-f_2$. So if $f_1 \leq 3K-2(a-b)-f_2$, then $w = 0$ and the payoff function of firm 1 is given by:

$$\pi_1(f_1, f_2) = \hat{p}(g_1(f_1, f_2) + g_2(f_1, f_2))\, g_1(f_1, f_2) - c(g_1(f_1, f_2)),$$

where g_1, g_2 and q are given by Proposition 5.2.1 as

$$g_1 = \frac{a-b+2f_1-f_2}{3}, \quad g_2 = \frac{a-b+2f_2-f_1}{3}, \quad \text{and } \hat{p} = \frac{a+2b-f_1-f_2}{3}.$$

Substituting and simplifying gives the following payoff function for firm 1, given forward sales f_2:

$$\pi_1(f_1, f_2) = \frac{1}{9}[(a-b)^2+2f_2^2-2af_2+(a-b-f_2)f_1-2f_1^2], \quad \text{if } f_1 \leq 3K-2(a-b)-f_2.$$

Next, we consider the case where the line is congested in the real-time market, i.e., $f_1 > 3K - 2(a-b) + f_2$. In this case, the payoff function is

$$\pi_1(f_1, f_2) = \hat{p}(g_1(f_1, f_2) + g_2(f_1, f_2))\, g_1(f_1, f_2) - c(g_1(f_1, f_2)) - w g_1(f_1, f_2),$$

where g_1, g_2, $\hat{p}$ and w are given by Proposition 5.2.1 as

$$g_1 = \frac{K + f_1 - f_2}{2}, \quad g_2 = \frac{K + f_2 - f_1}{2}, \quad \hat{p} = a - K, \quad w = \frac{2(a-b) + f_1 + f_2 - 3K}{2}.$$

Substituting and simplifying gives the following payoff function:

$$\pi_1(f_1, f_2) = \frac{1}{4}[K^2 - Kf_2 - f_1 f_2 - f_1^2], \quad \text{if } f_1 > 3K - 2(a-b) - f_2.$$

Putting both cases together, the payoff function of firm 1, given f_2, is:

$$\pi_1(f_1, f_2) = \begin{cases} \frac{1}{9}[(a-b)^2 + 2f_2^2 - 2af_2 \\ \qquad + (a - b - f_2)f_1 - 2f_1^2] & \text{if } f_1 \leq 3K - 2(a-b) - f_2, \\ \frac{1}{4}[K^2 - Kf_2 - f_1 f_2 - f_1^2] & \text{otherwise.} \end{cases}$$

Similarly, we can derive

$$\pi_2(f_1, f_2) = \begin{cases} \frac{1}{9}[(a-b)^2 + 2f_1^2 - 2af_1 \\ \qquad + (a - b - f_1)f_2 - 2f_2^2] & \text{if } f_2 \leq 3K - 2(a-b) - f_1, \\ \frac{1}{4}[K^2 - Kf_1 - f_1 f_2 - f_2^2] & \text{otherwise.} \end{cases}$$

Given transmission capacity K and forward sales f_j, $j \neq i$, firm i chooses forward sales f_i to maximize π_i. As in Proposition 5.2.2, optimal forward sales for firm i are given by

$$f_i = \max(0, 3K - 2(a-b) - f_j), \quad \text{if } \frac{2}{3}(a-b) < K \leq \frac{3}{4}(a-b).$$

This means, if $\frac{2}{3}(a-b) < K < \frac{4}{5}(a-b)$, there will be multiple equilibria for which

$f_1+f_2 = 3K-2(a-b)$. If one firm is able to obtain a large share in the forward market, the best the other firm can do is to take the rest. Factors outside the simplified model considered here will determine the distribution of sales and profits between the firms.

We will summarize the equilibrium outcome in the constrained case in the following Proposition.

Proposition 5.2.3. *The forward market equilibrium outcome with limited transmission capacity K is given by:*

(a) If $K < \frac{2}{3}(a-b)$, then the firms will not commit to any forward sales so that

$$\begin{aligned} f_1 &= f_2 = 0, \\ g_1 &= g_2 = \frac{1}{2}K, \\ \hat{p} &= a - K, \\ w &= \frac{2(a-b) - 3K}{2}. \end{aligned}$$

(b) If $\frac{2}{3}(a-b) \leq K < \frac{4}{5}(a-b)$, then multiple equilibria exist such that

$$\begin{aligned} f_1, f_2 &\geq 0, \\ f_1 + f_2 &= 3K - 2(a-b), \\ g_1 &= a - b + f_1 - K, \\ g_2 &= a - b + f_2 - K, \\ \hat{p} &= a - K, \\ w &= 0. \end{aligned}$$

(c) If $K \geq \frac{4}{5}(a-b)$*, then*

$$g_1 = g_2 = \frac{2(a-b)}{5},$$
$$f_1 = f_2 = \frac{a-b}{5},$$
$$\hat{p} = b + \frac{a-b}{5},$$
$$w = 0.$$

Proof: The proof is very similar to that of Proposition 5.2.2. □

Chapter 6

A General Forward Market Model

We have seen so far that even in the simplest case of one transmission line and no capacity constraints, it takes some effort to find equilibria for the two-stage game of forward and spot-market electricity sales. More general situations can only be analyzed by more general models. In this chapter we will present such a model that is scalable in the number of firms, nodes and transmission lines and includes transmission and generation capacity limits as well as an accurate network representation using Kirchhoff's laws. Although our model is quite general, further extensions are possible and will be mentioned later.

We will first present a Cournot model of the spot market, which generalizes the simple example we dealt with so far. After that we will consider the payoff function in the forward market and present a mathematical formulation of the forward problem in terms of a mathematical program with equilibrium constraints. Notice that a summary of the notation and commonly used definitions is given in the appendix.

6.1 Spot-market equilibrium

In this section, we will summarize the assumptions about market participants in our model. We will derive the firms' and grid owner's optimization problem, which together with the market clearing conditions will be the equilibrium constraints representing the spot-market in the forward-market model. The following Nash-Cournot model is based on the model presented by Hobbs in [Hob00] (see also [Met00]), but has some notable differences.

First, in our context, the model represents only the spot market and not the equilibrium outcome of the whole market. We assume that firms have committed to binding forward sales contracts before they enter the spot market. In this context, the spot-market equilibrium will serve as the constraints in an optimization problem. As an extension, one could also consider uncertainty in the demand, represented by scenarios in the spot market. To make the model more realistic, we have added increasing marginal cost functions (in contrast to constant marginal cost) for the firms, although the same effect could have been achieved by increasing the number of generation units and varying marginal cost.

Firm ℓ owns power generating facilities located at nodes $i \in N$ of the network. The indices i and j designate nodes, $c_{i\ell} + b_{i\ell}\, g_{i\ell}$ is the marginal cost (in \$/MW) of generating $g_{i\ell}$ (in MW). The capacity of a generator is $CAP_{i\ell}$.

Collectively, consumers at a node i consume quantity q_i, which is price responsive. In this model, we will assume linear demand functions $p_i(q_i) = P_i^0 - (P_i^0/Q_i^0)q_i$, with P_i^0 (in \$/MWh) and Q_i^0 (in MW) being the positive price and quantity intercept, respectively. Nonlinear demand functions would give rise to a nonlinear complementarity problem as the subproblem, which generally results in models that are more complicated to solve.

In real time, $s_{j\ell}$ is the quantity shipped by firm ℓ to consumers at node j. Furthermore, there is a forward market in which producer ℓ contracts to deliver $f_{j\ell}$ to consumers at node j in real time. Therefore, in real time, the producer receives revenues for the amount $(s_{j\ell} - f_{j\ell})$. Assuming market clearing and no arbitrage, $\sum_\ell s_{j\ell} = q_j$. If there is arbitrage, then $\sum_\ell s_{j\ell} + a_j = q_j$, where a_j is the net amount of power sold by arbitragers to node j. The generators determine the level of sales to each node, and then request transmission service from the grid. We impose an energy balance on each firm ($\sum_i g_{i\ell} = \sum_j s_{j\ell}$).

The owner of the grid charges a wheeling fee w_i (in \$/MWh) for transmitting power from node i to an arbitrary hub node. (For simplicity, it is assumed that there is neither generation nor consumption at the hub). Because of the linearity of the DC network, all generation and sales can be modeled as being routed through the hub node. A firm pays w_i to get power to the hub from a generator at i and then pays $-w_j$ to convey power for sale from the hub to customers at j. Thus, the total cost of transmitting power from a generator at i to the point of sale at j is $w_i - w_j$. The total transmission service that the grid provides for power transferred from node i to the hub is t_i MW (which may be negative). Consistent with the linear DC approximation, flows through transmission line k are modeled using *power transmission distribution factors* (we will denote the matrix of PTDF's by D); i.e., the net MW flow through k is $\sum_i D_{ik}\, t_i$. The lower and upper bounds on real power flows through an interface k are $-T_k^-$ and T_k^+. We assume there are no losses and that congestion is the basis for pricing.

The owner of the grid is assumed to ration limited interface capacity to maximize the value of the transmission services t_i, as expressed by generators' willingness to pay. This behavior can be shown to be equivalent to having the grid to choose values of t_i to maximize its revenue $\sum_i w_i\, t_i$ as if the w_i were fixed, while respecting interface constraints, see [Hob00, Sto99a].

Arbitrage

A final assumption concerns arbitragers. In the present context, arbitrage is the practice of buying power at one node and selling it for profit at another node. This can occur when the cost of shipping is smaller than the price difference between two nodes. If the price at B is more than the price at A plus the cost of shipping it from A to B, then it is profitable to sell from A to B. If it is strictly less, then it would be profitable to sell power generated at B to A. By allowing arbitrage to take place, the price difference between two nodes will be exactly the cost of shipment between the two nodes at equilibrium. If this is not the case, then a profitable opportunity for the arbitrager would have been ignored.

We will model arbitragers in the presented framework by recognizing that arbitragers will erase price differences between nodes so that in equilibrium $p_i(q_i) + w_i = p_j(q_j) + w_j = p_H$, the hub price. We will denote the amount of arbitrage sold at node

j by a_j, so $a_j < 0$ means that power is bought and transferred out of node j. This assumptions leads also to simplifications in the firms' problem since we can assume that all sales in the spot market are at the hub. Transmission only has to be paid to get energy to the hub.

In the following, we will summarize the firms' and grid's optimization problem as well as the market clearing conditions. The KKT conditions of these problem will result in the equilibrium problem for the spot market.

Firm's problem

The following quadratic optimization model states that in real time, given contracted forward sales $f_{i\ell}$, firm ℓ chooses generation $g_{i\ell}$ and sales $s_{i\ell}$ in order to maximize profit:

$$
\begin{array}{lll}
\max\limits_{s_\ell, g_\ell, a_\ell, p_\ell} & \sum\limits_{i \in N} p_\ell \left(s_{i\ell} - f_{i\ell}\right) - \sum\limits_{i \in N} \left[c_{i\ell} + \frac{1}{2} b_{i\ell}\, g_{i\ell} + w_i\right] g_{i\ell} & \\
\text{subject to} & g_{i\ell} \le CAP_{i\ell}, & \forall i \in N \\
& P_i^0 - (P_i^0/Q_i^0)(\sum\limits_{t \in F} s_{jt} + a_{i\ell}) = p_\ell - w_i, & \forall i \in N \\
& \sum\limits_{i \in N} (s_{i\ell} - g_{i\ell}) = 0 & \\
& \sum\limits_{i \in N} a_{i\ell} = 0 & \\
& s_{i\ell},\, g_{i\ell} \ge 0, & \forall i \in N.
\end{array}
$$

Grid owner's problem

The grid owner chooses t_i to maximize its profit from bilateral transactions, adopting the naive Nash-Bertrand assumption that it cannot affect the fees it gets for providing transmission services:

$$\begin{aligned} \max_{y} \quad & \sum_{i\in N} w_i\, t_i \\ \text{subject to} \quad & \sum_{i\in N} D_{ik}\, t_i \le T_k^+, \quad \forall k \in A \\ & -\sum_{i\in N} D_{ik}\, t_i \le T_k^-, \quad \forall k \in A. \end{aligned}$$

Market clearing

The total transmission service demanded by each generator ℓ and arbitragers from node i to the hub must equal the transmission service the grid provides to that node:

$$\sum_{\ell\in F} (g_{i\ell} - s_{i\ell}) - a_{i\ell} = t_i, \quad \forall i \in N, \ell \in F.$$

Elimination of sales variables

To simplify the model, the firms' problem can also be stated in the generation variables $g_{i\ell}$ alone, resulting in an optimization problem with only half the number of variables. Recall the original optimization problem

$$\max \quad \sum_{i\in N} p_\ell\,(s_{i\ell} - f_{i\ell}) - \sum_{i\in N} \left[c_{i\ell} + \frac{1}{2}\, b_{i\ell}\, g_{i\ell} + w_i \right] g_{i\ell} \tag{6.1}$$

$$\text{subject to} \quad g_{i\ell} \le CAP_{i\ell}, \qquad \forall i \in N \tag{6.2}$$

$$P_i^0 - (P_i^0/Q_i^0)(\sum_{t\in F} s_{jt} + a_{i\ell}) = p_\ell - w_i, \qquad \forall i \in N \tag{6.3}$$

$$\sum_{i\in N} (s_{i\ell} - g_{i\ell}) = 0 \tag{6.4}$$

$$\sum_{i\in N} a_{i\ell} = 0 \tag{6.5}$$

$$s_{i\ell},\ g_{i\ell} \ge 0, \qquad \forall i \in N. \tag{6.6}$$

It can be shown that if $x^* = (s_\ell^*, g_\ell^*, a_\ell^*, p_\ell^*)$ is an optimal solution to the firms optimization problem, then the point $x' = (g_\ell^*, g_\ell^*, a_\ell^* + s_\ell^* - g_\ell^*, p_\ell^*)$ is a possibly alternate optimal solution to the problem.

Proposition 6.1.1. *If* $x^* = (s^*_\ell, g^*_\ell, a^*_\ell, p^*_\ell)$ *is an optimal solution to the firms optimization problem, then the point* $x' = (g^*_\ell, g^*_\ell, a^*_\ell + s^*_\ell - g^*_\ell, p^*_\ell)$ *is an alternate optimal solution to the problem.*

Proof: First, we show the points yield the same objective value, then we will show that x' is feasible. Consider the objective function at x':

$$\sum_{i\in N}\left[P_i^0 - \frac{P_i^0}{Q_i^0}\left(g^*_{i\ell} + \sum_{t\in F\setminus f} s_{it} + a^*_{i\ell} + s^*_{i\ell} - g^*_{i\ell}\right) + w_i\right](g^*_{i\ell} - f_{i\ell})$$
$$- \sum_{i\in N}\left[c_{i\ell} + \frac{1}{2}\, b_{i\ell}\, g^*_{i\ell} + w_i\right] g^*_{i\ell}$$
$$= \sum_{i\in N}\left[P_i^0 - \frac{P_i^0}{Q_i^0}\left(s^*_{i\ell} + \sum_{t\in F\setminus f} s_{it} + a^*_{i\ell}\right) + w_i\right](g^*_{i\ell} - f_{i\ell})$$
$$- \sum_{i\in N}\left[c_{i\ell} + \frac{1}{2}\, b_{i\ell}\, g^*_{i\ell} + w_i\right] g^*_{i\ell}$$

(by constraint (6.4))

$$= \sum_{i\in N} p^*_\ell (g^*_{i\ell} - f_{i\ell}) - \sum_{i\in N}\left[c_{i\ell} + \frac{1}{2}\, b_{i\ell}\, g^*_{i\ell} + w_i\right] g^*_{i\ell}$$
$$= p^*_\ell \sum_{i\in N}(g^*_{i\ell} - f_{i\ell}) - \sum_{i\in N}\left[c_{i\ell} + \frac{1}{2}\, b_{i\ell}\, g^*_{i\ell} + w_i\right] g^*_{i\ell}$$

(by constraint (6.3))

$$= p^*_\ell \sum_{i\in N}(s^*_{i\ell} - f_{i\ell}) - \sum_{i\in N}\left[c_{i\ell} + \frac{1}{2}\, b_{i\ell}\, g^*_{i\ell} + w_i\right] g^*_{i\ell}$$
$$= \sum_{i\in N}\left[P_i^0 - \frac{P_i^0}{Q_i^0}\left(s^*_{i\ell} + \sum_{t\in F\setminus f} s_{it} + a^*_{i\ell}\right) - w_i\right](s^*_{i\ell} - f_{i\ell})$$
$$- \sum_{i\in N}\left[c_{i\ell} + \frac{1}{2}\, b_{i\ell}\, g^*_{i\ell} + w_i\right] g^*_{i\ell}.$$

Thus, the objective values at x^* and x' are equal. Also, within the derivation above, one can see that the price functions yield the same values at x^* and x'. Thus the point x' will also satisfy constraint (6.4). Now, we show that x' satisfies constraint

(6.5):

$$\sum_{i\in N}(a^*_{i\ell} + s^*_{i\ell} - g^*_{i\ell}) = \sum_{i\in N} a^*_{i\ell} + \sum_{i\in N}(s^*_{i\ell} - g^*_{i\ell})$$
$$\text{(By equation (6.3):)} = \sum_{i\in N} a^*_{i\ell}.$$

It is trivial to show that x' also satisfies the other constraints. □

This means that for any optimal solution, there is an alternate solution for which $s = g$. This allows one to reduce the problem by eliminating the s variables in each firm's optimization problem. As a result, the exogenous sales variables for the other firms can now also be written in terms of g, leading to a reduced optimization problem.

For each $\ell \in F$,

$$\max \quad \sum_{i\in N} p_\ell\,(g_{i\ell} - f_{i\ell}) - \sum_{i\in N}\left[c_{i\ell} + \frac{1}{2}\,b_{i\ell}\,g_{i\ell} + w_i\right] g_{i\ell} \tag{6.7}$$

$$\text{subject to} \quad g_{i\ell} \le CAP_{i\ell}, \qquad \forall i \in N \tag{6.8}$$

$$P^0_i - (P^0_i/Q^0_i)(\sum_g g_{jg} + a_{i\ell}) = p_\ell - w_i, \qquad \forall i \in N \tag{6.9}$$

$$\sum_{i\in N} a_{i\ell} = 0 \tag{6.10}$$

$$g_{i\ell} \ge 0, \qquad \forall i \in N. \tag{6.11}$$

Simplified firm's problem

Before considering the KKT conditions of the above optimization problems, we will eliminate further variables in the firm's problem. In particular, it is possible to eliminate the arbitrage variables $a_{i\ell}$ and the hub prices p_ℓ, see [Met00]. To do so, consider equations (6.9) and (6.10) in matrix notation:

$$\begin{aligned} P^0 - Q(a_\ell + \hat{g}) + w - p_\ell \, e_n &= 0, \\ e_n^T \, a_\ell &= 0, \end{aligned}$$

where Q is the $n \times n$ diagonal matrix with entries $Q_{ii} = \frac{P_i^0}{Q_i^0}$, $\hat{g}$ is the n vector of sums of generation at node i, i.e., $\hat{g}_i = \sum_\ell g_{i\ell}$, and e_n is the vector of n ones. Solving for a_ℓ and p_ℓ, we obtain

$$\begin{bmatrix} a_\ell \\ p_\ell \end{bmatrix} = \begin{bmatrix} Q & e_n \\ e_n^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} -Q\hat{g} + w + P^0 \\ 0 \end{bmatrix}.$$

Since Q is a positive definite diagonal matrix, it is easy to see that

$$\begin{bmatrix} Q & e_n \\ e_n^T & 0 \end{bmatrix}^{-1} = \begin{bmatrix} L & d \\ d^T & -\hat{d} \end{bmatrix},$$

where L is an $n \times n$ matrix, d is an $n \times 1$ vector, and $\hat{d}$ is a constant. In particular, $\hat{d}$, d and L are defined as follows:

$$\begin{aligned} \hat{d} &:= \frac{1}{\sum_{k=1}^n \frac{Q_k^0}{P_k^0}}, \\ d_i &:= \hat{d}\frac{Q_i^0}{P_i^0}, \quad \text{for } i \in N, \\ L_{ii} &:= \hat{d}\frac{Q_i^0}{P_i^0} \sum_{j \in N \setminus \{i\}} \frac{Q_j^0}{P_j^0}, \quad \text{for } i \in N, \\ L_{ij} &:= -\hat{d}\frac{Q_i^0 Q_j^0}{P_i^0 P_j^0}, \quad \text{for } i, j \in N, i \neq j. \end{aligned}$$

Note that this implies

$$LQ + de_n^T = I_n.$$

Since Q is positive definite and $-\hat{d} < 0$, it is also clear that L is positive definite. This will be important later on when we analyze the whole model, see Proposition 6.1.3.

Now, we can write a_ℓ and p_ℓ explicitly as follows:

$$\begin{bmatrix} a_\ell \\ p_\ell \end{bmatrix} = \begin{bmatrix} -LQ\hat{g} + Lw + LP^0 \\ -d^T Q\hat{g} + d^T w + d^T P^0 \end{bmatrix} = \begin{bmatrix} (de_n^T - I_n)(\hat{g}) + Lw + LP^0 \\ -\hat{d}^T e_n^T(\hat{g}) + d^T w + d^T P^0 \end{bmatrix}.$$

Rewriting in subscript form, we have:

$$a_{i\ell} = -\sum_{\ell \in F} g_{i\ell} + d_i G + \sum_{j \in N} L_{ij} w_j + \sum_{j \in N} L_{ij} P_j^0 \quad \forall i \in N, \quad \text{and} \tag{6.12}$$

$$p_\ell = -\hat{d}G + \sum_{j \in N} d_j w_j + \sum_{j \in N} d_j P_j^0, \tag{6.13}$$

where G is equal to total generation in the market, i.e., $G := \sum_{i \in N} \sum_{\ell \in F} g_{i\ell}$. These two equations show that both $a_{i\ell}$ and p_ℓ are independent of ℓ. Next, we substitute for p_ℓ and $a_{i\ell}$ in the reduced optimization problem and arrive at the following system:

$$\begin{aligned} \max \quad & \left[-\hat{d}G + \sum_{j \in N} d_j w_j + \sum_{j \in N} d_j P_j^0\right] \sum_{i \in N} (g_{i\ell} - f_{i\ell}) - \sum_{i \in N} \left[c_{i\ell} + \frac{1}{2} b_{i\ell}\, g_{i\ell} + w_i\right] g_{i\ell} \\ \text{subject to} \quad & g_{i\ell} \leq CAP_{i\ell}, \quad \forall i \in N \\ & g_{i\ell} \geq 0, \quad \forall i \in N. \end{aligned} \tag{6.14}$$

Without the sales variables, the market clearing conditions reduce to $-a_{i\ell} = t_i$ so that without the arbitrage variables they take the following form:

$$t_i = \sum_{\ell \in F} g_{i\ell} - d_i G - \sum_{j \in N} L_{ij} w_j - \sum_{j \in N} L_{ij} P_j^0 \quad \forall i \in N, \tag{6.15}$$

where G denotes total generation in the market.

Spot-market LCP

The first-order optimality conditions for the reduced firms' problem (6.14) are then:

For all $i \in N$, $\ell \in F$:

$$\begin{aligned} 0 \leq g_{i\ell} \perp & - \hat{d} \sum_{j \in N} g_{jf} + \hat{d} \sum_{j \in N} f_{j\ell} - \hat{d}G + \sum_{j \in N} d_j w_j + \sum_{j \in N} d_j P_j^0 \\ & - w_i - c_{i\ell} - b_{i\ell} g_{i\ell} - \gamma_{i\ell} \leq 0, \\ 0 \leq \gamma_{i\ell} \perp & \ g_{i\ell} - CAP_{i\ell} \leq 0. \end{aligned}$$

We can write down the KKT conditions for the linear optimization problem of the grid owner. After we introduce multipliers λ_k^- and λ_k^+ for the lower and upper bounds on transmission through interface k, respectively, the KKT conditions for the grid owner are:

For all $i \in N$, $k \in A$:

$$0 = w_i - \sum_{k \in A} D_{ik} (\lambda_k^+ - \lambda_k^-) - \Delta, \tag{6.16}$$

$$0 \leq \lambda_k^+ \perp \sum_{i \in N} D_{ik}\, t_i \leq T_k^+, \tag{6.17}$$

$$0 \leq \lambda_k^- \perp - \sum_{i \in N} D_{ik}\, t_i \leq T_k^-, \tag{6.18}$$

$$0 = \sum_{i \in N} t_i. \tag{6.19}$$

We can drop Δ in the above system by defining $w_i' = w_i - \Delta$, $\alpha_\ell' = \alpha_\ell + \Delta$, and $p_\ell' = p_\ell + \Delta$.

Before putting the whole model together, we can also eliminate the w and t variables using (6.16) and (6.15):

$$w_i = \sum_{k \in A} D_{ik} (\lambda_k^+ - \lambda_k^-), \tag{6.20}$$

$$t_i = \sum_{\ell \in F} g_{i\ell} - d_i G - \sum_{j \in N} L_{ij} w_j - \sum_{j \in N} L_{ij} P_j^0. \tag{6.21}$$

Since $t_i = -a_{i\ell}$ and by (6.10), $\sum_{i \in N} a_{i\ell} = 0$, for all ℓ, we can drop (6.19). The resulting equilibrium model for the spot market, given forward sales $f_{i\ell}$, is a linear complementarity problem in the variables $(g_\ell, \gamma_\ell, \lambda_k^+, \lambda_k^-)$:

For all $i \in N$, $\ell \in F$, and $k \in A$:

$$\begin{aligned}
0 \le g_{i\ell} \perp & - \hat{d} \sum_{j \in N} g_{jf} + \hat{d} \sum_{j \in N} f_{jf} - \hat{d} G + \sum_{k \in A} \sum_{j \in N} d_j D_{jk} (\lambda_k^+ - \lambda_k^-) + \sum_{j \in N} d_j P_j^0 \\
& - \sum_{k \in A} D_{ik} (\lambda_k^+ - \lambda_k^-) - c_{i\ell} - b_{i\ell} g_{i\ell} - \gamma_{i\ell} \le 0, \\
0 \le \gamma_{i\ell} \perp & \; g_{i\ell} - CAP_{i\ell} \le 0, \\
0 \le \lambda_k^+ \perp & \sum_{j \in N} \sum_{\ell \in F} D_{jk} \, g_{jf} - \sum_{j \in N} d_j D_{jk} \, G - \sum_{j \in N} \sum_{j' \in N} D_{jk} \, L_{jj'} \, P_{j'}^0 \\
& - \sum_{h \in A} L_{kh}^D (\lambda_k^+ - \lambda_k^-) - T_k^+ \le 0, \\
0 \le \lambda_k^- \perp & - \sum_{j \in N} \sum_{\ell \in F} D_{jk} \, g_{jf} + \sum_{j \in N} d_j D_{jk} \, G + \sum_{j \in N} \sum_{j' \in N} D_{jk} \, L_{jj'} \, P_{j'}^0 \\
& + \sum_{h \in A} L_{kh}^D (\lambda_k^- - \lambda_k^+) - T_k^- \le 0,
\end{aligned}$$

where L^D is the $l \times l$ matrix $D^T L D$, and L_{kh}^D is the $(k,h)^{th}$ entry of L^D. Note that L^D is positive semidefinite since L is positive definite.

Spot-market LCP in matrix notation

To ease the analysis in the following, we will also consider the above formulation in vector/matrix form. To do so, we will first introduce some more notation. Let

$$B(if, jh) := \begin{cases} 2\hat{d} + b & \text{if } f = h, \\ \hat{d} & \text{otherwise.} \end{cases}$$

We will denote the vector of all $g_{i\ell}$, $i \in N$, $\ell \in F$ by g and similar for γ and f. We will also make use of the following block structured matrix $\tilde{D}$:

$$\tilde{D} := \begin{bmatrix} D \\ D \\ \vdots \\ D \end{bmatrix}.$$

Furthermore, let E_D denote the $nm \times l$ matrix

$$E_D := \begin{bmatrix} e_n d^T D \\ e_n d^T D \\ \vdots \\ e_n d^T D \end{bmatrix},$$

and E_P the $nm \times 1$ vector

$$E_P := \begin{bmatrix} e_n d^T P^0 \\ e_n d^T P^0 \\ \vdots \\ e_n d^T P^0 \end{bmatrix}.$$

Also, if we let E_n be the matrix of all ones of size $n \times n$, then we define the following block matrix

$$\tilde{E} := \begin{bmatrix} E_n & 0 & \cdots & 0 \\ 0 & E_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_n \end{bmatrix}.$$

Notice that we switch the inequality signs from before at this point to express the problem in standard form for LCP's:

$$0 \le g \quad \perp -\hat{d}\tilde{E}f + Bg + \gamma + (\tilde{D} - E_D)(\lambda^+ - \lambda^-) - E_P + c \ge 0, \tag{6.22}$$

$$0 \le \gamma \quad \perp \quad -g + CAP \ge 0, \tag{6.23}$$

$$0 \le \lambda^+ \perp -(\tilde{D} - E_D)^T g + L^D(\lambda^+ - \lambda^-) + T^+ + D^T LP^0 \ge 0, \tag{6.24}$$

$$0 \le \lambda^- \perp (\tilde{D} - E_D)^T g - L^D(\lambda^+ - \lambda^-) + T^- - D^T LP^0 \ge 0. \tag{6.25}$$

In complete matrix notation we have a multi-parametric LCP of the form

$$0 \le y \perp +My + q + Nx \ge 0,$$

with $x = f$, $y = (g, \gamma, \lambda^+, \lambda^-)$,

$$N := \begin{bmatrix} -\hat{d}\tilde{E} \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$M := \begin{bmatrix} B & I_{nm\times nm} & (\tilde{D} - E_D) & -(\tilde{D} - E_D) \\ -I_{nm\times nm} & 0 & 0 & 0 \\ -(\tilde{D} - E_D)^T & 0 & L^D & -L^D \\ (\tilde{D} - E_D)^T & 0 & -L^D & L^D \end{bmatrix}, \quad \text{and}$$

$$q := \begin{bmatrix} -E_P + c \\ CAP \\ T^+ + D^T LP^0 \\ T^- - D^T LP^0 \end{bmatrix}.$$

Mathematical properties of the spot-market LCP

We will now summarize some of the mathematical properties of the spot market LCP.

Proposition 6.1.2. *The matrix B is positive semidefinite.*

Proof: Let $B_1 = e_{mn\times mn} e^T_{mn\times mn}$, $B_2 = \text{diag}(e_n e_n^T, e_n e_n^T, \ldots, e_n e_n^T)$, and $B_3 = \text{diag}(b)$.

Since $B = \hat{d}B_1 + B_2 + B_3$ and $\hat{d} > 0$, the assertion follows. □

As a consequence of Proposition 6.1.2, the coefficient matrix M is also positive semidefinite.

Proposition 6.1.3. *The matrix M is positive semidefinite.*

Proof: The coefficient matrix M is the sum of the positive semidefinite matrix

$$M^1 := \begin{bmatrix} B & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and the skew-symmetric matrix

$$M^2 := \begin{bmatrix} 0 & I_{nm \times nm} & (\tilde{D} - E_D) & -(\tilde{D} - E_D) \\ -I_{nm \times nm} & 0 & 0 & 0 \\ -(\tilde{D} - E_D)^T & 0 & L^D & -L^D \\ (\tilde{D} - E_D)^T & 0 & -L^D & L^D \end{bmatrix}.$$

It is well known and easy to see that the sum of a positive semidefinite matrix and a skew-symmetric matrix is positive semidefinite. □

The spot market LCP is therefore a monotone problem. This is a favorable property for the convergence of the algorithms described in Section 6.3.

6.2 Firm's forward-market problem

In this section, we will analyze the optimization problem of firm ℓ in the forward market. We will assume that each firm predicts the outcome of the spot market equilibrium and considers this in its decision in the forward market. Therefore, the decision variables in the forward market are the forward sales $f_{i\ell}$ of firm ℓ taking forward sales f_{is} of the other firms $s \in L \backslash \{\ell\}$ as given. The other variables are resulting from the equilibrium depending on the forward sales.

Forward-market payoff function

We assume that each firm wants to maximize its profit resulting from forward ($f_{i\ell}$) and spot-market ($g_{i\ell} - f_{i\ell}$) sales. In addition to the spot-market hub price $\hat{p}_h$ and transmission prices w, let p_h^f and w^f denote the hub and transmission prices in the forward market. A possible way to express the objective function π_ℓ in the forward market is then

$$\pi_\ell = \sum_{i \in N} p_h^f \, f_{i\ell} + \sum_{i \in N} \hat{p}_h \, (g_{i\ell} - f_{i\ell}) - \sum_{i \in N} \left[c_{i\ell} + \frac{1}{2} b_{i\ell} \, g_{i\ell} + w_i \right] g_{i\ell}.$$

Furthermore, if we assume that the forward price is a perfect forecast of the (deterministic) spot price, since otherwise arbitrage would take place, then the prices in the forward and spot market are the same. In this case, we can eliminate the forward sales f and can express the profit in terms of real-time generation variables only. The profit of firm ℓ is therefore:

$$\pi_\ell = \sum_{i \in N} \hat{p}_h \, g_{i\ell} - \sum_{i \in N} \left[c_{i\ell} + \frac{1}{2} b_{i\ell} \, g_{i\ell} + w_i \right] g_{i\ell}. \tag{6.26}$$

Notice that forward sales f do not appear in the objective function any more. This has a consequence for the spot-market equilibrium, since there, only aggregated forward sales are important. It seems therefore, we should only consider total forward sales by each firm; the locational aspect of forward sales is not important. To avoid introducing extra notation, we will use f_ℓ as the total amount of forward sales of firm ℓ from this point on. Since f_s, $s \neq \ell$, is given in firm ℓ's problem, we will denote these parameters as $\bar{f}_s$. Recall from (6.13) that $\hat{p}_h$ was given by:

$$\hat{p}_h = -\hat{d}G + \sum_{j \in N} d_j w_j + \sum_{j \in N} d_j P_j^0,$$

and w_i from (6.20) as

$$w_i = \sum_{k \in A} D_{ik} (\lambda_k^+ - \lambda_k^-).$$

Substituting this into (6.26) gives

$$\pi_\ell = \sum_{i\in N}\left[-\hat{d}G + \sum_{j\in N} d_j \sum_{k\in A} D_{jk}(\lambda_k^+ - \lambda_k^-) + \sum_{j\in N} d_j P_j^0\right] g_{i\ell}$$
$$- \sum_{i\in N}\left[c_{i\ell} + \frac{1}{2} b_{i\ell}\, g_{i\ell} + \sum_{k\in A} D_{ik}(\lambda_k^+ - \lambda_k^-)\right] g_{i\ell}. \qquad (6.27)$$

Note that in the current form, the objective function (6.27) is neither convex nor concave due to the bilinear terms $\lambda_k^+ g_{if}$ and $\lambda_k^- g_{if}$. Fortunately, it is possible to use the constraints of the spot-market equilibrium LCP to reformulate the objective function as a concave function in the decision variables.

Proposition 6.2.1. *The objective function π^ℓ for firm ℓ in the forward market can be expressed as a concave function of the spot market variables $y = (g, \gamma, \lambda^+, \lambda^-)$.*

Proof: We will again use the vector/matrix notation introduced before. The objective takes the following form:

$$\begin{aligned}\pi_\ell =& - \hat{d} g_\ell^T e_n G + g_\ell^T (e_n d^T D)(\lambda^+ - \lambda^-) + g_\ell^T e_n d^T P^0 \\ & - c_\ell^T g_\ell - \frac{1}{2} g_\ell^T B_\ell^m g_\ell - g_\ell^T D(\lambda^+ - \lambda^-) \\ =& - (\lambda^+ - \lambda^-)^T [D - e_n d^T D] g_\ell - \hat{d} g_\ell^T e_n G + g_\ell^T e_n d^T P^0 - c_\ell^T g_\ell - \frac{1}{2} g_\ell^T B_\ell^m g_\ell,\end{aligned}$$

where $B_\ell^m := \mathrm{diag}(b_\ell)$. The m in B_ℓ^m is used to denote marginal cost. Using (6.24) and (6.25) for firm ℓ, we can derive the following identity:

$$\begin{aligned}-(\lambda^+ - \lambda^-)^T [D - e_n d^T D] g_\ell =& \sum_{s\in F\setminus\{\ell\}} (\lambda^+ - \lambda^-)^T [D - e_n d^T D] g_s \\ & - (\lambda^+ - \lambda^-)^T L^D (\lambda^+ - \lambda^-) \\ & - (T^-)^T \lambda^- - (T^+)^T \lambda^+ - (\lambda^+ - \lambda^-)^T D^T L P^0.\end{aligned}$$

Next, we use (6.22) for firms $s \neq \ell$ to get

$$\sum_{s\in F\setminus\{\ell\}} (\lambda^+ - \lambda^-)^T[D - e_n d^T D]g_s = \sum_{s\in F\setminus\{\ell\}} \Big[-\hat{d}g_s^T E_n g_s + \hat{d}g_s^T e_n \bar{f}_s - \hat{d}g_s^T e_n G + g_s^T e_n d^T P^0 - c_s^T g_s - g_s^T B_s^m g_s - \gamma_s^T g_s \Big],$$

where $\bar{f}_s$ is used as a reminder that f_s is fixed in producer ℓ's problem. Further using (6.22) and putting everything together yields:

$$\begin{aligned}
\pi_\ell =& -(\lambda^+ - \lambda^-)^T[D - e_n d^T D]g_\ell - \hat{d}g_\ell^T e_n G + g_\ell^T e_n d^T P^0 - c_\ell^T g_\ell - \tfrac{1}{2} g_\ell^T B_\ell^m g_\ell \\
=& \sum_{s\in F\setminus\{\ell\}} \Big[-\hat{d}g_s^T E_n g_s + \hat{d}g_s^T e_n \bar{f}_s - \hat{d}g_s^T e_n G + g_s^T e_n d^T P^0 \\
& \quad - c_s^T g_s - g_s^T B_s^m g_s - \gamma_s^T CAP_s \Big] \\
& - \hat{d}g_\ell^T e_n G + g_\ell^T e_n d^T P^0 - c_\ell^T g_\ell - \frac{1}{2} g_\ell^T B_\ell^m g_\ell \\
& - (\lambda^+ - \lambda^-)^T L^D (\lambda^+ - \lambda^-) - (T^-)^T \lambda^- - (T^+)^T \lambda^+ - (\lambda^+ - \lambda^-)^T D^T L^D P^0 \\
=& -\frac{1}{2} g^T B_{MPEC}\, g - c^T g + E_P - \sum_{s\in F\setminus\{\ell\}} \Big[-\hat{d}g_s^T e_n \bar{f}_s + \gamma_s^T CAP_s \Big] \\
& - (\lambda^+ - \lambda^-)^T L^D (\lambda^+ - \lambda^-) - (T^-)^T \lambda^- - (T^+)^T \lambda^+ - (\lambda^- - \lambda^+)^T D^T L P^0,
\end{aligned}$$

where we use B_{MPEC} to denote the matrix

$$B_{MPEC}(i\ell, js) := \begin{cases} 2\hat{d} + b_\ell & \text{if } s = \ell, \\ 4\hat{d} + 2b_s & \text{otherwise.} \end{cases}$$

We can write the objective as

$$\pi_\ell(x, y) = -\frac{1}{2}[x, y] H \begin{bmatrix} x \\ y \end{bmatrix} - c^T x - d^T y, \tag{6.28}$$

where $x = f_\ell$, $y = (g, \gamma, \lambda^+, \lambda^-)$, H is the Hessian of the firm ℓ's payoff function,

$c := 0$, and

$$d = \begin{bmatrix} d^g \\ d^\gamma \\ d^{\lambda^+} \\ d^{\lambda^-} \end{bmatrix},$$

where

$$d^g_s := \begin{cases} -e_n d^T P^0 + c_s & \text{if } s = \ell, \\ -e_n d^T P^0 + c_s - \hat{d} e_n \bar{f}_s & \text{otherwise,} \end{cases}$$

$$d^\gamma_s := \begin{cases} 0 & \text{if } s = \ell, \\ CAP_s & \text{otherwise,} \end{cases}$$

and

$$\begin{bmatrix} d^{\lambda^+}_s \\ d^{\lambda^-}_s \end{bmatrix} := \begin{bmatrix} T^+ + D^T L P^0 \\ T^- - D^T L P^0 \end{bmatrix}.$$

The Hessian of the reformulated payoff function π_ℓ takes the following form:

$$H := \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & B_{MPEC} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & L^D & -L^D \\ 0 & 0 & 0 & -L^D & L^D \end{bmatrix}.$$

The key observation is that B_{MPEC} is the sum of three positive semidefinite matrices,

$$B_1 := 2\hat{d} E_{nm},$$

$$B_2 := 2\hat{d} \begin{bmatrix} 0 & 0 & 0 \\ 0 & E_s & 0 \\ 0 & 0 & E_s \end{bmatrix}, \qquad \text{for } s \neq \ell,$$

and B_3 is a diagonal matrix with

$$B_3(is, is) := \begin{cases} b_{is} & \text{if } s = \ell, \\ 2b_{is} & \text{if } s \neq \ell. \end{cases}$$

This makes B_{MPEC} positive semidefinite (remember that $\hat{d} > 0$). We already argued earlier that L^D is positive semidefinite. Again, the sum of a positive semidefinite and a skew-symmetric matrix is positive semidefinite, so that we have shown that the objective function can be reformulated as a concave function in the model variables. □

Complete forward-market problem for firm ℓ

The overall problem for a single firm ℓ, given forward sales $\bar{f}_s$ for $s \neq \ell$, is the following MPEC:

$$\begin{aligned} \min \quad & \frac{1}{2}[x, y] H \begin{bmatrix} x \\ y \end{bmatrix} + c^T x + d^T y \\ \text{subject to} \quad & 0 \leq y \perp Nx + My + q \geq 0, \end{aligned} \tag{6.29}$$

where $x = f_\ell$, $y = (g, \gamma, \lambda^+, \lambda^-)$, H, c, and d have been defined above. The matrices N, M and the vector q have following form:

$$N := \begin{bmatrix} -\hat{d}\tilde{E}_\ell \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$M := \begin{bmatrix} B & I_{nm \times nm} & \tilde{D} - E_D & -\tilde{D} + E_D \\ -I_{nm \times nm} & 0 & 0 & 0 \\ -(\tilde{D} - E_D)^T & 0 & L^D & -L^D \\ (\tilde{D} - E_D)^T & 0 & -L^D & L^D \end{bmatrix},$$

and

$$q := \begin{bmatrix} q^g \\ q^\gamma \\ q^{\lambda^+} \\ q^{\lambda^-} \end{bmatrix},$$

where

$$q_s^g := \begin{cases} -e_n d^T P^0 + c_s & \text{if } s = \ell, \\ -e_n d^T P^0 + c_s - \hat{d} e_n \bar{f}_s & \text{otherwise,} \end{cases}$$

$$q_s^\gamma := \begin{cases} 0 & \text{if } s = \ell, \\ CAP_s & \text{otherwise,} \end{cases}$$

and

$$\begin{bmatrix} q_s^{\lambda^+} \\ q_s^{\lambda^-} \end{bmatrix} := \begin{bmatrix} T^+ + D^T LP^0 \\ T^- - D^T LP^0 \end{bmatrix}.$$

6.3 Algorithms for forward-market equilibria

In the same way as we did in Sections 5.1 and 5.2, we will again consider first the case where only one firm is allowed to sell in the forward market. In the framework of the last section, this amounts to setting $\bar{f}_s = 0$ for all $s \neq \ell$ and solving the resulting MPEC (6.29) for firm ℓ.

The single-firm algorithm

To solve the single-firm problem, we will use ECOPT. Recall that ECOPT solves problems of the general form:

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{subject to} \quad & g(x,y) \geq 0, \\ & y \geq 0, \;\; F(x,y) \geq 0, \;\; y^T F(x,y) = 0. \end{aligned}$$

In the context of our model, $x = f_\ell$, y are the real-time equilibrium variables, g is empty and $F(x,y) = Nx + My + q$. Through the reformulations in the last section, the objective function is concave and equilibrium constraint matrix M is positive semidefinite. Even with these nice properties, the problem is possibly nonconcave and hard to solve.

To take this into account, we use a heuristic and start with different starting points to avoid possible local maxima. A general procedure for solving the single-firm problem for firm ℓ is the following:

Initialization

(1) Generate four different random values of forward sales f_ℓ for firm ℓ, and set $\bar{f}_s = 0$ for all $s \neq \ell$. Solve the spot-market equilibrium using an LCP solver for each value of forward sales.

(2) For each equilibrium y found in Step (1), determine the value of the objective function $\pi_\ell(f_\ell, y)$ in (6.28). Let (f_ℓ, y) be the solution giving the largest value of π_ℓ.

General Step

(2) Use (f_ℓ, y) as a starting point to solve the single-firm problem for firm $\ell \in L$ with $\bar{f}_s = 0$ using ECOPT.

The multi-firm algorithm

If more than one firm is participating in the forward market, the problem will be a Nash game with multiple players. Each player has perfect foresight about the spot-market equilibrium and takes other firms decision in the forward market as given (Cournot assumption). In this game, each firm is solving an MPEC instead of a regular optimization problem.

An equilibrium for the multi-firm problem is a set of forward sales such that no firm can increase profits by changing its forward sales unilaterally. Another way of stating this is:

$(f_1, f_2, \ldots, f_m)$ is a multi-firm equilibrium if

$$f_\ell \in SOL(f_\ell), \quad \text{for } \ell = 1, \ldots, m,$$

where $SOL(f_\ell)$ is the amount of forward sales that are optimal for firm ℓ.

In general, one can not expect to find pure strategy equilibria in the forward market. We have encountered examples where in a six-node network with two producers, there exists the possibility of cycling between different strategy pairs. In particular, if the second firm does not sell in the forward market, the first may do so profitably. This will, in turn, force the second firm to sell forward which in the end leads the first firm to go back to zero forward sales. But for this case, the second firm in turn prefers to also not sell forward which brings the cycle back to the start. This behavior indicates the existence of mixed strategy equilibria and ways to detect these by our algorithm. Although these mixed equilibria are possible, in many cases pure strategy equilibria exist and are identified by our algorithm.

In the following, we summarize the multi-firm algorithm. We use a diagonalization algorithm similar to [HMP00] and [CHH97]:

Initialization

(1) Solve the single-firm problem for each firm $\ell \in L$ with $\bar{f}_s = 0$ for all $s \neq \ell$ using ECOPT. To save computation, we set ECOPT to a loose optimality tolerance, e.g., $tol = 1.e-1$. Update the set of forward sales $f = (f_1, \ldots, f_m)$ with the found solutions.

General Step

(2) Use (f_ℓ, y), where y is the equilibrium vector from the last solved problem, as a starting point to solve the single-firm problem for firm $\ell \in L$ with $\bar{f}_s = f_s$ for all $s \neq \ell$ using ECOPT to find a new f_ℓ. Update firm ℓ's forward sales as f_ℓ. Repeat for $\ell = 1, \ldots, m$.

(2) Repeat step (2), increasing the optimality tolerance appropriately, until the maximum number of iterations is reached or a satisfactory solution is found.

In our practical tests, we run the algorithm until a satisfactory tolerance level is reached for ECOPT, and forward sales and profits do not differ in 5 significant digits in consecutive iterations in the General Step.

6.4 Six-node example

We will now present an example for the general model from Section 6 and apply the algorithms just presented to find equilibria in the two-stage game.

Consider the network presented in Figure 6.1. There are two firms owning generation facilities at nodes 1,2 and 4, and three demand points at nodes 3,5 and 6. We assume constant marginal cost to make the interpretation of results easier. Increasing marginal cost as described in the general model would be easily handled. The demand and cost data is summarized next to the nodes. Notice that the transmission line from node 1 to 6 has limited capacity and is possibly congested, limiting transmission from the northern zone of the network to the south. All other lines are assumed to have enough capacity so that they will not be congested.

We analyze the market for three different values of transmission capacity. First, we consider a medium capacity of 200 MW on the line from node 1 to 6. This will serve as our base case. A medium capacity will be enough to make it profitable for firms to sell forward. The transmission line will be congested in equilibrium and we will see the prisoner's dilemma derived for the two-node example we considered in Section 5. After that, we will contrast the situation with a low capacity of 100 MW and the case where the transmission capacity is unconstrained between node 1 and 6.

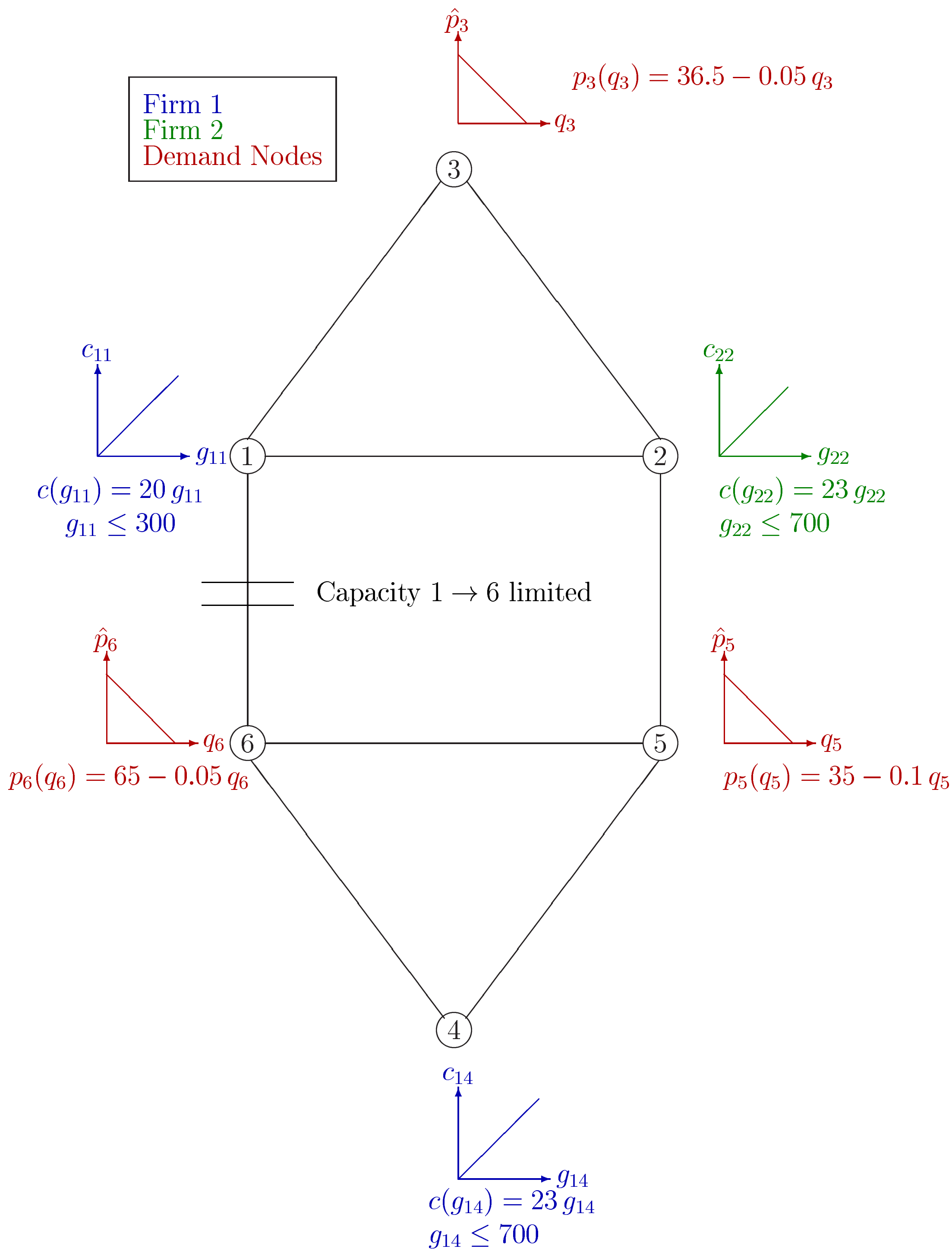

Figure 6.1: Six-node network with two firms and three demand nodes

Iterations for the Multi-Firm Algorithm				
Loop	Forward Sales Firm 1	π_1	Forward Sales Firm 2	π_2
1	349.167	4876.69	261.666	2738.78
2	283.750	3220.57	190.729	1455.10
3	301.484	3635.71	186.296	1388.24
4	302.593	3662.50	186.018	1384.11
5	302.662	3664.17	186.001	1383.86
6	302.666	3664.28	186.000	1383.84

Table 6.1: Results for the multi-firm algorithm on the six-node example

6.4.1 Medium transmission capacity

In the following, we will contrast four different cases. First, we find the Cournot equilibrium in the spot market if no forward sales are contracted by either of the firms. This involves solving a linear complementarity problem for which we use the algorithm presented in [KP99], although any algorithm for solving LCP's would be able to find a solution to this problem efficiently. Next, we solve the single-firm problems for each of the two firms using the single-firm algorithm. To find the multi-firm equilibrium in the forward market, we use the multi-firm algorithm.

The results for medium transmission capacity are summarized in Tables 6.1–6.5. Table 6.1 shows the iterations of the multi-firm algorithm. It is easy to see that the algorithm converges to an equilibrium (which need not be unique). Note that although we solve the single-firm problems in iteration 1, we use a loose accuracy, so that the results do not coincide with the single-firm outcomes solved to full accuracy (see Tables 6.4 and 6.5).

We summarize some observations for the four different cases. First, looking at Table 6.2, we notice that, generally, prices are declining if players participate in the forward market, while transmission prices (Table 6.3) stay constant in the case where only one line is congested. The decrease in prices is, of course, caused by an increase in generation (see Table 6.4) when players start selling in the forward market. In contrast to the Cournot outcome without forward sales, each firm can increase production by

Cases	Delivered Prices $/MWh					
	p_1	p_2	p_3	p_4	p_5	p_6
Cournot Competition	29.3	30.0	29.6	32.3	32.0	32.6
Only Firm 1 Forward	27.0	27.7	27.3	30.0	29.7	30.3
Only Firm 2 Forward	27.6	28.2	27.9	30.6	30.2	30.9
Both Firms Forward	26.1	26.7	26.4	29.1	28.7	29.4

Table 6.2: Prices for medium capacity

Cases	Transmission Price to Hub ($/MWh)					Power Transmitted	Interface Dual λ_{3+} $/MWh
	W_1	W_2	W_3	W_4	W_5	t_{16}	
Cournot Competition	3.3	2.7	3.0	0.3	0.7	200	5.3
Only Firm 1 Forward	3.3	2.7	3.0	0.3	0.7	200	5.3
Only Firm 2 Forward	3.3	2.7	3.0	0.3	0.7	200	5.3
Both Firms Forward	3.3	2.7	3.0	0.3	0.7	200	5.3

Table 6.3: Transmission prices for medium capacity

Cases	Forward Sales by each Firm		Quantities Demanded			Generation by Firm/Node		
	f_1	f_2	q_3	q_5	q_6	g_{11}	g_{14}	g_{22}
Cournot Competition	0	0	137	30	647	138	328	349
Only Firm 1 Forward	349	0	184	53	694	267	431	233
Only Firm 2 Forward	0	262	172	48	682	31	347	523
Both Firms Forward	303	186	202	63	712	174	431	372

Table 6.4: Quantities for medium capacity

being the only player in the forward market.

Total production is highest when both firms participate in the forward market and each generator produces more electricity than in a Cournot market with no forward sales. Notice that the transmission line is congested in all cases in contrast to the simple model in Section 5.2, where no firm sold forward if the transmission line was congested in the Cournot equilibrium with no forward sales.

Higher output leads to higher consumer surplus, as can be seen in Table 6.5. But similar to the observation we made for the simple two node example studied earlier, the situation is again a prisoner's dilemma where both firms benefit ($\approx +12.5\%$) if they are the only players in the forward market, but their profits decrease by a higher amount ($\approx -15\%$ for firm 1 and -43% for firm 2) if they both compete in the forward market. Transmission owner profits are constant in all four cases for the medium capacity of 200 MW on the line from node 1 to 6. The overall social benefit, the sum of firms' profits, grid owner's revenue and consumer surplus, increase if firms sell in the forward market.

6.4.2 Low transmission capacity

In the second scenario we discuss, the transmission capacity on the line from node 1 to node 6 is only 100 MW. In the simple two-node network discussed in Section 5,

Cases	Profits $/MWh		Grid Owner Revenue $/hr	Consumer Surplus $/hr	Net Social Welfare $/hr
	$\ell = 1$	$\ell = 2$			
Cournot Competition	4335	2434	1067	10984	18820
Only Firm 1 Forward	4877	1081	1067	13016	20040
Only Firm 2 Forward	2863	2739	1067	12481	19149
Both Firms Forward	3664	1384	1067	13903	20018

Table 6.5: Profits and benefits for medium capacity

such a low capacity forced the firms to refuse to sell forward and only trade in the spot market. Although we did not force forward sales to be nonnegative, under the simple assumptions there, no negative forward sales were encountered even for very low transmission capacities. This situation changes under the assumptions incorporated into the general model.

Negative forward sales allowed

With low transmission capacity, if firms have the opportunity to buy (= negative forward sales) electricity in the forward market, they can do so profitable under the current assumptions. It is not clear which counter party will enter into the trade, but a similar argument as before could bring hedgers into the market to take this counter position in the forward market. We will see that this behavior is not beneficial for the overall market.

For completeness, we present the same model output we showed for the medium capacity case in Tables 6.6 through 6.9. As mentioned above, if firms are allowed to buy electricity in the forward market, they will choose do so, see Table 6.8.

In contrast to higher transmission capacity, in this case actually both players benefit from such a behavior. Profits, see Table 6.9, increase for both firms if only one firm buys in the forward market. This gives a strong incentives to firms to collude.

Cases	Delivered Prices $/MWh					
	p_1	p_2	p_3	p_4	p_5	p_6
Cournot Competition	28.3	29.3	28.8	32.9	32.4	33.4
Only Firm 1 Forward	27.9	29.5	28.7	35.3	34.5	36.2
Only Firm 2 Forward	30.0	30.6	30.3	33.0	32.6	33.3
Both Firms Forward	31.5	32.4	32.0	35.6	35.1	36.0

Table 6.6: Prices for low capacity

Cases	Transmission Price to Hub ($/MWh)					Power Transmitted	Interface Dual λ_{3+} $/MWh
	W_1	W_2	W_3	W_4	W_5	t_{16}	
Cournot Competition	5.1	4.1	4.6	0.5	1.0	100	8.1
Only Firm 1 Forward	8.3	6.6	7.5	0.8	1.7	100	13.3
Only Firm 2 Forward	3.3	2.7	3.0	0.3	0.7	100	5.3
Both Firms Forward	4.5	3.6	4.0	0.4	0.9	100	7.1

Table 6.7: Transmission prices for low capacity

Cases	Forward Sales by each Firm		Quantities Demanded			Generation by Firm/Node		
	f_1	f_2	q_3	q_5	q_6	g_{11}	g_{14}	g_{22}
Cournot Competition	0	0	153	26	632	0	494	317
Only Firm 1 Forward	−203	0	156	5	577	0	413	326
Only Firm 2 Forward	0	−99	124	24	634	0	498	283
Both Firms Forward	−207	−222	90	−1	580	0	421	249

Table 6.8: Quantities for low capacity

Cases	Profits $/MWh		Grid Owner Revenue $/hr	Consumer Surplus $/hr	Net Social Welfare $/hr
	$\ell = 1$	$\ell = 2$			
Cournot Competition	4889	2012	810	10611	18321
Only Firm 1 Forward	5087	2120	1328	8935	17470
Only Firm 2 Forward	4970	2161	533	10459	18123
Both Firms Forward	5278	2346	714	8616	16954

Table 6.9: Profits and benefits for low capacity

If they both buy forward, they can increase their profits even more. Buying and selling in the forward market have opposite effects on prices. We can see in Table 6.6 that prices increase; this causes consumer surplus and total social benefit to decrease (Table 6.9). The effect on transmission prices, see Table 6.7, depends on the location of the generators and the firms acting in the forward market. Transmission prices tend to go down with lower overall generation. It is interesting to see that firm 1 is not using its cheaper generation facility at node 1 (see Table 6.8). The reason for this is the high transmission costs from node 1 compared to node 4 which offset the lower cost in all four cases, see Tables 6.6 and 6.7.

No negative forward sales allowed

We have seen that allowing firms to buy in the forward market could have negative effects on overall benefits. The fear of such an effect may have been one the reasons why generators in the initial California market design where not allowed to participate in the forward market at all; generators were forced to sell into the spot market. If generation firms could successfully buy instead of sell in the forward market and manipulate the overall market in the way we have just seen, this is one obvious, but certainly not the best way, to deal with this problem. With higher transmission capacity, forward trading is, as we have seen, beneficial to all market participants, and should not be forbidden.

Assuming that an effective control of generators could be implemented, it seems the best way is to force generators to only sell, but not buy in the forward market. This allows for the positive effect of positive forward sales but prevents the negative effect of negative forward sales.

To simulate such a rule, we ran the same model of low transmission capacity with a nonnegative bound on the forward sales quantity. As expected, in this case firms do not sell forward any electricity. Similar to the small example in Section 5, the Cournot outcome (the first case in Tables 6.6 through 6.9) prevails in all situations. As we have seen, firms have an incentive to manipulate the market, so firms would have to be monitored in cases of very limited transmission capacity.

Cases	Delivered Prices $/MWh					
	p_1	p_2	p_3	p_4	p_5	p_6
Cournot Competition	31.2	31.2	31.2	31.2	31.2	31.2
Only Firm 1 Forward	29.1	29.1	29.1	29.1	29.1	29.1
Only Firm 2 Forward	29.2	29.2	29.2	29.2	29.2	29.2
Both Firms Forward	27.9	27.9	27.9	27.9	27.9	27.9

Table 6.10: Prices for unlimited capacity

6.4.3 Unlimited transmission capacity

The results for the six-node network with unlimited transmission capacity are summarized in Tables 6.10 through 6.13. In essence, the results are the same as for the small example in Section 5.

If firms have the chance to sell forward, they can do so profitably, see Table 6.13. We again observe the prisoners dilemma; both firms are worse off if they both go into the forward market. From our experience with different demand, generation and transmission data on the six-node network, this will generally be true, although in more complicated cases the strategic interaction between firms gets more difficult to analyze. Consumers' surplus and social benefit increase through positive forward sales. Prices (Table 6.10) decrease, while the marginal transmission prices for unlimited transmission capacity (Table 6.11) are obviously zero in all cases.

6.5 Transmission expansion analysis

The last section showed the application of the general model to the six-node network. The inclusion of the physical laws governing electricity transmission on realistic networks allow the study of complicated relationships between the forward and spot

Cases	Transmission Price to Hub ($/MWh)					Power Transmitted	Interface Dual λ_{3+} $/MWh
	W_1	W_2	W_3	W_4	W_5	t_{16}	
Cournot Competition	0.0	0.0	0.0	0.0	0.0	335	0.0
Only Firm 1 Forward	0.0	0.0	0.0	0.0	0.0	271	0.0
Only Firm 2 Forward	0.0	0.0	0.0	0.0	0.0	405	0.0
Both Firms Forward	0.0	0.0	0.0	0.0	0.0	340	0.0

Table 6.11: Transmission prices for unlimited capacity

Cases	Forward Sales by each Firm		Quantities Demanded			Generation by Firm/Node		
	f_1	f_2	q_3	q_5	q_6	g_{11}	g_{14}	g_{22}
Cournot Competition	0	0	106	38	676	300	110	410
Only Firm 1 Forward	308	0	147	59	717	300	315	307
Only Firm 2 Forward	0	307	147	58	717	300	8	615
Both Firms Forward	246	246	172	71	742	300	192	492

Table 6.12: Quantities for unlimited capacity

Cases	Profits $/MWh		Grid Owner Revenue $/hr	Consumer Surplus $/hr	Net Social Welfare $/hr
	$\ell = 1$	$\ell = 2$			
Cournot Competition	4262	3362	0	11778	19402
Only Firm 1 Forward	4682	1891	0	13564	20137
Only Firm 2 Forward	2791	3782	0	13564	20137
Both Firms Forward	3321	2421	0	14736	20477

Table 6.13: Profits and benefits for unlimited capacity

market in deregulated electricity markets. The strategic interaction between firms can also be analyzed using the new model.

In this section, we will look at the results from a higher perspective and analyze more general implications. In particular, our results show an intrinsic relationship between transmission capacity and the development of forward markets in electricity. In the following, we will show how this can be used to accomplish a better understanding of transmission investments in deregulated electricity markets.

To do this, we summarize the main results in a slightly different form than before. Recall that we were looking at four different cases, depending on whether each firm was allowed to participate in the forward or not. This can be summarized in the form of a two-by-two table as follows:

In Figure 6.2, we summarize the strategic options given to each firm. Each firm can either not participate in the forward market (No) or do so (Yes). The four strategic outcomes, Cournot competition in the spot market, one firm only in the forward market and both firms in the forward market, are summarized in the appropriate boxes of the table.

Using this format, we can summarize the three transmission capacity scenarios of the last section. In particular, we focus on the overall social benefit as a function of the transmission capacity. In Figure 6.3, the social benefits for the three transmission

		Firm 2	
		No	Yes
Firm 1	No	Cournot Competition	Only Firm 2 Forward
	Yes	Only Firm 1 Forward	Both Firms Forward

Figure 6.2: Matrix representation of different strategic outcomes

capacity scenarios and all strategy pairs are rounded to the next hundred. The values in parenthesis in the table for low capacity of 100 MW correspond to negative forward sales. For this discussion, we will focus on the Cournot market outcome in the low capacity scenario and assume that negative forward sales are prohibited.

Transmission expansion from 100 MW to 200 MW

As an example of how our results can be used to analyze transmission investments in deregulated electricity markets, we consider the case where a decision maker is faced with the evaluation of an expansion from 100 MW to 200 MW on the line between nodes 1 and 6.

Using a Cournot model of competition, one would compare the social benefit in the two upper left corners of the two-by-two matrices. The result, see Figure 6.4, would be an increase of 500 in social benefit due to the increase of 100 MW in transmission capacity.

On the other hand, using our forward market model, one would compare the lower right corners of the matrices. This box corresponds to the outcome if both firms are allowed to sell/buy in the forward market. If firms are not allowed to buy in the

100 MW

18,300	(18,100)
(17,500)	(17,000)

200 MW

18,800	20,000
19,100	20,000

Unlimited

19,400	20,100
20,100	20,500

Figure 6.3: Social benefit as a function of transmission capacity

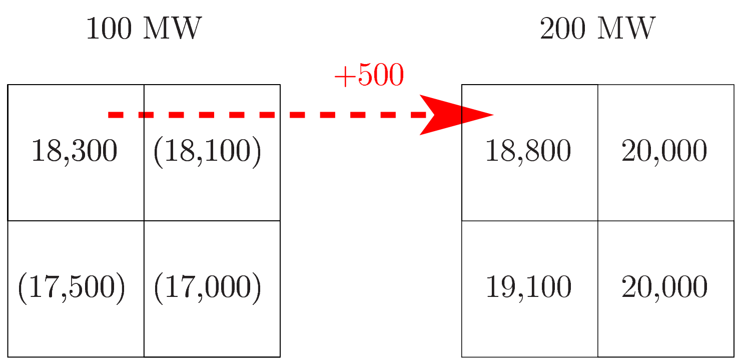

Figure 6.4: Transmission expansion: Cournot model analysis

forward, they will choose not to sell and the Cournot outcome (upper left corner) will prevail, as argued earlier. So, given our model and the assumption of no negative forward sales, the increase in social benefit can be expected to be 1,700, see Figure 6.5.

This is more than three times as high as predicted by a pure Cournot model analysis. If negative forward can't be avoided effectively, this difference could increase even more (to +3000), as a comparison of the lower right corners immediately reveals.

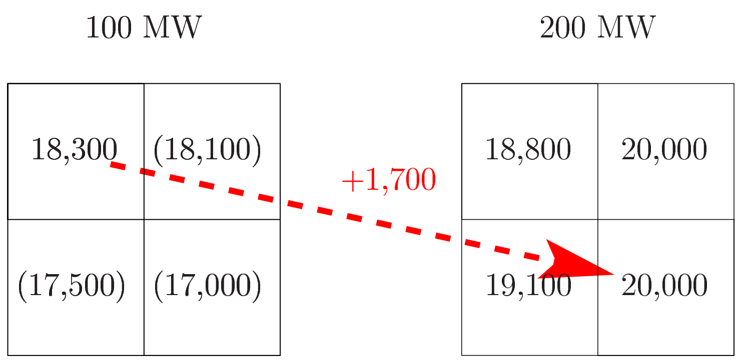

Figure 6.5: Transmission expansion: Forward model analysis

6.6 Summary

We have shown that even with transmission constraints, generation companies in a deregulated electricity environment have incentives to sell energy in the forward market with the assumption of certainty and perfect foresight. The analytical treatment of a small example with only one transmission line revealed the existence of threshold of transmission capacity under which no firm has an incentive to sell in the forward market. Over this threshold, forward sales take place but the transmission price stays at zero under the simple assumptions made.

We have also presented a general model to analyze larger electricity networks and demonstrated the use on a six-node example. Numerical results indicate that firms would decide to sell in the forward market and total social benefit increases in the process, even though the transmission price for the line is positive in some cases.

We showed that a closer look at the strategic interactions between firms is necessary to understand complex electricity markets. In particular, we discovered an intrinsic relationship between transmission capacity and forward markets in electricity. We also demonstrated the use of our model to provide a better understanding of the transmission investment problem in deregulated electricity markets.

Chapter 7

Conclusions and Future Research

7.1 MPEC Algorithms

In the first part of this dissertation, we presented a new algorithm for mathematical programs with equilibrium constraints. The discussion of the mathematical properties of the MPEC in Chapter 2 showed the difficulty of the problem and the need for specialized algorithms due to the failure of standard constraints qualifications at any feasible point. In Chapter 3, we first reviewed existing methods for MPECs and then described our new algorithm ECOPT. The idea and theory of ECOPT is closely related to the explicit smoothing algorithm proposed by Jiang and Ralph in [JR99]. In contrast to their work, we also implemented and tested our method on a wide range of test problems.

The implementation details of ECOPT, in particular an update strategy for the approximation of the Lagrangian Hessian using exact second derivatives of the smoothing function and a specialized termination criterion are given in the first part of Chapter 4. In the second part of that chapter, we compare the numerical performance of ECOPT to state of the art nonlinear optimization algorithms as well as some other algorithms specifically designed to solve MPEC problems. The test showed that ECOPT and the SQP algorithm NPSOL perform best on the problems tested while the sequential linearization algorithm MINOS cannot be recommended for the solution of MPECs. The other methods tested (PIPA, PSQP and SNOPT) are slightly less efficient than ECOPT and NPSOL and fail in some instances. Overall, we feel that the use of smoothing function improves efficiency and reliability of MPEC solvers and recommend ECOPT to solve small to medium scale MPEC problems. Although theoretically problematic, a general SQP algorithm like NPSOL can be used as an alternative with good results. For large-scale MPEC problems, we suggest using SNOPT, although we feel that the QP solver SQOPT used in SNOPT does not handle degenerate subproblems efficiently and could be improved.

Future research on ECOPT will extend the Lagrangian Hessian update to nonlinear constraint functions and the use of a sparse QP solver to enable the solution of large-scale MPEC problems. We also feel that QP solvers handling degenerate QP subproblems efficiently will improve the performance of standard nonlinear optimization algorithms like NPSOL and SNOPT.

7.2 Forward market model in electricity

In the second part of this dissertation, we introduced new models for electricity forward markets. The analytical discussion of a small-scale example in Chapter 5 revealed the intrinsic relationship between transmission capacity and the development of forward markets in electricity. A general model for forward markets in electricity is developed in Chapter 6.

The general model is flexible in the number of firms, nodes and transmission lines and includes transmission and generation capacity limits as well as an accurate network representation using Kirchhoff's laws. We also apply ECOPT to solve the resulting MPEC problems for a six-node example. The study of different transmission capacity scenarios extends the results found for the small-scale case and showed how our model can be used to gain insights into complicated strategic interactions between electricity producers and between forward and spot markets in electricity. A further application to transmission investment evaluations is suggested and demonstrated.

Possible extensions of the model could include the consideration of extra constraints such as flowgate constraints on the transmission lines, constraints on forward sales, different transmission pricing schemes or nonlinear functions. Also the expansion of generation capacity by the firms in the forward market could be modeled by including extra variables.

One particularly interesting extension is the inclusion of uncertainty. The analysis could give insight into whether producers benefit from uncertain demand in the current framework, and are therefore better off not to participate in forward markets for electricity despite the apparent advantages revealed by the results of our model. In this more general setting, it might also be possible to obtain uniqueness results about equilibria, which do not hold for our model.

Appendix A

Summary of Notation for the General Model

Parameters

n:	number of nodes, indexed by i, j (not including the hub)
m:	number of firms, indexed by ℓ, s
l:	number of bidirectional arcs
D_{ik}:	power distribution factor i on arc k
T_k^+:	transmission capacity on arc k
T_k^-:	transmission capacity in the reverse direction of arc k
$c_{i\ell}$:	marginal cost intercept at node i by firm ℓ
$b_{i\ell}$:	marginal cost slope at node i by firm ℓ
$CAP_{i\ell}$:	production capacity at node i for firm ℓ
P_i^0:	positive price intercept of demand curve at node i
Q_i^0:	positive quantity intercept of demand curve at node i
N:	set of nodes
A:	set of arcs in the full network
F:	set of firms

Variables for Hub Network

$s_{i\ell}$: amount of sales at node i by firm ℓ

$g_{i\ell}$: generation at node i by firm ℓ

t_i: amount of transmission service (MW) from i to the hub

w_i: wheeling cost from i to the hub

p_ℓ: price at the hub, determined by firm ℓ

$a_{i\ell}$: amount of arbitrage at i, determined by firm ℓ

$\hat{p}_i$: spot market price at node i

Vectors & Matrices

e_n: vector of n ones

E_n: $n \times n$ matrix of ones

s_ℓ: vector of length n of $s_{i\ell}$ for fixed ℓ

g_ℓ: vector of length n of $g_{i\ell}$ for fixed ℓ

$\hat{s}$: sum of s_t for $t \in F$

$\hat{g}$: sum of g_t for $t \in F$

P^0: vector of P_i^0

t: vector of t_i

CAP_ℓ: vector of length n of $CAP_{i\ell}$ for fixed ℓ

CAP: vector of $CAP_{i\ell}$ (length nm)

c: vector of $c_{i\ell}$

T^+: vector of flow capacity T_k^+

T^-: vector of flow capacity T_k^-

D: matrix of size $n \times l$ of PTDF values

Q: positive definite diagonal matrix of size $n \times n$, $Q_{ii} = \frac{P_i^0}{Q_i^0}$

Commonly Used Definition

S: defined as the quantity $\sum_{i\in N}\sum_{\ell\in F} s_{i\ell}$

G: defined as the quantity $\sum_{i\in N}\sum_{\ell\in F} g_{i\ell}$

$\hat{d}$: defined as: $\frac{1}{\sum_{k=1}^{n}\frac{Q_k^0}{P_k^0}}$

d_i: defined as: $\hat{d}\frac{Q_i^0}{P_i^0}$, for $i \in N$

L_{ii}: defined as: $\hat{d}\frac{Q_i^0}{P_i^0}\sum_{j\in N\setminus\{i\}}\frac{Q_j^0}{P_j^0}$, for $i \in N$

L_{ij}: defined as: $-\hat{d}\frac{Q_i^0 Q_j^0}{P_i^0 P_j^0}$, for $i, j \in N, i \neq j$

L: defined as the $n \times n$ matrix of L_{ij}

L^D: defined as the $l \times l$ matrix $D^T L D$

L^D_{kh}: defined as the $(k,h)^{th}$ entry of L^D, for $k, h \in A$

$\tilde{D}$: block-structured matrix of size $nm \times l$, column of m D's:

$$\tilde{D} := \begin{bmatrix} D \\ D \\ \vdots \\ D \end{bmatrix}.$$

E_D: the $n \times l$ matrix defined as $e_n d^T D$

E_P: the vector defined as $e_n d^T P^0$

$\tilde{E}$: block-structured matrix of size $nm \times nm$, E_n's on the diagonal:

$$\tilde{E} := \begin{bmatrix} E_n & 0 & \cdots & 0 \\ 0 & E_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_n \end{bmatrix}.$$

Bibliography

[Aba67] J. Abadie. On the Kuhn-Tucker theorem. In J. Abadie, editor, *Nonlinear Programming*, pages 19–36. North Holland Publishing Company, Amsterdam, 1967.

[All92] B. Allaz. Oligopoly, uncertainty and strategic forward transactions. *International Journal of Industrial Organization*, 10:297–308, 1992.

[AV93] B. Allaz and J.L. Vila. Cournot competition, forward markets and efficiency. *Journal of Economic Theory*, 59(1):1–16, 1993.

[BBS98] S. Borenstein, J. Bushnell, and S. Stoft. The competitive effects of transmission capacitiy in a deregulated electricity industry. Technical report, University of California Energy Institute, October 1998.

[Bil95] S.C. Billups. *Algorithms for Complementarity Problems and Generalized Equations*. PhD thesis, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1995.

[Bom99] Erik G. Boman. *Infeasibility and Negative Curvature in Optimization.* PhD thesis, Scientific Computing and Computational Mathematics, Stanford University, Stanford, California, 1999.

[CHH97] J. Cardell, C.C. Hitt, and W.W. Hogan. Market power and strategic interaction in electricity networks. *Resource and Energy Econonomics*, 19(1-2):109–137, 1997.

[Cla90] F. H. Clarke. *Optimization and nonsmooth analysis.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, second edition, 1990.

[CP96] H.P. Chao and S. Peck. A market mechanism for electric power transmission. *Journal of Regulatory Econonmics*, 10(1):25–59, 1996.

[CPS92] R.W. Cottle, J.S. Pang, and R.E. Stone. *The Linear Complementarity Problem.* Academic Press, Boston, 1992.

[DF99] S.P. Dirkse and M.C. Ferris. Modeling and solution environments for MPEC: GAMS & MATLAB. In *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997)*, pages 127–147. Kluwer Academic Publisher, Dordrecht, 1999.

[FER00] FERC. Market order proposing remedies for California wholesale electrics. Document No. EL00-95-000, November 2000. Available online at: http://www.ferc.fed.us/electric/bulkpower/EL00-95-000.pdf.

[Fis92] A. Fischer. A special Newton-type optimization method. *Optimization*, pages 269–284, 1992.

[FJQ99] F. Facchinei, H. Jiang, and L. Qi. A smoothing method for mathematical programs with equilibrium constraints. *Mathematical Programming*, (85):107–134, 1999.

[FKM99] M. C. Ferris, C. Kanzow, and T.S. Munson. Feasible descent algorithms for mixed complementarity problems. *Mathematical Programming*, 86(3):475–498, 1999.

[FLP98] M. Fukushima, Z.-Q. Luo, and J.-S. Pang. A globally convergent sequential programming algorithm for mathematical programs with linear complementarity constraints. *Computational Optimization and Application*, 10:5–30, 1998.

[FP66] M. Fiedler and V. Pták. Some generalizations of positive definiteness and monotonicity. *Numerische Mathematik*, 9:163–172, 1966.

[FP98] M. Fukushima and J.-S. Pang. Some feasibility issues in mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 8:673–681, 1998.

[FP99] M. Fukushima and J.-S. Pang. Convergence of a smoothing continuation method for mathematical programs with complementarity constraints. In M. Théra and R. Tichatschke, editors, *Ill-Posed Variational Problems and Regularization Techniques*, pages 99–110. Springer-Verlag, 1999.

[GMS97] P.E. Gill, W. Murray, and M.A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. Technical Report SOL 97-3, Department of Engineering-Economic Systems and Operations Research, Stanford University, Stanford, California, 1997.

[GMSW86] P.E. Gill, W. Murray, M.A. Saunders, and M.H. Wright. User's guide for NPSOL (Version 4.0): a Fortran package for nonlinear programming. Technical Report SOL 86-2, Department of Operations Research, Stanford University, Stanford, California, 1986.

[GMW81] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Academic Press, London and New York, 1981.

[Gol99] M. J. Goldsmith. *Sequential Quadratic Programming Methods Based on Indefinite Hessian Approximations*. PhD thesis, Department of Operations Research, Stanford University, Stanford, California, 1999.

[Gre96] R. Green. Increasing competition in the British electricity spot market. *Journal of Industrial Economics*, 44:205–216, 1996.

[HMP00] B.F. Hobbs, C.B. Metzler, and J.-S. Pang. Strategic gaming analysis for electric power systems: An MPEC approach. *IEEE Power Engineering Transactions 15*, 15:638–645, 2000.

[Hob00] B.F. Hobbs. Linear complementarity models of Nash-Cournot competition in bilateral and POOLCO power markets. Technical report, Department of Geography & Environmental Engineering, The Johns Hopkins University, Baltimore, Maryland, 2000.

[Hol99a] K. Holmström:. New optimization algorithms and software. *Theory of Stochastic Processes*, (1-2):55–63, 1999.

[Hol99b] K. Holmström:. The TOMLAB optimization environment in MATLAB. *Advanced Modeling and Optimization*, (1):47–69, 1999.

[JR99] H. Jiang and D. Ralph. QPECgen, a MATLAB generator for mathematical programs with quadratic objective and affine variational inequality constraints. *Computational Optimization and Application*, (13):25–59, 1999.

[JR00] H. Jiang and D. Ralph. Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. *SIAM Journal on Optimization*, 10(3):779–808, 2000.

[Kan96] C. Kanzow. Some noninterior continuation methods for linear complementarity problems. *SIAM Journal on Matrix Analysis and Applications*, (17):851–868, 1996.

[KP99] C. Kanzow and H. Pieper. Jacobian smoothing methods for nonlinear complementarity problems. *SIAM Journal on Optimization*, 9:342–372, 1999.

[KS86] M. Kojima and S. Shindo. Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. *Journal of Operations Research Society of Japan*, 29:352–374, 1986.

[LF00] S. Leyffer and R. Fletcher. Mathematical programs with equilibrium constraints: Much ado about nothing? Conference Presentation, SIAM Annual Meeting, Atlanta, Georgia, September 2000.

[LFK96] T. De Luca, F. Facchinei, and C. Kanzow. A semismooth equation approach to the solution of nonlinear complementarity problems. *Mathematical Programming*, (75):407–439, 1996.

[LFK00] T. De Luca, F. Facchinei, and C. Kanzow. A theoretical and numerical comparison of some semismooth algorithms for complementarity problems. *Computational Optimization an Applications*, 16:173–203, 2000.

[LPR96] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical programs with equilibrium constraints*. Cambridge University Press, Cambridge, 1996.

[LPR98] Z.Q. Luo, J.-S. Pang, and D. Ralph. Piecewise sequential quadratic programming for mathematical programs with nonlinear complementarity constraints. In *Multilevel optimization: algorithms and applications*, pages 209–229. Kluwer Academic Publisher, Dordrecht, 1998.

[Met00] C.B. Metzler. *Complementarity Models of Competitive Oligopolistic Electric Power Generation Markets*. PhD thesis, Department of Mathematical Sciences, Johns Hopkins University, 2000.

[MP99] W. Murray and F. J. Prieto. A second-derivative method for nonlinearly constrained optimization. Technical Report SOL 95-3, Department of Operations Research, Stanford University, Stanford, California, 1999.

[MR73] J.J. Moré and W.C. Rheinboldt. On **P**- and **S**-functions and related classes of nonlinear mappings. *Linear Algebra and Its Applications*, 6:45–63, 1973.

[MS95] B.A. Murtagh and M.A. Saunders. MINOS 5.4 User's Guide. Technical report, Department of Operations Research, Stanford University, Stanford, California, Revised 1995.

[Mur69] W. Murray. An algorithm for constrained minimization. In R. Fletcher, editor, *Optimization*. Academic Press, London, 1969.

[Mur97] W. Murray. Sequential quadratic programming methods for large-scale problems. *Computational Optimization and Applications*, 7:127–142, 1997.

[Nas51] J.F. Nash. Non-cooperative games. *Annals of Mathematics*, 54:286–295, 1951.

[OKZ98] J. Outrata, M. Kočvara, and J. Zowe. *Nonsmooth approach to optimization problems with equilibrium constraints.* Kluwer Academic Publishers, Dordrecht, 1998.

[OZ95] J. Outrata and J. Zowe. A numerical approach to optimization problems with variational inequality constraints. *Mathematical Programming*, (68):105–130, 1995.

[PHR91] J.S. Pang, S.P. Han, and N. Rangaraj. Minimization of locally Lipschitzian functions. *SIAM Journal on Optimization*, 1:57–82, 1991.

[Pie98] H. Pieper. Newton-type algorithms for mixed complementarity problems. Technical report, Department of Engineering-Economic Systems and Operations Research, Stanford University, Stanford, California, 1998.

[Pow83] M.J.D. Powell. Variable metric methods for constrained optimization. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical Programming: The State of the Art*, pages 288–311. Springer-Verlag, London, 1983.

[Ral96] D. Ralph. Sequential quadratic programming for mathematical programs with linear complementarity constraints. In R.L. May and A.K. Easten, editors, *Computational Techniques and Applications: CTAC95*, pages 663–668. World Scientific, 1996.

[Ras94] E. Rasmussen. *Games and Information: An Introduction to Game Theory.* Basil Blackwell, Cambridge, Massachusetts, 1994.

[Sch01] S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001.

[SCTB88] F.C. Schweppe, M.C. Caramanis, R.E. Tabors, and R.E. Bohn. *Spot Pricing in Electricity*. Kluwer, Norwell, Massachusetts, 1988.

[SS00] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25:1–22, 2000.

[Sta52] H. Van Stackelberg. *The Theory of Market Economy*. Oxford University Press, Oxford, 1952.

[Sto99a] S.E. Stoft. Financial transmission rights meet Cournot: How TCCs curb market power. *The Energy Journal*, 20(1):1–23, 1999.

[Stö99b] M. Stöhr. *Nonsmooth Trust Region Methods and Their Applications to Mathematical Programs with Equilibrium Constraints*. PhD thesis, University of Karslruhe, Germany, published by Shaker-Verlag, Aachen, Germany, 1999.

[VC94] L.N. Vincente and P.H. Calamai. Bilevel and multilevel programming: A bibliography review. *Journal of Global Optimization*, 5:291–306, 1994.

[Wil63] R.B. Wilson. *A Simplicial Method for Convex Programming*. PhD thesis, Harvard University, Cambridge, Massachusetts, 1963.