TWO DECOMPOSITION ALGORITHMS FOR NONCONVEX OPTIMIZATION PROBLEMS WITH GLOBAL VARIABLES

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

Angel-Victor de Miguel
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.

Richard W. Cottle

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.

Michael A. Saunders

Approved for the University Committee on Graduate Studies:

Abstract

A feature common to many optimization problems is a weak connectivity between component systems. Decomposition algorithms exploit this feature by breaking the problem into a set of smaller independent problems. One type of connectivity occurs when only a few of the variables, known as global variables, are relevant to all systems, while the remainder are local to a single component. We term these problems Optimization Problems with Global Variables. Examples arise in the design of complex systems such as an aircraft or automobile and in the solution of stochastic problems such as portfolio management.

Collaborative Optimization (CO) is a promising decomposition algorithm that transforms an Optimization Problem with Global Variables into an equivalent master problem and a set of subproblems. Unfortunately, both the CO master problem and the subproblems are degenerate. Nondegeneracy is a common assumption when proving convergence for most optimization algorithms. Not surprisingly, CO fails to solve some simple test problems.

We propose two novel decomposition algorithms that circumvent some of the difficulties associated with CO. The first algorithm, named Inexact Penalty Decomposition (IPD), uses an inexact penalty function. The second algorithm, termed Exact Penalty Decomposition (EPD), employs an exact penalty function and a barrier function. The main advantage is that these new approaches result in nondegenerate problems. Consequently, there exist algorithms that are fast locally convergent for both the master problem and the subproblems.

To test the new algorithms we present a new quadratic programming test-problem set. The user can choose problem size, convexity, degeneracy, and degree of coupling. All test-problem minimizers are known a priori. Both IPD and EPD successfully solve the test set for a wide variety of circumstances.

Acknowledgements

When I first met Professor Walter Murray during his short visit to Madrid in 1997, I was impressed by both his bright technical comments and his interesting, friendly conversation. Later, at Stanford, Professor Murray became not only my advisor, but also one of my best friends. Our long weekly meetings were lively discussions encompassing subjects as diverse as optimization, travel, and kids. He patiently taught me all I know about research, and much regarding other important aspects of life. His thought will have a long-lasting influence on me. For all this, I am deeply indebted to him.

Many thanks are due to Professor Richard W. Cottle. He greatly encouraged me to transfer from the Master's to the Ph.D. program, and therefore shares with Professor Murray the responsibility for me staying at Stanford. In his lectures, I learned the fundamentals of optimization. Finally, as a member of my reading committee, Dick helped to greatly improve the accuracy and style of this dissertation.

I would like to express my gratitude to Professor Michael A. Saunders. He patiently read this dissertation in great detail and suggested many changes that dramatically improved this document as well as my ability to write English. He was always glad to answer my many questions on the computational aspects of optimization and, on some occasions, his hints took my research out of a dead end.

I would like to thank all the faculty and staff of the Department of Management Science and Engineering for their help and advice over the years. Financial support for this research was provided by The La Caixa Fellowship Program, the National Science Foundation Grant CCR-9988205, and the Department of Navy Grant 00014-96-1-0274-P00009.

My office mates (Antonino del Gatto, Kien-Ming Ng, Heiko Pieper, Michael P. Friedlander, and Vinayak Vishnu Shanbhag) provided the most exciting working environment. Their influence is apparent in all parts of this dissertation. They made my time at the Systems Optimization Laboratory an enriching experience from both a professional and a personal perspective. In addition to my office mates, at Stanford I made many true friends from all continents—too many to name here! The diversity of their cultural backgrounds broadened my view of the world. There will always be a place for them in my heart.

My girlfriend Esperanza Amann is the best I take from Stanford. Her wisdom and patience have no limits. With her bright sense of humor, she often turns my worst worries into plain laughter. I hope our time at Stanford is just the beginning of the best love.

Last, but not least, this dissertation is dedicated to my parents, whose many sacrifices allowed me pursue an academic career.

A mis padres,
por todo.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Many optimization problems combine objective and constraint functions corresponding to a set of weakly connected systems. One class of connectivity occurs when only a few of the variables, known as global variables, appear in all systems, while the remainder occur in only a single component. These problems are known as Optimization Problems with Global Variables (OPGVs).

Decomposition algorithms break an OPGV into a set of smaller independent subproblems. The advantage is usually that the subproblems are easier to solve than the original problem. In this dissertation we propose two new decomposition algorithms for the nonconvex OPGV and investigate their analytical and numerical properties.

This chapter is organized as follows. In Section 1.1, we state the OPGV and the closely related Optimization Problem with Global Constraints (OPGC). In Section 1.2, we describe two problems from business and engineering that can be modeled as OPGVs. In Section 1.3, we discuss how decomposition algorithms work and review some of the most popular algorithms available. In Section 1.4, we describe the main contributions of this dissertation. Finally, Section 1.5 gives an overview of the remaining chapters.

1.1 Problem Statement

We distinguish two types of problems according to the modality of coupling among systems: (i) optimization problems with global variables and (ii) optimization problems with global constraints. Although this dissertation focuses on the OPGV, in this chapter, we discuss the characteristics of both types of problems.

1.1.1 Optimization Problems with Global Variables

In an OPGV, constraints are naturally classified as belonging to N different systems. Then some of the variables (known as global variables) are needed to evaluate all of the constraints, whereas the rest of variables (known as local variables) are needed only in the evaluation of the constraints belonging to one of the systems. Likewise, the objective function is the summation of N different terms, one per system. Again, while the global variables are needed in the evaluation of all of the terms, the local variables are needed only in the evaluation of one term. The OPGV may be stated as:

$$\begin{array}{lll} \min\limits_{x,y_i} & \sum\limits_{i=1}^{N} F_i(x,y_i) & \\ s.t. & c_i(x,y_i) \;\geq\; 0, & i = 1{:}N, \end{array} \tag{1.1}$$

where $x \in \mathbb{R}^n$ are the *global* variables, $y_i \in \mathbb{R}^{n_i}$ are the *local* variables, $F_i(x,y_i) : \mathbb{R}^{n+n_i} \to \mathbb{R}$ is the objective function term corresponding to the ith system, and $c_i(x,y_i) : \mathbb{R}^{n+n_i} \to \mathbb{R}^{m_i}$ are the constraints corresponding to the ith system.

If, in an OPGV, we set the global variables to a fixed value, the problem breaks into N independent subproblems. Decomposition algorithms use information gathered at the solution of these N independent subproblems to determine the optimal value of the global variables.

1.1.2 Optimization Problems with Global Constraints

In an OPGC, variables are naturally classified as belonging to N different systems. Then a few of the constraints (known as global constraints) depend on all of the variables, whereas the rest of the constraints (known as local constraints) only depend on variables corresponding to one of the systems. The objective function is a summation of N different terms, one per system. The OPGC may be stated as:

$$
\begin{aligned}
\min_{y_i} \quad & \sum_{i=1}^{N} F_i(y_i) \\
s.t. \quad b(y_1, \ldots, y_N) &\geq 0, \\
c_i(y_i) &\geq 0, \quad i = 1{:}N,
\end{aligned}
\tag{1.2}
$$

where $y_i \in \mathbb{R}^{n_i}$ are the variables corresponding to the ith system, $F_i(y_i) : \mathbb{R}^{n_i} \to \mathbb{R}$ is the objective function term corresponding to the ith system, $b(y_1, \ldots, y_N) : \mathbb{R}^{\sum_{i=1}^{N} n_i} \to \mathbb{R}^m$ are the global constraints, and $c_i(y_i) : \mathbb{R}^{n_i} \to \mathbb{R}^{m_i}$ are the local constraints.

If, in an OPGC, we ignore the global constraints the problem breaks into N independent subproblems. Decomposition algorithms use information gathered at the solution of these N independent subproblems to minimize the objective function while ensuring that the global constraints are satisfied.

1.2 Applications

1.2.1 Aircraft Wing Design

When designing an aircraft wing, we wish to minimize the drag, that is, air resistance to aircraft movement, subject to constraints corresponding to two different analysis disciplines: (i) aerodynamics and (ii) structures. The aerodynamic constraints are the discretization of the partial differential equations that describe the air flow around the wing. Likewise, the structure constraints are the discrete version of the partial differential equations that model the stress distribution in the wing interior.

In essence, the air flowing around the wing creates a pressure distribution that determines the load on the wing. This load causes a deflection on the wing that alters the pressure distribution. Overall, we want to minimize the drag while keeping the wing weight bounded and ensuring that the structure will resist the load imposed on it. This problem is depicted in Figure 1.1.

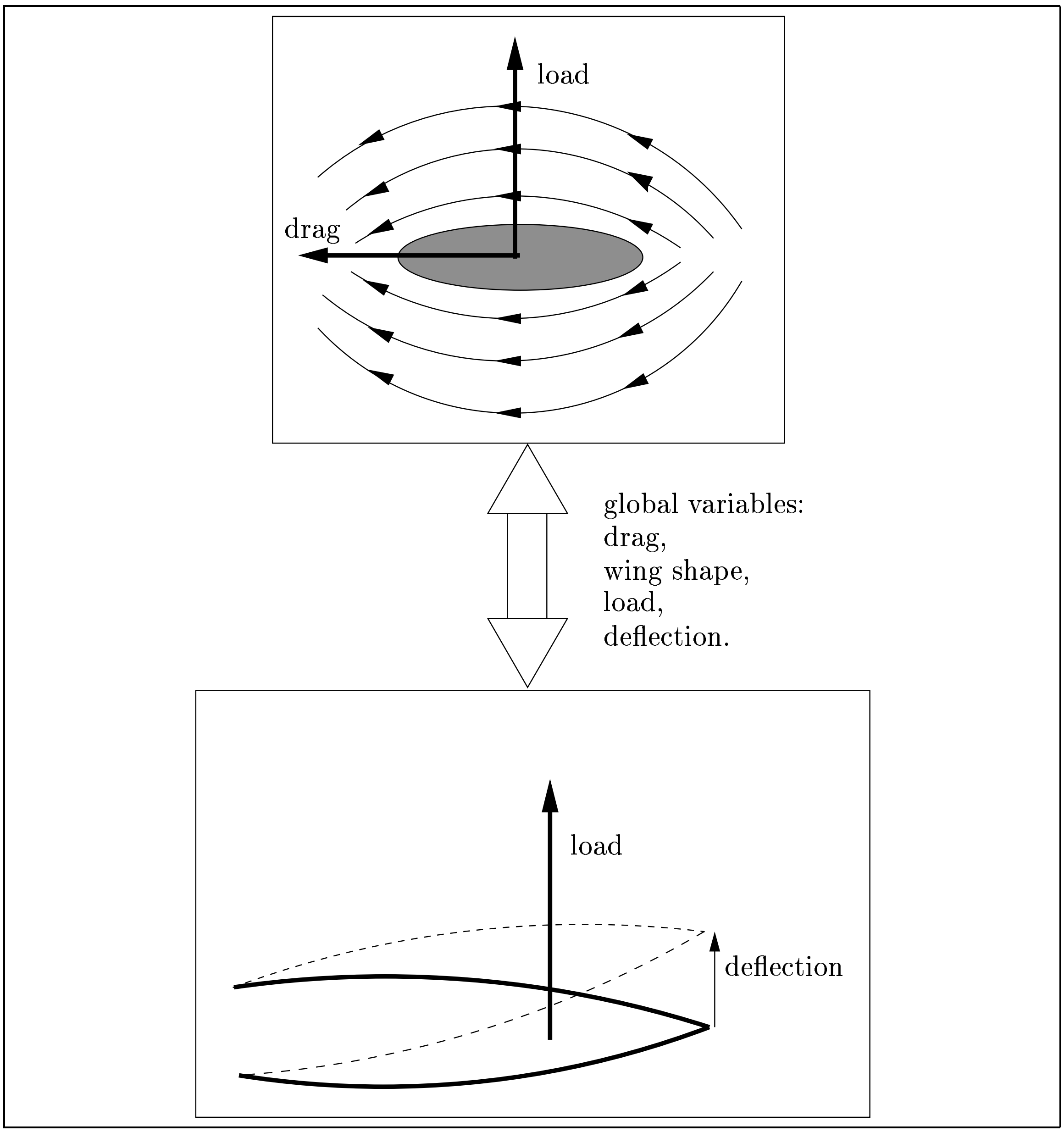

Figure 1.1: Aircraft wing design

The aircraft wing design is an instance of OPGV. Variables global to both the aerodynamic and the structure systems are the undeflected wing shape, the wing deflection, the load on the wing, and the drag. On the other hand, variables local to the aerodynamic analysis are the air pressure and speed distributions around the wing surface. Finally, variables local to the structure analysis are the wing structure geometry and the stress distribution in the wing interior. For a more detailed treatment of the aircraft wing design problem see Cramer et al. [CDF$^+$94] and the references therein.

1.2.2 Electricity Generation Planning

Consider an electricity producer operating a mix of thermal and hydro power plants. The operation of the thermal plants (nuclear, coal, and gas) involves important variable costs such as fuel and maintenance. In contrast, the variable costs associated with hydro plants are usually negligible. Consequently, the producer wishes to produce as much energy as possible using its hydro plants. Unfortunately, the amount of water available in the reservoir system is not enough in general to satisfy all the electricity demand. To complicate matters, in most cases water inflows into the reservoirs are highly stochastic. Scheduling how much water should be released for generation now and how much should be stored for future use is a major challenge for electricity producers.

To make a good decision the producer has to take into account the stochastic nature of the water inflows and the electricity demand. The probability distribution of the water inflows into the reservoirs is usually available. Moreover, we assume that the probability distribution of the electricity demand that the producer will serve is also known. This is not a strong assumption even in deregulated electricity markets, where producers often use historical data to forecast the total demand and their market share. Given the probability distributions for water inflows and electricity demand, the target is to compute the generation schedule that minimizes the expected total generation costs along a one-year time horizon.

Escudero et al. [EdGP96] proposed a scenario tree approach to model the stochasticity. Basically, the one-year time horizon is divided into T time periods. The water inflows and the demand for the first time period are considered deterministic. On the other hand, four different events are considered for the second time period: (i) high water inflows and high demand, (ii) high water inflows and low demand, (iii) low water inflows and high demand, and (iv) low water inflows and low demand. The probability of each of these events is assigned using the probability distributions for water inflows and demand. The scenario tree is the result of replicating this procedure for each of the remaining time periods. A typical scenario tree is depicted in Figure 1.2.

The producer uses the scenario tree to compute the production schedule that minimizes the expected total generation costs. The scenario tree is updated and the analysis rerun on a weekly basis.

This problem is an example of OPGV. Each node in the scenario tree can be considered as a component system within the optimization problem. The only variables global to all nodes are the amounts of water stored in the reservoirs at the end of each time period. Variables local to each node include the amount of water released for generation at each hydro power plant and the energy generated using each of the thermal plants.

1.3 Decomposition Algorithms

1.3.1 How Decomposition Algorithms Work

If, in an OPGV or OPGC, we eliminate the global variables or constraints, the problem breaks into N independent subproblems. Decomposition algorithms coordinate the solution to these N subproblems to find the minimizer to the original problem.

The coordination is carried out by a so-called *master problem*, an optimization problem whose objective and/or constraint functions are defined using information gathered at the subproblem solutions. At each iteration of the optimization algorithm solving the master problem, all of the N subproblems are solved and information is exchanged between the master problem and the subproblems.

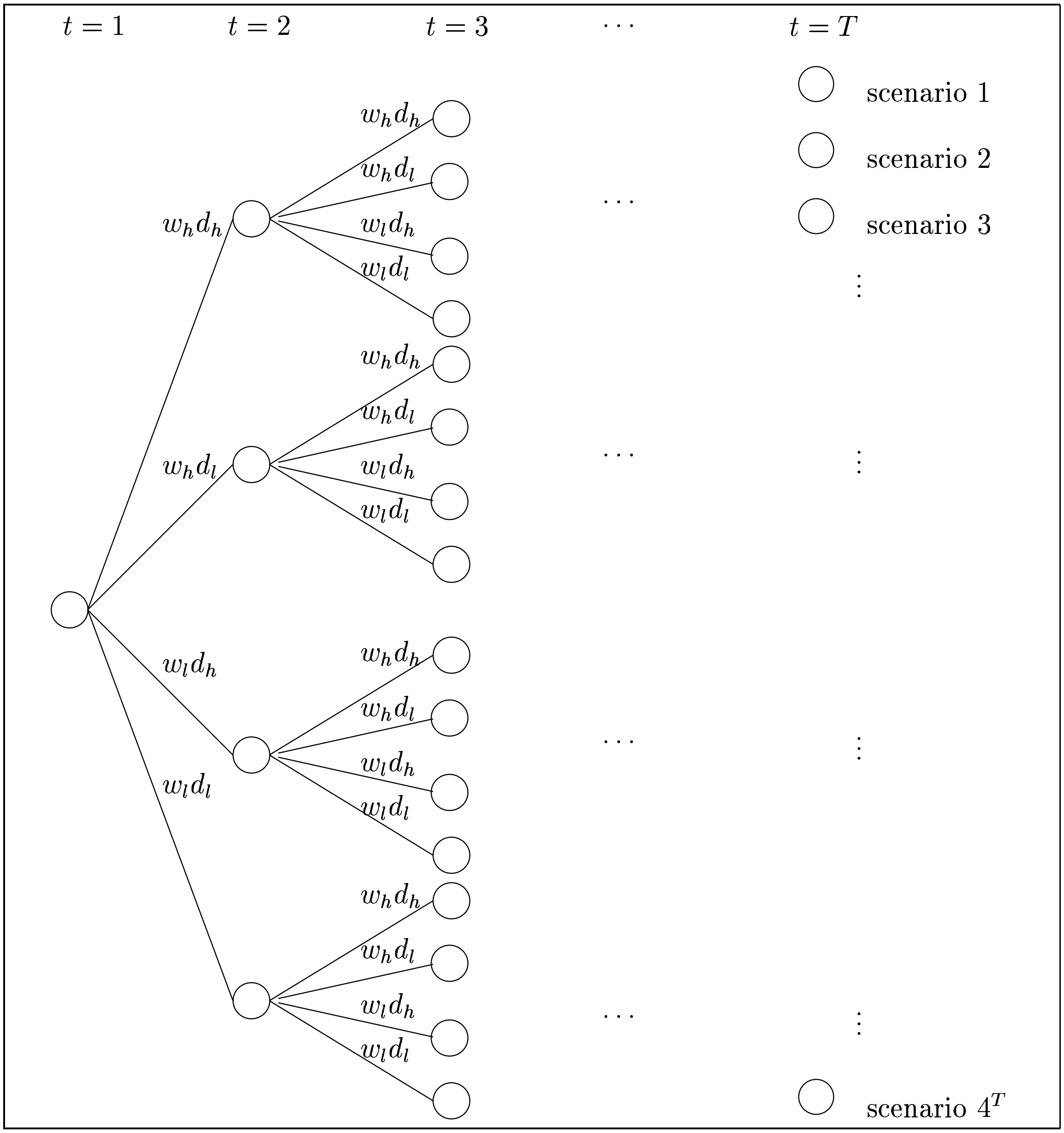

Figure 1.2: Scenario tree for electricity generation planning. The event high water inflows and high demand is represented as $w_h d_h$.

In the OPGV case, if we set the global variables to a fixed value, the problem breaks into N independent subproblems. A master problem is used to find the optimal value of the global variables. Local variables and constraints corresponding to the ith system are kept within the ith subproblem. In the OPGC case, N independent subproblems are obtained by ignoring the global constraints. A master problem subject only to the global constraints is used to ensure that the global constraints are satisfied. Local constraints corresponding to the ith system are kept within the ith subproblem.

1.3.2 Why Decompose?

There are computational and organizational advantages in the use of decomposition algorithms. From a computational perspective, the advantage is that the subproblems are usually easier to solve than the original problem. The subproblems are, by definition, smaller than the original problem. Moreover, in many cases the subproblems have special properties such as convexity, sparsity, or network constraints that enable the use of efficient specialized algorithms to solve them. By decomposing the original problem, we can take advantage of the efficient solution method available for the subproblems.

For example, in the electricity generation planning problem discussed, the original problem has hundreds of thousands of variables. By using decomposition algorithms, we can break the problem into one subproblem per tree node. Each of the subproblems has only hundreds of variables. Moreover, Escudero et al. show that, unlike the original problem, the subproblems have network structure. The advantage in the use of decomposition algorithms is that, in addition to the subproblems having only hundreds of variables, now we can apply specialized network algorithms to solve them.

In some cases the main motivation for the use of decomposition algorithms is related to their organizational aspects. Most engineering design problems involve the participation of different design groups who work largely in isolation [Kro97].

For instance, in the aircraft wing design problem, two different departments carry out the aerodynamic and the structure analysis. Each department must rely on complex software codes whose method of use is subject to constant modification. Porting all the code to a specific machine is judged to be impractical (sometimes the source code is not available). Also it would raise the issue of how local modifications to the use of such codes would be incorporated into the integrated code. What is required is some procedure to optimize the whole design while keeping the work of the different departments as independent as possible.

Decomposition algorithms allow these problems to be solved in a distributed environment in the manner described above. The key point in the design of a decomposition algorithm in this environment is that only a limited communication between the subproblems and the master problem is required. The aim is that different engineering teams should solve only their own subproblem and only a small amount of communication should be required with the central coordinator. For a survey of the application of decomposition algorithms to aerospace design problems see Haftka and Sobieszczanski-Sobieski [HSS97].

1.3.3 Analyzing Decomposition Algorithms

Alexandrov and Lewis [AL99] distinguish two different ingredients in the analysis of a decomposition algorithm:

- The equivalence between the minimizers to the original problem and the minimizers to the proposed master problem must be shown. Otherwise, when finding a minimizer to the master problem, we would not be able to claim that we found a minimizer to the original problem.

- The existence of globally and fast locally convergent optimization algorithms for both the master problem and the subproblems must be shown. Global convergence means the iterates will converge to a minimizer from any starting point, possibly far from any minimizer. Fast local convergence means the algorithm must converge to the minimizer at a high rate (superlinear or quadratic) once the iterates are sufficiently close to it. To prove the existence of globally and fast locally convergent optimization algorithms, we usually need to make assumptions about the optimization problem such as smoothness and nondegeneracy

[MP95]. When analyzing a decomposition algorithm, we would like to prove that if those assumptions are satisfied for the original problem, then they are also satisfied for the proposed master problem and the subproblems.

In the remainder of this section, we review some of the most popular decomposition algorithms proposed for the OPGV and the OPGC. For each of these algorithms, we study the two analysis ingredients introduced above.

1.3.4 State of the Art: Convex Problems

In 1960, Dantzig and Wolfe [DW60] developed an efficient decomposition algorithm to deal with the linear programming OPGC. A few years later, Benders [Ben62] developed a decomposition algorithm for OPGVs whose objective and constraint functions are linear in the local variables.

Both Dantzig-Wolfe and Benders decompositions are widely used to solve linear programs. The key to their widespread use is twofold. Firstly, we can prove global minimizer equivalence between the original problem and the proposed master problem. Secondly, there exist algorithms that find the minimizer to the master problem in a finite number of steps from any starting point.

Benders Decomposition

Benders decomposition [Ben62] deals with OPGVs whose objective and constraint functions are linear in the local variables. Geoffrion [Geo72] extended Benders decomposition to problems whose objective and constraints are convex in the local variables. Both the Benders and Geoffrion algorithms are efficient only when the problem functions are separable with respect to global and local variables.

In a Benders decomposition approach, local variables and constraints corresponding to the ith system are kept within the ith subproblem. The master problem depends only on the global variables. However, the master problem includes a possibly large number of newly generated constraints known as *cuts*.

At each iteration, a relaxed version of the master problem (that is, a master problem with only a few of the cuts) is solved and the current estimate of the global variables x_k is obtained. Then all subproblems are solved using x_k as a parameter. When a given subproblem is feasible at x_k, a new *optimality cut* is generated. If, on the other hand, a subproblem is infeasible at x_k, a *feasibility cut* is generated. The current relaxed master problem is updated by including all cuts generated together with all or some of the cuts available previously.

Benders showed that a minimizer to the master problem is a minimizer to the original problem. He also proved that, for linear programs, a finite number of iterations suffices to find the minimizer from any starting point. For the convex case Geoffrion showed that an approximate solution can be found in a finite number of iterations.

The main limitation of Benders decomposition is that it can only be used when the problem functions are convex in the local variables, the reason being that the cuts are generated by means of convex duality theory. A more detailed discussion of Benders decomposition is given in Chapter 2.

Dantzig-Wolfe Decomposition

Dantzig-Wolfe decomposition [DW60] deals with linear programming OPGCs. In a Dantzig-Wolfe scheme, local constraints are kept within the ith subproblem. The master problem includes only the global constraints. The difficulty is that a change of variables is introduced in the master problem that results in a possibly large number of variables.

At each iteration, a relaxed version of the master problem (that is, a master problem including only a few of the variables) is solved. Then the N subproblems are solved using the reduced costs of the master linear program as parameters. As a result, each subproblem generates a candidate variable to be introduced in the master problem. The current relaxed master problem is updated by including all candidate variables found by the subproblems.

Dantzig and Wolfe showed that a minimizer to the master problem is a minimizer to the original linear program. They also proved that a finite number of iterations suffices to find the minimizer from any starting point.

The Dantzig-Wolfe decomposition is applicable only to the linear case, the reason being that to introduce the change of variables in the master problem we make use of the Resolution Theorem for convex polyhedra [Gol56], which can only be applied in the linear case. For detailed discussions of Dantzig-Wolfe decomposition see [Dan63, Chapter 23] and [DT, Chapter 2].

1.3.5 State of the Art: Nonconvex Problems

Decomposition algorithms transform a weakly connected optimization problem into a master problem and a set of subproblems. Optimization problems involving a master problem and subproblems are known as *bilevel programs* [Bar98, FL95, SIB97]. Unfortunately, bilevel programs are difficult to solve.

Dantzig and Wolfe, and Benders developed efficient ways to deal with the bilevel programs resulting from the decomposition of linear programs. The situation is a lot more complicated for nonconvex problems. Few decomposition algorithms have been proposed for nonconvex problems. Although local minimizer equivalence can be proven for most approaches, we do not know of any globally convergent algorithms. Some fast locally convergent algorithms have been proposed but they rely on strong nondegeneracy assumptions or the use of optimization algorithms for nonsmooth problems.

Tammer's Decomposition

Tammer proposed a decomposition algorithm for the nonconvex OPGV [Tam87]. The subproblems are obtained by setting the global variables to a fixed value. The optimal objective function to a subproblem as a function of the global variables is known as the optimal-value function. The master problem is the unconstrained problem whose objective function is the summation of the optimal-value functions corresponding to all subproblems.

At each iteration, all of the subproblems are solved and a master problem search direction is generated. Then, a line search is performed to ensure a sufficient descent for the master problem objective while keeping feasibility with respect to the original OPGV constraints.

Tammer showed that for every OPGV minimizer satisfying the Strong Linear Independence Constraint Qualification[1] (SLICQ) and the Strong Second Order Sufficient Conditions (SSOSC) there exists an equivalent master problem minimizer. Moreover, assuming the SLICQ, the Second Order Sufficient Conditions (SOSC), and the Strict Complementarity Slackness Conditions (SCSC) hold at the OPGV minimizer, he showed that optimization algorithms for smooth problems achieve fast local convergence when applied to the master problem and the subproblems.

The main limitation to this approach is that the nondegeneracy conditions assumed to prove fast local convergence include the SLICQ. This condition implies that, at the minimizer to the OPGV, for any small perturbation of the global variables, we can find values of the local variables that are feasible with respect to the OPGV constraints. The SLICQ is not likely to hold for many real problems.

Collaborative Optimization

Collaborative Optimization (CO) is a decomposition algorithm proposed by Braun [Bra96] for nonconvex OPGVs. In a CO scheme, the global variables are allowed to take different values within each of the subproblems. However, inexact (quadratic) penalty functions are used as the subproblem objective functions to ensure that the value of the global variables in each of the subproblems converges to the so-called *target variables*.

As in Tammer's decomposition, the subproblem solutions are used to compute a search direction for the master problem. Then, a line search is performed to ensure we achieve a sufficient descent in the master problem objective. At the same time, as a result of the use of quadratic penalty terms, the global variables in each subproblem may be adjusted to ensure feasibility. This is an important advantage compared to Tammer's decomposition.

Braun showed minimizer equivalence between the original problem and the proposed master problem. Unfortunately, both the CO master problem and the subproblems are degenerate [AL00, dM00]. In particular, the subproblems do not satisfy the SCSC and therefore the master problem is not differentiable in general. The

[1] See Appendix A for a review of nondegeneracy conditions.

SCSC and the smoothness of the objective and constraint functions are common assumptions to prove fast local convergence. Therefore we can expect difficulty when trying to solve the CO master problem and subproblems using algorithms for smooth problems.

Despite the fact that no convergence proof is known for CO, Braun's approach has an important advantage over Tammer's decomposition. Namely, the subproblems are feasible for any value of the target variables even if the SLICQ does not hold. In Chapter 2, we give a detailed analysis of CO. We also propose two alternative decomposition algorithms that overcome the degeneracy and nonsmoothness difficulties inherent in Braun's approach.

Nonsmooth Bilevel Programming Approaches

Tammer showed that if the SLICQ, SCSC, and SOSC hold at the OPGV minimizer, then the optimal-value functions for the subproblems he proposed are twice continuously differentiable and therefore optimization algorithms for smooth problems show fast local convergence when applied to the master problem [Tam87].

Under weaker nondegeneracy assumptions it is possible to prove that the subproblem optimal-value functions are locally Lipschitz. Then an optimization algorithm for nonsmooth problems such as a bundle method [Mif77, SZ92, HUL93] can be used to solve the resulting Lipschitz master problem. In particular, assuming the OPGV minimizer satisfies the SLICQ and the SSOSC, Tammer showed that the master problem is Lipschitz. Moreover, he proved that a method of feasible directions is locally convergent when applied to the master problem.

Shimizu, Ishizuka, and Bard [SIB97, Chapter 8] propose a master problem for the nonconvex separable OPGC. Assuming the OPGC minimizer satisfies the SLICQ and that the subproblem feasible region is uniformly compact near the minimizer they show that their master problem objective function is Lipschitz. They also prove that the bundle method by Mifflin [Mif77] is locally convergent when applied to solve the resulting Lipschitz master problem.

1.4 Main Contributions

There is a recognized demand for an efficient decomposition algorithm for the nonconvex OPGV. Firstly, the severe SLICQ is required to show local convergence for Tammer's decomposition. Secondly, no convergence proof is known for CO because of the degeneracy difficulties associated with it. Finally, there exist several nonsmooth bilevel programming approaches but they preclude the use of the efficient and reliable optimization software available for smooth problems such as NPSOL or SNOPT [GMSW86, GMS97].

In this dissertation we propose two novel decomposition algorithms based on optimization techniques for smooth problems and show that they overcome the difficulties associated with CO even when only the LICQ holds instead of the more restrictive SLICQ. As a consequence, fast local convergence can be expected from optimization algorithms for smooth problems when applied to the proposed master problems and subproblems.

A major difficulty in developing a decomposition algorithm for the nonconvex OPGV is the lack of an adequate test-problem set. To fill this gap, we introduce a new quadratic programming OPGV test-problem set. The user can control problem characteristics such as dimension, convexity, degeneracy, and degree of coupling among systems. We use the test-problem set to investigate the numerical behavior of the new decomposition algorithms and show that both of them behave satisfactorily on the test set for a wide range of circumstances.

1.5 Overview of Remaining Chapters

Chapter 2 gives a detailed analysis of the most relevant decomposition algorithms available for the OPGV. We also propose two novel decomposition algorithms that we term Inexact Penalty Decomposition (IPD) and Exact Penalty Decomposition (EPD).

In Chapter 3, we give nondegeneracy results proving that optimization algorithms for smooth problems are fast locally convergent on the IPD and EPD master problems and subproblems. We also discuss the difficulties encountered when we try to prove global convergence for decomposition algorithms for the nonconvex OPGV.

In Chapter 4, we present a new quadratic programming OPGV test-problem set. In Chapter 5, we discuss the numerical behavior of IPD and EPD on the new test-problem set. Finally, we give some concluding remarks and directions of future research in Chapter 6.

Chapter 2

Decomposition Algorithms

We review several decomposition algorithms for the OPGV. We describe how generalized Benders decomposition efficiently deals with the convex OPGV. We discuss the difficulties associated with several decomposition algorithms available for nonconvex OPGVs and propose two new decomposition algorithms that overcome some of these difficulties.

2.1 Generalized Benders Decomposition

In this section we first show how to transform an OPGV whose objective and constraint functions are convex in the local variables into a master problem and a set of subproblems. Then we give an algorithm to find an approximate minimizer to the resulting master problem. Finally, we prove that the algorithm converges in a finite number of iterations. For a more detailed study of Benders decomposition see Benders [Ben62] and Geoffrion [Geo72].

2.1.1 Problem Formulation

The master problem is obtained from the OPGV by a sequence of two manipulations: (i) projection and (ii) dualization. Projection is the result of setting the global variable vector to a fixed value. As a consequence, the OPGV breaks into the N independent subproblems,

$$\begin{array}{rlcl} \min\limits_{y_i} & F_i(x, y_i) & & \\ s.t. & c_i(x, y_i) & \geq & 0, \end{array} \qquad i = 1{:}N. \tag{2.1}$$

Note that x is a parameter in each of the subproblems. The optimal objective for the ith subproblem as a function of the global variables is called the optimal-value function $F_i^*(x)$. The optimal-value functions are used to form the following master problem:

$$\min_x \sum_{i=1}^{N} F_i^*(x). \tag{2.2}$$

The equivalence between the master problem (2.2) and the OPGV in terms of global minimizers follows from Theorem 2.1 in Geoffrion [Geo72]. The main benefit obtained from projection is that now we deal with a master problem that only depends on the global variables x. Local variables and constraints belonging to the ith system are kept within the ith subproblem.

The second transformation is called dualization because it makes use of convex duality theory. Let X_i be the set of values of x such that the ith subproblem defined in (2.1) is feasible. When F_i and $-c_i$ are convex for fixed x, and c_i satisfies some additional mild conditions, Geoffrion [Geo72] showed, using convex duality theory, that $x \in X_i$ iff

$$\sup_{y_i} \quad u_i^T c_i(x, y_i) \geq 0, \qquad \forall u_i \geq 0. \tag{2.3}$$

Moreover, if in addition to F_i and $-c_i$ being convex for fixed $x \in X_i$, for all $x \in X_i$ either the ith subproblem is unbounded or there exists a minimizer satisfying the LICQ, then

$$F_i^*(x) = \sup_{v_i \leq 0} \left[\inf_{y_i} F_i(x, y_i) + v_i^T c_i(x, y_i) \right], \quad \forall x \in X_i. \tag{2.4}$$

Let

$$\mathcal{F}_i(x, u_i) = \sup_{y_i} \quad u_i^T c_i(x, y_i)$$

and

$$\mathcal{O}_i(x, v_i) = \inf_{y_i} \quad F_i(x, y_i) + v_i^T c_i(x, y_i).$$

Then, from (2.3), (2.4), and the definition of the infimum as the greatest lower bound the master problem can be written as

$$\begin{array}{llcll} \min\limits_{x,\gamma_i} & \sum\limits_{i=1}^{N} \gamma_i & & & \\ s.t. & \mathcal{F}_i(x,u_i) & \geq & 0, & \forall u_i \geq 0, \quad i = 1{:}N, \\ & \mathcal{O}_i(x,v_i) & \leq & \gamma_i, & \forall v_i \leq 0, \quad i = 1{:}N, \end{array} \tag{2.5}$$

where the constraints $\mathcal{F}_i(x,u_i) \geq 0$ are known as *feasibility cuts* and the constraints $\mathcal{O}_i(x,v_i) \leq \gamma_i$ are known as *optimality cuts*. The optimality cuts build an approximation of the optimal-value functions $F_i^*(x)$. Likewise, the feasibility cuts approximate the region formed by the values of x that make all the subproblems feasible, namely $\bigcap_{i=1}^{N} X_i$. Note that, in general, there might exist an infinite number of feasibility and optimality cuts.

2.1.2 Computational Procedure

In this section, we outline an algorithm leading to an approximate minimizer of the master problem (2.5) in a finite number of iterations. The main difficulty in solving (2.5) is that it has an infinite number of constraints. A natural strategy to deal with a large number of constraints is relaxation. First, solve a relaxed master problem including only a few of the constraints. If the solution to the relaxed master problem is not feasible with respect to all constraints, then add some of the violated constraints to the relaxed master problem and compute a new solution. We iterate this procedure until a solution feasible to all constraints is found.

When the solution to the kth relaxed master problem x_k is computed, the N subproblems are solved using x_k as a parameter. If a given subproblem is feasible, the computed Lagrange multipliers v_i are used to generate a new optimality cut for the master problem. If, on the other hand, a subproblem is infeasible, we compute $u_i \geq 0$ such that $\mathcal{F}_i(x,u_i) < 0$ and a new feasibility cut is generated.

The minimum to the relaxed master problem is a lower bound on the original OPGV because we take only a few of the cuts into account. In contrast, the minimum to the subproblems is an upper bound because we do satisfy all of the OPGV constraints, but we set the global variables x to a fixed value x_k. The procedure is terminated when the lower and upper bounds are sufficiently close. The procedure is described by the following three steps:

- **Step 1.** Given the current iterate for the global variables x_1, solve the N subproblems defined by (2.1). Set $p = 1$ and generate $u_i^p \geq 0$ for $i \in \Phi_p = \{i : i\text{th subproblem is infeasible for } x_1\}$ and $v_i^p \leq 0$ for $i \in \Theta_p = \{i : i\text{th subproblem is feasible for } x_1\}$. If all subproblems are feasible, then set the objective upper bound $U_b = \sum_{i=1}^N F_i^*(x_k)$, otherwise set $U_b = \infty$. Select the convergence tolerance ϵ.

- **Step 2.** Solve the current relaxed master problem,

$$\begin{array}{rlcl} \min\limits_{x,\gamma_i} & \sum\limits_{i=1}^{N} \gamma_i & & \\ s.t. & \mathcal{F}_i(x, u_i^j) & \geq & 0, \quad \forall j = 1{:}p, \quad i \in \Phi_j, \\ & \mathcal{O}_i(x, v_i^j) & \leq & \gamma_i, \quad \forall j = 1{:}p, \quad i \in \Theta_j. \end{array} \tag{2.6}$$

  Let $(\hat{x}, \hat{\gamma}_i)$ be a minimizer. Then $\sum_{i=1}^N \hat{\gamma}_i$ is a lower bound for the OPGV. If $U_b \leq \sum_{i=1}^N \hat{\gamma}_i + \epsilon$, stop.

- **Step 3.** Solve the N subproblems at $\hat{x}$. If $\sum_{i=1}^N F_i^*(\hat{x}) \leq \sum_{i=1}^N \hat{\gamma}_i + \epsilon$, stop. Otherwise, increase p by one and generate $\hat{u}_i^p \geq 0$ for $i \in \Phi_p = \{i : i\text{th subproblem is infeasible for } \hat{x}\}$ and $v_i^p \leq 0$ for $i \in \Theta_1 = \{i : i\text{th subproblem is feasible for } \hat{x}\}$. If all the subproblems are feasible, then set the objective upper bound $U_b = \min(U_b, \sum_{i=1}^N F_i^*(x_k))$. Return to Step 2.

Remark 2.1 *The computational procedure stated above is efficient only if the feasibility and optimality cuts are easy to evaluate. A condition sufficient for this is that the OPGV objective and constraint functions are separable in the global and local variables. Geoffrion [Geo72] discusses other situations under which the cuts are easy to evaluate.*

2.1.3 Finite Convergence

The following theorem is a consequence of Theorem 2.5 in Geoffrion [Geo72] and shows that the computational procedure defined converges to an approximate solution in a finite number of steps.

Theorem 2.2 *Assume that, for all* $i = 1{:}N$, X_i *is a nonempty compact set, the feasible region for the ith subproblem for all* $x \in X_i$ *is a nonempty compact set, the OPGV objective and constraint functions are convex and continuous for fixed* x, *and the set of optimal multiplier vectors for the ith subproblem is nonempty for any* x *and uniformly bounded in some neighborhood of each such point. Then, for any given* $\epsilon > 0$, *the generalized Benders decomposition procedure terminates in a finite number of steps.*

The most restrictive assumption used in the above theorem is the one regarding convexity of the OPGV. In the remainder of this dissertation we focus on nonconvex OPGVs.

2.2 Tammer's Decomposition

Geoffrion showed that Benders decomposition is a globally and fast locally convergent algorithm for the convex OPGV. Moreover, the computational procedure he proposed finds a global minimizer. The situation is far more complicated for the nonconvex OPGV. For the nonconvex case, one can only aspire to find local minimizers. Moreover, we believe global convergence can not be proven for a decomposition algorithm applied to the nonconvex OPGV. However, we would like to have fast locally convergent decomposition algorithms.

Nonconvex duality theory [TW81] leads to decomposition algorithms that require the solution of problems in infinite dimensional spaces [FK93]. Therefore, if finite dimensionality is to be preserved, out of the two manipulations used to obtain the Benders master problem, only projection can be used for nonconvex OPGVs. Tammer [Tam87] proposed a decomposition algorithm, based on projection, for which he showed local convergence under rather restrictive nondegeneracy assumptions.

2.2.1 Problem Formulation

Tammer proposed solving the following master problem:

$$\min_x \sum_{i=1}^{N} F_i^*(x) \tag{2.7}$$

where $F_i^*(x)$ is the optimal-value function corresponding to the ith subproblem,

$$\begin{aligned} F_i^*(x) \quad = \quad & \min_{y_i} \quad F_i(x, y_i) \\ & s.t. \quad c_i(x, y_i) \geq 0. \end{aligned} \tag{2.8}$$

The proposed master problem and subproblems are obtained by setting the global variables to a fixed value. As in Benders, constraints and terms in the objective function corresponding to the ith system are kept within the ith subproblem. The difference is that, in the absence of an adequate duality theory, cuts are not introduced in the master problem. Instead, the master problem is just the unconstrained problem whose objective function is the summation of the subproblem optimal-value functions.

2.2.2 Analysis

A major difficulty in Tammer's approach is that the algorithm fails when it arrives at a value of the global variables for which one of the subproblems is infeasible. To preclude this possibility, Tammer assumes the restrictive Strong Linear Independence Constraint Qualification (SLICQ).

To define the SLICQ we need to introduce a little notation. The OPGV Jacobian at a point $(x, y_1, \ldots, y_N)$ is

$$\begin{pmatrix} A_1 & B_1 & & & \\ A_2 & & B_2 & & \\ \vdots & & & \ddots & \\ A_N & & & & B_N \end{pmatrix},$$

where $A_i = \nabla_x \hat{c}_i(x, y_i)$, $B_i = \nabla_{y_i} \hat{c}_i(x, y_i)$, and $\hat{c}_i$ are the active constraints.

Definition 2.3 The SLICQ holds at a feasible OPGV point if for $i = 1{:}N$ the matrix B_i has full rank.

By the implicit function theorem, we know that if the SLICQ holds at a feasible point $(x, y_1, \ldots, y_N)$, then the subproblems defined in (2.8) are feasible for any value of the global variables in a neighborhood of x. By assuming the SLICQ holds at the OPGV minimizer, Tammer ensures that the subproblems are always feasible once the iterates are sufficiently close to the solution.

Minimizer Equivalence

Tammer proved that if the SLICQ and the Strong Second-Order Sufficient Conditions (SSOSC) hold at an OPGV minimizer $(x^*, y_1^*, \ldots, y_N^*)$, then x^* is an isolated minimizer to the master problem (2.7).

Fast Local Convergence

Assuming an OPGV minimizer satisfies the SSOSC and the SLICQ, Tammer showed local convergence for a method of feasible directions applied to the master problem. If, in addition to SSOSC and SLICQ, the SCSC hold, fast local convergence can also be proven for Newton-type algorithms for smooth problems.

Tammer gave a fast locally convergent algorithm for the nonconvex OPGV under the assumption that SLICQ holds at the minimizer to the OPGV. We propose two decomposition algorithms that are fast locally convergent under the less restrictive Linear Independence Constraint Qualification (LICQ). The SLICQ is in fact a sufficient but not necessary condition for LICQ.

2.3 Collaborative Optimization

Tammer assumed the SLICQ to ensure that the subproblems are always feasible once the iterates are sufficiently close to the minimizer. Unfortunately, the SLICQ does not hold for many problems. Braun [Bra96] proposed a decomposition algorithm, known as Collaborative Optimization (CO), that uses quadratic penalty functions to formulate subproblems that are always feasible even if the SLICQ does not hold.

However, in this section we show that the master problem and the subproblems proposed by Braun are degenerate at the solution. Moreover, the subproblem degeneracy implies that the master problem is nonsmooth. Nondegeneracy and smoothness are usual assumptions for most local convergence proofs available for optimization algorithms. Therefore, we expect numerical difficulty in the computation of the minimizers of the CO master problem and subproblems. This expectation is supported by numerical results that show how CO fails to solve some simple test problems [AK98, AL00].

2.3.1 Problem Formulation

Braun proposed the following master problem:

$$\begin{array}{ll} \min\limits_{z} & F(z) \\ s.t. & p_i^*(z) = 0, \quad i = 1{:}N, \end{array} \tag{2.9}$$

where $p_i^*(z)$ is the optimal-value function corresponding to the ith subproblem,

$$\begin{array}{rcll} p_i^*(z) & = & \min\limits_{x_i, y_i} & \frac{1}{2}\|x_i - z\|_2^2 \\ & & s.t. & c_i(x_i, y_i) \geq 0. \end{array} \tag{2.10}$$

Braun allows the global variables to take a different value x_i within each of the subproblems. A quadratic penalty is used as the subproblem objective function to force the x_i to converge to the so-called target variables z. At a solution to the master problem, $p_i^*(z) = 0$ and therefore $x_i = z$ for $i = 1{:}N$. Note that CO only works when the OPGV objective function depends exclusively on the global variables (that is, the OPGV objective function $\sum_i F_i(x, y_i)$ takes the specific form $F(x)$) since otherwise the master problem would also depend on the local variables. Finally, Braun in his numerical experiments used a slightly more elaborate form than the one given above, but we retain this form for expository purposes.

2.3.2 Minimizer Equivalence

It is easy to show that the OPGV feasible region and the CO master problem feasible region are identical. Since the objective function corresponding to both problems is the same, the OPGV and the CO master problem obviously have the same set of minimizers.

2.3.3 Fast Local Convergence

Subproblem Degeneracy

The ith subproblem objective function gradient is

$$\nabla_{x_i,y_i}\frac{1}{2}\|x_i - z\|_2^2 = \begin{pmatrix} x_i - z \\ 0 \end{pmatrix}.$$

At the solution, $x_i^* = z$ and therefore the ith subproblem objective function gradient is zero. Given that the original OPGV satisfies LICQ at its minimizers, this in turn implies that the subproblem Lagrange multipliers are zero and therefore the SCSC do not hold at the subproblem minimizer.

Master Problem Nonsmoothness

The degeneracy of the CO subproblem minimizers implies that the subproblem optimal-value functions $p_i^*(z)$ are not smooth in general. Fiacco and McCormick [FM68] showed that the optimal-value function to a parametric nonlinear program is smooth at a point z if the minimizer to the parametric nonlinear program satisfies LICQ, SCSC, and SOSC. Unfortunately, we have shown that the SCSC are not satisfied at the subproblem minimizer. Therefore, $p_i^*(z)$ is not differentiable in general.

Master Problem Degeneracy

Assuming the optimal-value functions $p_i^*(z)$ are smooth, Braun shows that the gradients for the master problem constraints can be computed analytically as

$$\nabla p_i^*(z) = -(x_i^* - z).$$

Clearly, even when each $p_i^*(z)$ is smooth the Jacobian for the master problem constraints is singular at the solution (indeed it becomes the zero matrix since at the solution $x_i = z$). Thus the LICQ does not hold at the master problem minimizer.

Master Problem Active Set

A somewhat more subtle problem is the difficulty in identifying the active set of the master problem. It may seem odd to suggest difficulty in identifying the active set when only equality constraints are present. However, a worrying feature of the subproblems is that they fail to distinguish the case when $x_i^* = z$ is only just feasible, from the case when any change in z of sufficiently small magnitude would still result in $x_i^* = z$. In other words, this formulation is not able to identify those constraints in the master problem that are truly constraining the solution from those that are not.

2.3.4 Global Convergence

Global convergence proofs for optimization algorithms for smooth problems can not be applied to the CO master problem because of the nonsmoothness of $p_i^*(z)$. This difficulty could be overcome by the use of optimization algorithms for nonsmooth problems such as bundle methods [Mif77, SZ92, HUL93]. However, this precludes the use of optimization algorithms for smooth problems.

Even if we used an optimization algorithm for nonsmooth problems, global convergence may still be hindered (for the nonconvex case) by the existence of multiple local minimizers to the CO subproblems. In that case $p_i^*(x)$ is not a function but rather a set-valued function.[1] No global convergence proof is known for optimization algorithms applied to set-valued functions.

2.4 Inexact Penalty Decomposition (IPD)

In Section 2.3, we showed that there are numerical and analytical difficulties associated with the use of CO. However, an advantage of CO is that its subproblems are feasible for any value of the target variables. Moreover, CO has been successfully applied to the solution of some real problems [Man99, Sob98]. In this section we propose a new decomposition algorithm based also on the use of quadratic penalty

[1] See Aubin and Frankowska [AF90] for a reference on set-valued analysis.

functions that overcomes some of the difficulties associated with CO. We term the algorithm Inexact Penalty Decomposition (IPD).

2.4.1 Problem Formulation

The degeneracy of the CO master problem and subproblems is due to the specific form in which quadratic penalty functions are used. Here, we propose a more classical use of quadratic penalty functions. Solve the following master problem:

$$\min_{z} \quad \sum_{i=1}^{N} F_i^*(\gamma, z) \tag{2.11}$$

where $F_i^*(z)$ is the optimal-value function corresponding to the ith subproblem,

$$\begin{aligned} F_i^*(\gamma, z) \quad = \quad & \min_{x_i, y_i} \quad F_i(x_i, y_i) + \gamma \|x_i - z\|_2^2 \\ & s.t. \quad c_i(x_i, y_i) \geq 0. \end{aligned} \tag{2.12}$$

Unlike CO, IPD keeps the OPGV objective function term $F_i(x_i, y_i)$ within the ith subproblem. Then, a penalty parameter γ is used to weight the quadratic penalty term $\|x_i - z\|_2^2$ with respect to $F_i(x, y_i)$. Quadratic penalty functions are *inexact* penalty functions in the sense that the exact solution ($x_i = z$) is retrieved only for $\gamma = \infty$. Thus, the IPD master problem needs be solved for a sequence of penalty parameters $\{\gamma_k\}$ such that $\lim_{k\to\infty} \gamma_k = \infty$. Another difference between IPD and CO is that IPD uses the subproblem optimal-value functions $F_i^*(z)$ as penalty terms within the objective function of an *unconstrained* master problem.

2.4.2 Minimizer Equivalence

In Chapter 3, we show that there exists a trajectory of minimizers to the IPD master problem converging to any OPGV minimizer satisfying the LICQ, SCSC, and SOSC. The OPGV minimizer can be found by solving the IPD master problem for a sequence of penalty parameters $\{\gamma_k\}$ such that $\lim_{k\to\infty} \gamma_k = \infty$.

2.4.3 Fast Local Convergence

The main advantage of IPD over CO is that, as we show in Chapter 3, the IPD master problem and subproblem minimizers satisfy the LICQ, SCSC,and SOSC. Moreover, the master problem is smooth in a neighborhood of the minimizer. Therefore, optimization algorithms for smooth problems will show fast local convergence when applied to the IPD master problem and subproblems. This improves on the result by Tammer because the restrictive SLICQ is not needed.

The difficulty is now we need to solve the master problem for a sequence of penalty parameters $\{\gamma_k\}$ such that $\lim_{k\to\infty}\gamma_k = \infty$. For large γ, numerical difficulty might be expected due to the ill conditioning introduced by quadratic penalty functions [Mur71, Wri99]. Nevertheless, we believe certain sequential quadratic programming algorithms [GMSW86, GMS97, Mur97] can resolve this ill conditioning satisfactorily.

2.5 Exact Penalty Decomposition (EPD)

An alternative to inexact penalty functions are the so-called exact penalty functions [NW99, Chapter 17]. The term *exact* refers to the fact that the exact solution ($x_i = z$) is computed for finite values of the penalty parameter γ. Thus exact penalty functions avoid the ill conditioning associated with large penalty parameters.

In this section, we propose a novel decomposition algorithm based on the use of an exact penalty function. The advantage is that the penalty parameter need no longer be driven to infinity. The difficulty is that, as a consequence of the use of exact penalty functions, the subproblem optimal-value functions become nonsmooth. To alleviate this difficulty, we propose solving a sequence of perturbed problems that have better smoothness properties. We call the new algorithm Exact Penalty Decomposition (EPD).

2.5.1 Problem Formulation

Solve the following master problem:

$$\min_{z} \sum_{i=1}^{n} F_i^*(z)$$

where $F_i^*(z)$ is the optimal-value function corresponding to the ith subproblem,

$$\begin{array}{rcll} F_i^*(z) & = & \min\limits_{x_i,y_i,r_i,s_i,t_i} & F_i(x_i,y_i) + \gamma e^T(s_i+t_i) \\ & & s.t. & c_i(x_i,y_i) - r_i = 0 \\ & & & x_i + s_i - t_i = z \\ & & & r_i, s_i, t_i \geq 0. \end{array} \tag{2.13}$$

The exact penalty function used is the l_1 penalty function $\|x_i - z\|_1 = \sum_{j=1}^{n} |x_{ij} - z_j|$. However, to avoid the nonsmoothness of the absolute value function, rather than using the l_1 exact penalty function explicitly, we introduce elastic variables s_i and t_i. Then, it can be shown that the l_1 exact penalty function may be computed as $\|x_i - z\|_1 = e^T(s_i + t_i)$. We also introduce slack variables r_i in the constraints so that the subproblems only have equality constraints and nonnegativity bounds.

The advantage of *exact* penalty functions is that it suffices to solve the master problem for a sufficiently large but finite value of γ. The difficulty is that, if only the LICQ holds at the OPGV minimizer, then only the Mangasarian-Fromovitz Constraint Qualification holds, in general, at the subproblem minimizer, and not the LICQ. This implies that the optimal-value function $F_i^*(z)$ is not smooth in general.

We overcome this difficulty by using barrier terms to remove the nonnegativity constraints from the subproblems. Then we solve the following master problem for a decreasing sequence of barrier parameters $\{\mu_k\}$ such that $\lim_{k\to\infty} \mu_k = 0$:

$$\min_{z} \sum_{i=1}^{N} F_i^*(\mu, z) \tag{2.14}$$

where $F_i^*(\mu, z)$ is the optimal-value function corresponding to the ith subproblem,

$$\begin{array}{rcll} F_i^*(\mu,z) & = & \min\limits_{x_i,y_i,r_i,s_i,t_i} & F_i(x_i,y_i) + \gamma e^T(s_i+t_i) - \mu\phi(r_i,s_i,t_i) \\ & & s.t. & c_i(x_i,y_i) - r_i = 0 \\ & & & x_i + s_i - t_i = z, \end{array} \tag{2.15}$$

where the barrier function ϕ is defined as

$$\phi(r_i, s_i, t_i) = \sum_{j=1}^{m_i} \log r_{ij} + \sum_{j=1}^{n} (\log s_{ij} + \log t_{ij}).$$

2.5.2 Minimizer Equivalence

In Chapter 3, we show that there exists a trajectory of minimizers of the EPD master problem (2.14) converging to any OPGV minimizer satisfying the LICQ, SCSC, and SOSC. Thus, we can find the OPGV minimizer by driving μ to zero.

2.5.3 Fast Local Convergence

As in the IPD case, the main advantage of EPD over CO is that the EPD minimizers are nondegenerate with respect to the master problem and the subproblems. Since the EPD subproblems (2.15) only have equality constraints, the SCSC are automatically satisfied at any minimizer. Moreover, the LICQ is obviously satisfied because the EPD subproblem Jacobian clearly has full rank. Finally, the SOSC are likely to hold at the minimizer to the subproblems because in a barrier formulation, usually isolated minimizers are attained. In fact, we show in Chapter 3 that if an OPGV minimizer satisfies LICQ, SCSC, and SOSC, then the equivalent EPD minimizer is nondegenerate with respect to the master problem and the subproblems.

Consequently, optimization algorithms for smooth problems will be fast locally convergent on the EPD master problem and subproblems. The difficulty is that, for small μ, both the master problem and the subproblems may become ill conditioned [Mur71, Wri99]. However, we believe that sequential quadratic programming algorithms [GMSW86, GMS97, Mur97] can be used to deal efficiently with the ill-conditioned master problems and primal-dual interior point methods can deal with the ill conditioning introduced by the barrier terms in the EPD subproblems.

Chapter 3

Convergence Analysis

We study the convergence properties of Inexact Penalty Decomposition (IPD) and Exact Penalty Decomposition (EPD). We show that for every OPGV minimizer satisfying the LICQ, SCSC, and SOSC, there exist trajectories of IPD and EPD minimizers converging to it. Moreover, we give nondegeneracy results on the IPD and EPD minimizers that imply there exist optimization algorithms for smooth problems that will show fast local convergence when applied to the master problem and the subproblems. In other words, we show that IPD and EPD are fast locally convergent decomposition algorithms.

This chapter is organized as follows. In Section 3.1, we show that IPD is fast locally convergent. In Section 3.2, we show that EPD is fast locally convergent. In Section 3.3, we discuss the difficulties encountered when we try to prove global convergence for decomposition algorithms for the nonconvex OPGV such as EPD or IPD.

3.1 Local Convergence Results for IPD

Let a *nondegenerate* minimizer be one that satisfies the LICQ, SCSC, and SOSC. In this section we show that for every nondegenerate OPGV minimizer X_O^* there exists a trajectory of IPD minimizers $X_I^*(\gamma)$ such that $\lim_{\gamma\to\infty} X_I^*(\gamma) = X_O^*$. Thus X_O^* can be computed by solving the IPD master problem for a sequence of penalty parameters $\{\gamma_k\}$ such that $\lim_{k\to\infty}\gamma_k = \infty$. Moreover, we show that for each value of the penalty parameter γ, the minimizer $X_I^*(\gamma)$ is nondegenerate with respect to the IPD master problem and subproblems. This in turn implies the existence of optimization algorithms for smooth problems that will converge at a fast rate when applied to the IPD master problem and subproblems.

To prove the result, we first show how the IPD master problem can be obtained from the OPGV. Then we use the nondegeneracy of the OPGV and the implicit function theorem to show that $X_I^*(\gamma)$ is a nondegenerate minimizer.

3.1.1 IPD Master Problem Derivation

The IPD master problem is obtained from the OPGV through a sequence of three manipulations: (i) introduction of target variables, (ii) introduction of an inexact penalty function, and (iii) projection. The OPGV is

$$\begin{array}{lrcl} \min\limits_{x,y_i} & \sum\limits_{i=1}^{N} F_i(x,y_i) & & \\ s.t. & c_i(x,y_i) & \geq & 0, \quad i = 1{:}N, \end{array} \tag{3.1}$$

where $x \in \mathbb{R}^n$, $y_i \in \mathbb{R}^{n_i}$, $F_i(x,y_i) : \mathbb{R}^{n+n_i} \to \mathbb{R}$, and $c_i(x,y_i) : \mathbb{R}^{n+n_i} \to \mathbb{R}^{m_i}$.

Target Variables

The first manipulation operated on the OPGV is the introduction of the target variables z. Then, a different vector x_i is used to represent the value of the global variables at each of the systems. Compatibility constraints ($x_i = z$) are introduced to force the global variables to take the same value, equal to the target variables, for all systems. The resulting problem is the Individual System Separable Problem (ISSP):

$$\begin{array}{lrcl} \min\limits_{z,x_i,y_i} & \sum\limits_{i=1}^{N} F_i(x_i,y_i) & & \\ s.t. & c_i(x_i,y_i) & \geq & 0, \quad i = 1{:}N, \\ & x_i - z & = & 0, \quad i = 1{:}N, \end{array} \tag{3.2}$$

where $x_i, z \in \mathbb{R}^n$. The ISSP objective and constraint functions are separable with respect to the ith system variables (x_i, y_i). Clearly, (3.1) and (3.2) have the same set of minimizers.

Inexact Penalty Function

The second transformation is the introduction of quadratic penalty terms in the ISSP objective function to remove the compatibility constraints $x_i = z$. The result is the Penalty Individual System Separable Problem (PISSP):

$$\begin{array}{lll} \min\limits_{z,x_i,y_i} & \sum\limits_{i=1}^{N} \left[F_i(x_i, y_i) + \gamma \|x_i - z\|_2^2\right] & \\ s.t. & c_i(x_i, y_i) \;\geq\; 0, \quad i = 1{:}N. & \end{array} \tag{3.3}$$

These quadratic penalty functions are *inexact* because the PISSP minimizers only satisfy the compatibility constraints $x_i = z$ asymptotically as $\gamma \to \infty$. For finite γ, the PISSP minimizer is only in general an approximation to the ISSP minimizer and hence the minimizer of (3.1).

Projection

Finally, if we set the target variables to a fixed value, the PISSP breaks into N independent subproblems. The subproblem optimal-value functions can be used to formulate a master problem that only depends on the target variables. This procedure is known as *projection*. The result is the IPD problem, namely solve the following master problem for a sequence of penalty parameters $\{\gamma_k\}$ such that $\lim_{k\to\infty} \gamma_k = \infty$:

$$\min_{z} \sum_{i=1}^{N} F_i^*(\gamma, z) \ ,$$

where $F_i^*(\gamma, z)$ is the optimal-value function corresponding to the ith subproblem,

$$\begin{array}{rcl} F_i^*(\gamma, z) & = & \min\limits_{x_i,y_i} \;\; F_i(x_i, y_i) + \gamma \|x_i - z\|_2^2 \\ & & s.t. \;\; c_i(x_i, y_i) \geq 0. \end{array}$$

3.1.2 IPD Nondegeneracy

We prove the existence of a trajectory of nondegenerate IPD minimizers converging to each nondegenerate OPGV minimizer. The result is obtained in three steps. First, we show that for any nondegenerate OPGV minimizer there exists an equivalent nondegenerate ISSP minimizer. Second, we show that there exists a trajectory of PISSP minimizers that converges to each nondegenerate ISSP minimizer as we drive the penalty parameter to infinity. Finally, we show that each nondegenerate PISSP minimizer is an IPD minimizer satisfying the nondegeneracy conditions for both the IPD master problem and subproblems.

ISSP Nondegeneracy

First, we introduce some notation. The subindex O is used to distinguish variables related to the OPGV, S is used for ISSP variables, P for PISSP variables, and I for IPD variables. For a given set of vectors $\{v_i\}_{i=1}^N$, we denote the column vector

$$V = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix}$$

by $V = (v_1, v_2, \ldots, v_N)$. Likewise, for a given a set of matrices $\{M_i\}_{i=1}^N$, we denote the matrix

$$M = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_N \end{pmatrix}$$

by $M = (M_1, M_2, \ldots, M_N)$.

Definition 3.1 An OPGV point is a vector $X_O = (x, Y)$, where $x \in \mathbb{R}^n$ and $Y = (y_1, y_2, \ldots, y_N)$ with $y_i \in \mathbb{R}^{n_i}$.

Definition 3.2 For a given OPGV point $X_O = (x, Y)$, the *equivalent ISSP point* is $X_S = (x, X, Y)$, where $X \in \mathbb{R}^{N \cdot n}$ is $X = (x, x, \ldots, x)$.

Note that the equivalent ISSP point is obtained from the OPGV point by setting $x_i = x$ for $i = 1{:}N$, and $z = x$. Roughly speaking, the equivalent ISSP point is the projection of an OPGV point onto the ISSP variable space.

The transposed OPGV Jacobian at a point X_O is

$$J_O^T = \begin{pmatrix} A_1^T & A_2^T & \dots & A_N^T \\ B_1^T & & & \\ & B_2^T & & \\ & & \ddots & \\ & & & B_N^T \end{pmatrix}, \tag{3.4}$$

where $A_i = \nabla_x \hat{c}_i(x, y_i)$, $B_i = \nabla_{y_i} \hat{c}_i(x, y_i)$, and $\hat{c}$ are the constraints active at X_O. This matrix is also written as $J_O^T = (\tilde{A} \; B)^T$, where $\tilde{A} = (A_1, A_2, \dots, A_N)$ and B is the block diagonal matrix $B = \text{diag}(B_1, B_2, \dots, B_N)$. The transposed ISSP Jacobian at a point X_S is

$$J_S^T = \begin{pmatrix} & Z^T \\ A^T & I_{N \cdot n} \\ B^T & \end{pmatrix}, \tag{3.5}$$

where A is the block diagonal matrix $A = \text{diag}(A_1, A_2, \dots, A_N)$, I_k is the k-dimensional identity matrix, and Z is the matrix $Z = (-I_n, -I_n, \dots, -I_n)$.

Finally, in the remainder of this section we assume that the functions $F_i(x, y_i)$ and $c_i(x, y_i)$ are three times continuously differentiable.

We now turn to proving that if an OPGV minimizer X_O^* is nondegenerate, then its equivalent ISSP point X_S^* is a nondegenerate ISSP minimizer. We start by proving that the LICQ holds at X_S^*.

Lemma 3.3 *The LICQ holds at an OPGV point iff the LICQ holds at its equivalent ISSP point.*

Proof: Suppose there exists $\lambda_O \neq 0$ such that $J_O^T \lambda_O = 0$. Let

$$\lambda_S = \begin{pmatrix} \lambda_O \\ -A^T \lambda_O \end{pmatrix}.$$

Then λ_S is obviously nonzero and it is easy to show that $J_S^T \lambda_S = 0$. Conversely, suppose there exists

$$\lambda_S = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \neq 0$$

such that $J_S^T \lambda_S = 0$; then from (3.5) we know that

$$Z^T \lambda_2 = 0, \tag{3.6}$$

$$A^T \lambda_1 + \lambda_2 = 0, \tag{3.7}$$

$$B^T \lambda_1 = 0. \tag{3.8}$$

From (3.7) we know that $\lambda_2 = -A^T \lambda_1$. Substituting λ_2 into (3.6) we get

$$-Z^T A^T \lambda_1 = 0. \tag{3.9}$$

This in turn implies by the definitions of A^T and Z^T that

$$\tilde{A}^T \lambda_1 = 0. \tag{3.10}$$

From (3.8) and (3.10) we know that $J_O^T \lambda_1 = 0$. Moreover, $\lambda_1 \neq 0$ because otherwise $\lambda_S = 0$ from (3.7). ■

In the following lemma we show that if an OPGV minimizer satisfies the first-order KKT conditions for the OPGV, then its equivalent ISSP point satisfies the first-order KKT conditions for the ISSP.

Lemma 3.4 *The point (X_O, λ_O) is a first-order KKT point for the OPGV iff (X_S, λ_S) is a first-order KKT point for the ISSP, where X_S is the ISSP point equivalent to X_O and*

$$\lambda_S = \begin{pmatrix} \lambda_O \\ \nabla_X \sum_{i=1}^N F_i(x, y_i) - A^T \lambda_O \end{pmatrix}, \tag{3.11}$$

where

$$\nabla_X \sum_{i=1}^{N} F_i(x, y_i) = \begin{pmatrix} \nabla_x F_1(x, y_1) \\ \nabla_x F_2(x, y_2) \\ \vdots \\ \nabla_x F_N(x, y_N) \end{pmatrix}.$$

Proof: We need to show that the first-order KKT conditions A.2–A.6 are satisfied for the OPGV at X_O iff they are satisfied for the ISSP at X_S. From the definition of the equivalent ISSP point, it is obvious that the feasibility conditions A.2 and A.3 are satisfied at X_O iff they are satisfied at X_S. It remains to show that conditions A.4–A.6 are satisfied for the OPGV at X_O iff they are satisfied for the ISSP at X_S.

Assume there exists $\lambda_O \geq 0$ satisfying the complementarity condition A.4 at X_O such that

$$\begin{pmatrix} \sum_{i=1}^N \nabla_x F_i(x, y_i) \\ \nabla_Y \sum_{i=1}^N F_i(x, y_i) \end{pmatrix} = \begin{pmatrix} \tilde{A}^T \\ B^T \end{pmatrix} \lambda_O,$$

where

$$\nabla_Y \sum_{i=1}^N F_i(x, y_i) = \begin{pmatrix} \nabla_{y_1} F_1(x, y_1) \\ \nabla_{y_2} F_2(x, y_2) \\ \vdots \\ \nabla_{y_N} F_N(x, y_N) \end{pmatrix}.$$

Then it is easy to show that condition (A.6) holds at X_S with λ_S as defined in (3.11); that is,

$$\begin{pmatrix} 0 \\ \nabla_X \sum_{i=1}^N F_i(x, y_i) \\ \nabla_Y \sum_{i=1}^N F_i(x, y_i) \end{pmatrix} = \begin{pmatrix} & Z^T \\ A^T & I \\ B^T & \end{pmatrix} \lambda_S.$$

Moreover, if the nonnegativity and the complementarity conditions A.5 and A.4 hold at X_O^* for the OPGV, then they obviously also hold at (x_S, λ_S) for the ISSP.

Conversely, assume conditions A.4–A.6 are satisfied for the ISSP at X_S with $\lambda_S = (\lambda_1, \lambda_2)$. Then it is easy to prove by arguments identical to those used above that conditions A.4–A.6 are satisfied for the OPGV at X_O with $\lambda_O = \lambda_1$. ■

Definition 3.5 For a given OPGV first-order KKT point (X_O, λ_O), the *equivalent ISSP first-order KKT point* is (X_S, λ_S), where X_S is the ISSP point equivalent to X_O and λ_S is given by (3.11).

The following corollary is a consequence of Lemma 3.4.

Corollary 3.6 *The SCSC hold at a first-order KKT point for the OPGV iff the SCSC hold at its equivalent ISSP first-order KKT point.*

The following proposition shows the equivalence between the OPGV tangent cone $\mathcal{T}_O$ and the ISSP tangent cone $\mathcal{T}_S$.[1] This result is used later to prove equivalence between OPGV and ISSP in terms of SOSC.

Proposition 3.7 *For a given OPGV first-order KKT point (X_O, λ_O), there is a one-to-one correspondence between the set of vectors $\tau_O \in \mathcal{T}_O(X_O, \lambda_O)$ and the set of vectors $\tau_S \in \mathcal{T}_S(X_S, \lambda_S)$, where (X_S, λ_S) is the ISSP first-order KKT point equivalent to (X_O, λ_O).*

Proof: Given $\tau_O = (x, Y) \in \mathcal{T}_O(X_O, \lambda_O)$, we construct the ISSP vector

$$\tau_S = (x, X, Y),$$

where $X \in \mathbb{R}^{N \cdot n}$ is $X = (x, x, \ldots, x)$. Then

$$J_S \tau_S = \begin{pmatrix} & A & B \\ Z & I & \end{pmatrix} \begin{pmatrix} x \\ X \\ Y \end{pmatrix} = \begin{pmatrix} AX + BY \\ Zx + X \end{pmatrix}. \tag{3.12}$$

Moreover, it is obvious from the definitions of A, B, Z, X, and Y that

$$J_S \tau_S = \begin{pmatrix} AX + BY \\ Zx + X \end{pmatrix} = \begin{pmatrix} J_O \tau_O \\ 0 \end{pmatrix}. \tag{3.13}$$

Then from (3.13) and the definition of λ_S it follows that $\tau_S \in \mathcal{T}_S(X_S, \lambda_S)$.

Conversely, let $X = (x_1, x_2, \ldots, x_N)$ and assume $\tau_S = (x, X, Y)$ satisfies $J_S \tau_S = 0$; then

$$AX + BY = 0 \tag{3.14}$$

$$Zx + X = 0. \tag{3.15}$$

Then (3.15) implies that $X = (x, x, \ldots, x)$ and from (3.14) and the definition of λ_S by (3.11) we know that $\tau_O = (x, Y) \in \mathcal{T}_O(X_O, \lambda_O)$. ∎

[1] See A.12 for a definition of tangent cone.

Definition 3.8 Let (X_O, λ_O) be an OPGV first-order KKT point. For a given OPGV tangent cone vector $\tau_O = (x, Y) \in \mathcal{T}_O(X_O, \lambda_O)$, the *equivalent ISSP tangent cone vector* is $\tau_S = (x, X, Y)$, where $X \in \mathbb{R}^{N \cdot n}$ is $X = (x, x, \ldots, x)$.

Finally, the following theorem builds on the results given in Lemma 3.4 and Proposition 3.7.

Theorem 3.9 *An OPGV point X_O^* is a minimizer satisfying the SOSC for the OPGV iff its equivalent ISSP point X_S^* is a minimizer satisfying the SOSC for the ISSP.*

Proof: In Lemma 3.4 we showed that (X_O^*, λ_O) is a first-order KKT point for the OPGV iff its equivalent ISSP first-order KKT point (X_S^*, λ_S) satisfies the KKT conditions for the ISSP. Also, in Proposition 3.7 we showed that there is a one-to-one correspondence between vectors $\tau_O \in \mathcal{T}_O(X_O^*, \lambda_O)$ and vectors $\tau_S \in \mathcal{T}_S(X_S^*, \lambda_S)$. The proof will be complete if we show that for all $\tau_O = (x, Y) \in \mathcal{T}_O(X_O^*, \lambda_O)$,

$$\tau_O^T \nabla^2 \mathcal{L}_O(X_O^*, \lambda_O) \tau_O = \tau_S^T \nabla^2 \mathcal{L}_S(X_S^*, \lambda_S) \tau_S,$$

where τ_S is the ISSP tangent cone vector equivalent to τ_O and $\nabla^2 \mathcal{L}$ is the Hessian of the Lagrangian. Denote the OPGV Lagrange multiplier vector by $\lambda_O = ((\lambda_O)_1, (\lambda_O)_2, \ldots, (\lambda_O)_N)$, where $(\lambda_O)_i$ are the OPGV Lagrange multipliers corresponding to the ith system active constraints $\hat{c}_i(x, y_i)$. It follows that

$$\tau_S^T \nabla^2 \mathcal{L}_S(X_S^*, \lambda_S) \tau_S = (x, X, Y)^T \begin{pmatrix} 0 & & \\ & C & D \\ & D^T & E \end{pmatrix} \begin{pmatrix} x \\ X \\ Y \end{pmatrix}, \tag{3.16}$$

where C is the block diagonal matrix whose ith block is

$$\nabla_{xx}^2 F_i(x^*, y_i^*) - (\lambda_O)_i \nabla_{xx}^2 c_i(x^*, y_i^*),$$

D is the block diagonal matrix whose ith block is

$$\nabla_{x,y_i}^2 F_i(x^*, y_i^*) - (\lambda_O)_i \nabla_{x,y_i}^2 c_i(x^*, y_i^*),$$

and E is the block diagonal matrix whose ith block is

$$\nabla_{y_i,y_i}^2 F_i(x^*, y_i^*) - (\lambda_O)_i \nabla_{y_i,y_i}^2 c_i(x^*, y_i^*).$$

From (3.16) we deduce that

$$\begin{aligned}\tau_S^T \nabla^2 \mathcal{L}_S(X_S^*, \lambda_S)\tau_S &= \sum_{i=1}^N x^T \left(\nabla^2_{xx} F_i(x^*, y_i^*) - (\lambda_O)_i \nabla^2_{xx} c_i(x^*, y_i^*)\right) x \\ &+ \sum_{i=1}^N y_i^T \left(\nabla^2_{y_i,y_i} F_i(x^*, y_i^*) - (\lambda_O)_i \nabla^2_{y_i,y_i} c_i(x^*, y_i^*)\right) y_i \\ &+ 2\sum_{i=1}^N y_i^T \left(\nabla^2_{x,y_i} F_i(x_i^*, y_i^*) - (\lambda_O)_i \nabla^2_{x,y_i} c_i(x^*, y_i^*)\right) x \\ &= \tau_O^T \nabla^2 \mathcal{L}_O(X_O^*, \lambda_O)\tau_O.\end{aligned} \tag{3.17}$$

■

Theorem 3.10 *An OPGV point X_O^* is a minimizer satisfying the LICQ, SCSC, and SOSC for the OPGV iff its equivalent ISSP point X_S^* is a minimizer satisfying the LICQ, SCSC, and SOSC for the ISSP.*

Proof. The result is an immediate consequence of Lemma 3.3, Corollary 3.6, and Theorem 3.9. ■

PISSP Nondegeneracy

The following theorem follows from Theorems 14 and 17 in Fiacco and McCormick [FM68] and ensures the existence of a trajectory of nondegenerate PISSP minimizers converging to every nondegenerate ISSP minimizer.

Theorem 3.11 *If X_S^* is an ISSP minimizer satisfying the LICQ, SCSC and SOSC then, for γ sufficiently large, there exists a unique once continuously differentiable trajectory of PISSP minimizers $X_P^*(\gamma)$ satisfying the LICQ, SCSC, and SOSC such that $lim_{\gamma\to\infty} X_P^*(\gamma) = X_S^*$.*

IPD Nondegeneracy

We show that for any nondegenerate PISSP minimizer $X_P^*(\gamma)$, there exists an equivalent IPD minimizer $X_I^*(\gamma)$. Moreover, we show that $X_I^*(\gamma)$ is nondegenerate with respect to the IPD master problem and subproblems.

First we introduce some notation. Notice that PISSP points X_P and IPD points X_I have the same dimension. In particular, $X_P, X_I \in \mathbb{R}^{n_I}$, where $n_I = (N+1)n + \sum_{i=1}^N n_i$. We denote an IPD point by $X_I = (z, X, Y)$, where $z \in \mathbb{R}^n$, $X = (x_1, x_2, \ldots, x_N)$ with $x_i \in \mathbb{R}^n$, and $Y = (y_1, y_2, \ldots, y_N)$ with $y_i \in \mathbb{R}^{n_i}$. An IPD point can also be written as $X_I = P(z, (X_I)_1, (X_I)_2, \ldots, (X_I)_N)$, where $(X_I)_i = (x_i, y_i)$ and $P^T \in \mathbb{R}^{n_I \times n_I}$ is a permutation matrix that rearranges X_I so that the components corresponding to the ith system are contiguous.

Definition 3.12 A point $X_I^* = (z^*, X^*, Y^*)$ is a *semi-local IPD minimizer* if $(X_I^*)_i = (x_i^*, y_i^*)$ is a local minimizer for the ith IPD subproblem with $z = z^*$.

Definition 3.13 A point $X_I^* \in \mathbb{R}^{n_I}$ is a *strict local IPD minimizer* if: (i) X_I^* is a semi-local IPD minimizer, and (ii) there exists a neighborhood $\mathcal{N}_\epsilon(X_I^*)$ such that if $X_I \in \mathcal{N}_\epsilon(X_I^*)$ is a semi-local minimizer then

$$\sum_{i=1}^N \left[F_i(x_i^*, y_i^*) + \gamma \|x_i^* - z^*\|_2^2\right] < \sum_{i=1}^N \left[F_i(x_i, y_i) + \gamma \|x_i - z\|_2^2\right].$$

Lemma 3.14 *If X_P^* is a PISSP minimizer satisfying SOSC, then X_P^* is also a strict local IPD minimizer.*

Proof: We need to show that conditions (i) and (ii) in Definition 3.13 are satisfied at X_P^*. Assume X_P^* is a PISSP minimizer satisfying SOSC. Then there exists a neighborhood $\mathcal{N}_\epsilon(X_P^*)$ such that for all feasible points $X_P \in \mathcal{N}_\epsilon(X_P^*)$,

$$\sum_{i=1}^N \left[F_i(x_i^*, y_i^*) + \gamma \|x_i^* - z^*\|_2^2\right] < \sum_{i=1}^N \left[F_i(x_i, y_i) + \gamma \|x_i - z\|_2^2\right]. \tag{3.18}$$

In particular, for all

$$X_P = P(z^*, (X_P)_1^*, (X_P)_2^*, \ldots, (X_P)_i^* + \Delta(X_P)_i, \ldots, (X_P)_N^*)$$

such that $\|\Delta(X_P)_i^T\| \le \epsilon$ we know by (3.18) that

$$F_i(x_i^*, y_i^*) + \gamma \|x_i^* - z^*\|_2^2 < F_i(x_i^* + \Delta x_i, y_i^* + \Delta y_i) + \gamma \|x_i^* + \Delta x_i - z\|_2^2, \tag{3.19}$$

and therefore X_P^* is a semi-local IPD minimizer, i.e., condition (i) in Definition 3.13 holds at X_P^*. Also, every semi-local IPD minimizer is a feasible PISSP point. Therefore, by (3.18), we know that X_P^* satisfies condition (ii) in Definition 3.13. ■

Theorem 3.15 *If $X_P^* = (z^*, X^*, Y^*)$ is a PISSP minimizer satisfying the LICQ, SCSC, and SOSC, then $(X_P^*)_i = (x_i^*, y_i^*)$ is a minimizer satisfying the LICQ, SCSC, and SOSC for the ith IPD subproblem with $z = z^*$.*

Proof: Let λ_P be the unique Lagrange multiplier vector for the PISSP at X_P^*. Then

$$\begin{pmatrix} 2\gamma \sum_{i=1}^N (z - x_i) \\ \nabla_X \sum_{i=1}^N F_i(x_i^*, y_i^*) + 2\gamma(X + Zz) \\ \nabla_Y \sum_{i=1}^N F_i(x_i^*, y_i^*) \end{pmatrix} = \begin{pmatrix} 0 \\ A^T \\ B^T \end{pmatrix} \lambda_P. \tag{3.20}$$

From (3.20) we know that for $i = 1{:}N$,

$$\begin{pmatrix} \nabla_{x_i} F_i(x_i^*, y_i^*) + 2\gamma(x_i - z) \\ \nabla_{y_i} F_i(x_i^*, y_i^*) \end{pmatrix} = \begin{pmatrix} A_i^T \\ B_i^T \end{pmatrix} (\lambda_P)_i, \tag{3.21}$$

where $(\lambda_P)_i$ are the components in λ_P corresponding to the ith system constraints. Therefore, $((X_P^*)_i, (\lambda_P)_i)$ is a first-order KKT point for the ith IPD subproblem. Moreover, if the LICQ and SCSC hold at X_P^*, then they obviously hold also at $((X_P^*)_i, (\lambda_P)_i)$ for the ith IPD subproblem. It remains to show that the SOSC hold at $((X_P^*)_i, (\lambda_P)_i)$ for the ith IPD subproblem. The PISSP Jacobian at X_P^* is

$$J_S = \begin{pmatrix} 0 & (J_I)_1 & & & \\ 0 & & (J_I)_2 & & \\ 0 & & & \ddots & \\ 0 & & & & (J_I)_N \end{pmatrix} P^T, \tag{3.22}$$

where P^T is a permutation matrix that rearranges the columns of the PISSP Jacobian so that columns corresponding to the same system are contiguous and $(J_I)_i$ is the Jacobian of the ith IPD subproblem evaluated at $(X_P^*)_i$. From (3.22) it is clear that for any vector $(\tau_I)_i$ belonging to the ith subproblem tangent cone at $(X_P^*)_i$ we can form a vector

$$\tau_P = P(0, 0, \ldots, 0, (\tau_I)_i, 0, \ldots, 0) \tag{3.23}$$

belonging to the PISSP tangent cone at X_P^*. The PISSP Lagrangian Hessian is

$$\nabla^2 \mathcal{L}_P = P \begin{pmatrix} 0 & & & & \\ & (\nabla^2 \mathcal{L}_I)_1 & & & \\ & & (\nabla^2 \mathcal{L}_I)_2 & & \\ & & & \ddots & \\ & & & & (\nabla^2 \mathcal{L}_I)_N \end{pmatrix} P^T, \tag{3.24}$$

where $(\nabla^2 \mathcal{L}_I)_i$ is the ith subproblem Lagrangian Hessian at $(X_P^*)_i$. Because the SOSC hold at X_P^*, for any PISSP tangent cone vector τ_P we have

$$\tau_P^T (\nabla^2 \mathcal{L}_P) \tau_P > 0. \tag{3.25}$$

Then (3.23)–(3.25) imply

$$(\tau_I)_i^T (\nabla^2 \mathcal{L}_I)_i (\tau_I)_i > 0. \tag{3.26}$$

■

Lemma 3.16 *If the functions F_i and c_i are three times continuously differentiable and the PISSP minimizer $X_P^* = P(z^*, (X_P^*)_1, (X_P^*)_2, \ldots, (X_P^*)_N)$ satisfies the LICQ, SCSC, and SOSC, then the IPD master problem objective*

$$F^*(\gamma, z) = \sum_{i=1}^{N} F_i^*(\gamma, z)$$

can be defined as a twice continuously differentiable function in a neighborhood $\mathcal{N}_\epsilon(z^*)$ with $\epsilon > 0$.

Proof: From Theorem 3.15 we know that if X_P^* is a nondegenerate PISSP minimizer, then $(X_P^*)_i$ is a minimizer satisfying the LICQ, SCSC, and SOSC for the ith IPD subproblem with $z = z^*$. Therefore, by the implicit function theorem and Theorem 6 in [FM68] we know that if F_i and c_i are three times continuously differentiable, then there exists a unique twice continuously differentiable trajectory of minimizers to the ith IPD subproblem $(X_P^*)_i(z)$ defined in a neighborhood $\mathcal{N}_{\epsilon_i}(z^*)$ with $\epsilon_i > 0$. The $(X_P^*)_i(z)$ define in turn a unique twice continuously differentiable function $F^*(\gamma, z) = \sum_{i=1}^{N} F_i^*(\gamma, z)$ on $\mathcal{N}_\epsilon(z^*)$, where $\epsilon = \min(\epsilon_1, \epsilon_1, \ldots, \epsilon_N)$. ■

Definition 3.17 Let $X_P^* = P(z^*, (X_P^*)_1, (X_P^*)_2, \dots, (X_P^*)_N)$ be a PISSP point such that for $i = 1{:}N$ the vector $(X_P^*)_i$ is a minimizer satisfying the LICQ, SCSC, and SOSC for the ith IPD subproblem with $z = z^*$. Then z^* is a *strict IPD master problem minimizer* if there exists a neighborhood $\mathcal{N}_\epsilon(z^*)$ with $\epsilon > 0$ such that $\forall z \in \mathcal{N}_\epsilon(z^*)$ we have $F^*(\gamma, z) > F^*(\gamma, z^*)$, where F^* is the twice continuously differentiable master problem objective given by Theorem 3.16.

Theorem 3.18 *If $X_P^* = P(z^*, (X_P^*)_1, (X_P^*)_2, \dots, (X_P^*)_N)$ is a PISSP minimizer satisfying the LICQ, SCSC, and SOSC, then z^* is a strict IPD master problem minimizer satisfying the SOSC.*

Proof: From Theorem 3.15 and Lemma 3.16 we know that there exist twice continuously differentiable trajectories of IPD subproblem minimizers $(X_P^*)_i(z)$ defined in a neighborhood $\mathcal{N}_{\epsilon_1}(z^*)$. Then, by the differentiability of $(X_P^*)_i(z)$ we know that for all $\epsilon_2 > 0$ we can always find $\epsilon_3 > 0$ such that $\epsilon_3 < \epsilon_1$ and for all $z \in \mathcal{N}_{\epsilon_3}(z^*)$,

$$\overline{X}_P(z) = (z, (X_P^*)_1(z), (X_P^*)_2(z), \dots, (X_P^*)_N(z)) \in \mathcal{N}_{\epsilon_2}(X_P^*). \tag{3.27}$$

Because we know by Lemma 3.14 that X_P^* is a strict IPD local minimizer, (3.27) implies that there exists $\epsilon_4 < \epsilon_3$ such that $F^*(\gamma, z) > F^*(\gamma, z^*)$ for all $z \in \mathcal{N}_{\epsilon_4}(z^*)$, where F^* is the master problem objective given by Lemma 3.16. Thus z^* is a strict IPD master problem minimizer. It only remains to show that the SOSC hold at z^* for the IPD master problem. It suffices to show that for all $v \neq 0$,

$$\frac{d^2 F^*(\gamma, z^* + rv)}{dr^2} > 0.$$

But notice that

$$F^*(\gamma, z^* + rv) = F_P(\hat{X}_P(r)),$$

where F_P is the PISSP objective function and $\hat{X}_P(r) = \overline{X}_P(z^* + rv)$. Moreover, because (X_P^*, λ^*) satisfies the complementarity conditions for the PISSP and the implicit function theorem guarantees that the active set remains fixed at $X_P^*(r)$ for r small, we know that

$$F_P(\hat{X}_P(r)) = \mathcal{L}_P(\hat{X}_P(r), \lambda^*),$$

where $\mathcal{L}_P$ is the Lagrangian function. Therefore,

$$\frac{d^2F^*(\gamma, z^* + rv)}{dr^2} = \frac{d^2\mathcal{L}_P(\hat{X}_P(r), \lambda^*)}{dr^2}.$$

The first derivative of the PISSP Lagrangian function with respect to r is

$$\frac{d\mathcal{L}_P(\hat{X}_P(r), \lambda^*)}{dr} = \nabla_X \mathcal{L}_P(\hat{X}_P(r), \lambda^*) \frac{d\hat{X}_P(r)}{dr},$$

and the second derivative is

$$\begin{aligned} \frac{d^2\mathcal{L}_P(\hat{X}_P(r), \lambda^*)}{dr^2} =& \frac{d\hat{X}_P(r)}{dr}^T \nabla^2_{XX} \mathcal{L}_P(\hat{X}_P(r), \lambda^*) \frac{d\hat{X}_P(r)}{dr} \\ &+ \nabla_x \mathcal{L}_P(\hat{X}_P(r), \lambda^*) \frac{d^2\hat{X}_P(r)}{dr^2}. \end{aligned} \tag{3.28}$$

Evaluating (3.28) at $r = 0$ and because $\hat{X}_P(0)$ is a PISSP stationary point we get

$$\left.\frac{d^2F^*(\gamma, z^* + rv)}{dr^2}\right|_{r=0} = \frac{d\hat{X}_P(0)}{dr}^T \nabla^2_{XX} \mathcal{L}_P(\hat{X}_P(0), \lambda^*) \frac{d\hat{X}_P(0)}{dr}. \tag{3.29}$$

Because $\hat{X}_P(r)$ is twice continuously differentiable, $\hat{X}_P(r)$ remains PISSP feasible for r small, and the LICQ holds for the PISSP at X_P^*, we know that $\frac{d\hat{X}_P(0)}{dr}$ belongs to the PISSP tangent cone. Moreover, because (X_P^*, λ^*) satisfies the SOSC, (3.29) implies that

$$\left.\frac{d^2F^*(\gamma, z^* + rv)}{dr^2}\right|_{r=0} = \frac{d\hat{X}_P(0)}{dr}^T \nabla^2_{XX} \mathcal{L}_P(\hat{X}_P(0), \lambda^*) \frac{d\hat{X}_P(0)}{dr} > 0.$$

■

3.2 Local Convergence Results for EPD

In this section we show that EPD is a fast locally convergent decomposition algorithm. In particular, we show that for any nondegenerate OPGV minimizer X_O^*, there exists a trajectory of nondegenerate EPD minimizers $X_E^*(\mu)$ such that $\lim_{\mu \to 0} X_E^*(\mu) = X_O^*$.

The proof parallels the one given for IPD, but the notation is complicated by the presence of the elastic variables R, S, and T. We first show how the EPD master problem can be obtained from the OPGV. Then we use the nondegeneracy of the OPGV and the implicit function theorem to show that $X_E^*(\mu)$ is a nondegenerate minimizer.

3.2.1 EPD Master Problem Derivation

The EPD master problem is obtained from the OPGV through a sequence of three manipulations: (i) introduction of an exact penalty function, (ii) introduction of barrier terms, and (iii) projection.

Exact Penalty Function

The first manipulation operated on the OPGV is the introduction of the target variables z. Then a different vector x_i is used to represent the value of the global variables at each system. The l_1 exact penalty function $\gamma\|x_i - z\|_1 = \gamma \sum_{j=1}^{n} |x_{ij} - z_j|$ is used to force the global variables to take the same value, equal to the target variables, for all systems. To avoid the nonsmoothness of the absolute value function, rather than using the exact penalty function explicitly, we introduce elastic variables s_i and t_i. The l_1 exact penalty function can then be computed as $\gamma\|x_i - z\|_1 = \gamma e^T(s_i + t_i)$. Slack variables r_i are also introduced so that only equality constraints and nonnegativity bounds remain. The resulting problem is the Individual System Feasible Problem (ISFP):

$$\begin{array}{lrcl} \min\limits_{z,x_i,y_i,r_i,s_i,t_i} & \sum\limits_{i=1}^{N} \left[F_i(x_i, y_i) + \gamma e^T(s_i + t_i)\right] & & \\ s.t. & c_i(x_i, y_i) - r_i & = & 0, \quad i = 1{:}N, \\ & x_i + s_i - t_i & = & z, \quad i = 1{:}N, \\ & r_i, s_i, t_i & \geq & 0, \quad i = 1{:}N, \end{array} \tag{3.30}$$

where γ is the penalty parameter, $e \in \mathbb{R}^n$ is the vector of ones, $r_i \in \mathbb{R}^{m_i}$ and $s_i, t_i, z \in \mathbb{R}^n$. The term Individual System Feasible refers to the fact that, if the original OPGV is feasible, then the ISFP is feasible for any value of z.

Barrier Terms

The second transformation is the introduction of barrier terms in the ISFP objective function to remove the nonnegativity constraints. The result is the Barrier Individual

System Feasible Problem (BISFP):

$$\begin{array}{lrcl} \min\limits_{z,x_i,y_i,r_i,s_i,t_i} & \sum\limits_{i=1}^{N} \left[(F_i(x_i,y_i) + \gamma e^T(t_i+s_i) - \mu\phi(r_i,s_i,t_i)) \right] & & \\ s.t. & c_i(x_i,y_i) - r_i & = & 0, \quad i = 1{:}N, \\ & x_i + s_i - t_i & = & z, \quad i = 1{:}N, \end{array} \tag{3.31}$$

where μ is the barrier parameter and the barrier function $\phi(r_i,s_i,t_i) = \sum_{j=1}^{m_i} \log r_{ij} + \sum_{j=1}^{n} (\log s_{ij} + \log t_{ij})$. Notice that the BISFP is an equality constrained problem and therefore the SCSC hold at any KKT point.

Projection

Finally, if we project the BISFP onto the target variable space we get the EPD problem. We solve the following master problem for a sequence of barrier parameters $\{\mu_k\}$ such that $\lim_{k\to\infty} \mu_k = 0$:

$$\min_{z} \quad \sum_{i=1}^{N} F_i^*(\mu, z) \ ,$$

where $F_i^*(\mu, z)$ is the optimal-value function corresponding to the ith subproblem,

$$\begin{array}{lrcl} \min\limits_{x_i,y_i,r_i,s_i,t_i} & F_i(x_i,y_i) + \gamma e^T(s_i+t_i) - \mu\phi(r_i,s_i,t_i) & & \\ s.t. & c_i(x_i,y_i) - r_i & = & 0, \\ & x_i + s_i - t_i & = & z. \end{array}$$

3.2.2 EPD Nondegeneracy

We prove the existence of a trajectory of nondegenerate EPD minimizers converging to each nondegenerate OPGV minimizer. The result is obtained in three steps. First, we show that for any nondegenerate OPGV minimizer there exists an equivalent nondegenerate ISFP minimizer. Second, we show that there exists a trajectory of BISFP minimizers that converges to each nondegenerate ISFP minimizer as we drive the barrier parameter to zero. Finally, we show that each nondegenerate BISFP minimizer is an EPD minimizer satisfying the nondegeneracy conditions for the EPD master problem and subproblems.

ISFP Nondegeneracy

First, we introduce some notation. The subindex O is used to distinguish variables related to the OPGV, F is used for ISFP variables, B for BISFP variables, and E for EPD variables.

Definition 3.19 An OPGV point is a vector $X_O = (x, Y)$, where $x \in \mathbb{R}^n$ and $Y = (y_1, y_2, \ldots, y_N)$ with $y_i \in \mathbb{R}^{n_i}$.

Definition 3.20 For a given OPGV point $X_O = (x, Y)$, the *equivalent ISFP point* is $X_F = (x, X, Y, R, S, T)$, where $X \in \mathbb{R}^{N \cdot n}$ is $X = (x, x, \ldots, x)$, $R \in \mathbb{R}^{\sum_{i=1}^N m_i}$ is $R = (c_1(x_1, y_1), c_2(x_2, y_2), \ldots, c_3(x_3, y_3))$, and $S, T \in \mathbb{R}^{N \cdot n}$ are zero vectors.

Note that the equivalent ISFP point is obtained from the OPGV point by setting $r_i = c_i(x, y_i)$, $x_i = x$, and $s_i, t_i = 0$ for $i = 1{:}N$, and $z = x$. Roughly speaking, the equivalent ISFP point is the projection of an OPGV point onto the ISFP variable space.

Because the ISFP constraints $c_i(x_i, y_i) - r_i = 0$ are equality constraints, they are active at all feasible points. However, for the sake of clarity, we only consider as active those constraints $c_{ij}(x_i, y_i) - r_{ij} = 0$ for which $r_{ij} = 0$. There is no loss of generality in doing so because when $r_{ij} > 0$, the constraint $c_{ij}(x, y_i) - r_{ij} = 0$ plays no role in the nondegeneracy conditions because its gradient is linearly independent with respect to the other active constraint gradients and its corresponding Lagrange multiplier is zero. Hence, the transposed ISFP Jacobian at X_F is

$$J_F^T = \begin{pmatrix} & Z^T & & & \\ A^T & I_{N \cdot n} & & & \\ B^T & & & & \\ -I_{\hat{m}} & & I_{\hat{m}} & & \\ & I_{N \cdot n} & & I_{N \cdot n} & \\ & -I_{N \cdot n} & & & I_{N \cdot n} \end{pmatrix}, \tag{3.32}$$

where A is the block diagonal matrix $A = \operatorname{diag}(A_1, A_2, \ldots, A_N)$, I_k is the k-dimensional identity matrix, Z is the matrix $Z = (-I_n, -I_n, \ldots, -I_n)$, $\hat{m} = \sum_{i=1}^N \hat{m}_i$, and $\hat{m}_i$ is the number of active constraints in the ith system.

Finally, in the remainder of this section we assume that the functions F_i and c_i are three times continuously differentiable.

We now turn to proving that provided an OPGV minimizer X_O^* is nondegenerate, its equivalent ISFP point X_F^* is a nondegenerate ISFP minimizer. We start by proving that the LICQ holds at X_F^*.

Lemma 3.21 *The LICQ holds at an OPGV point iff the LICQ holds at its equivalent ISFP point.*

Proof: Suppose there exists $\lambda_O \neq 0$ such that $J_O^T \lambda_O = 0$. Let

$$\lambda_F = \begin{pmatrix} \lambda_O \\ -A^T \lambda_O \\ \lambda_O \\ A^T \lambda_O \\ -A^T \lambda_O \end{pmatrix}.$$

Then λ_F is obviously nonzero and it is easy to show that $J_F^T \lambda_F = 0$. Conversely suppose there exists

$$\lambda_F = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{pmatrix} \neq 0$$

such that $J_F^T \lambda_F = 0$. Then from (3.32) we know that

$$Z^T \lambda_2 = 0, \tag{3.33}$$

$$A^T \lambda_1 + \lambda_2 = 0, \tag{3.34}$$

$$B^T \lambda_1 = 0, \tag{3.35}$$

$$-\lambda_1 + \lambda_3 = 0, \tag{3.36}$$

$$\lambda_2 + \lambda_4 = 0, \tag{3.37}$$

$$-\lambda_2 + \lambda_5 = 0. \tag{3.38}$$

From (3.34) we know that $\lambda_2 = -A^T\lambda_1$. Substituting λ_2 into (3.33) we get

$$-Z^T A^T \lambda_1 = 0. \tag{3.39}$$

This in turn implies by the definitions of A^T and Z^T that

$$\tilde{A}^T \lambda_1 = 0. \tag{3.40}$$

From (3.35) and (3.40) we know that $J_O^T\lambda_1 = 0$. Moreover, $\lambda_1 \neq 0$ because otherwise $\lambda_F = 0$ from (3.34), (3.36), (3.37), and (3.38). ∎

In the following lemma we show that an OPGV point satisfies the first-order KKT conditions for the OPGV if and only if its equivalent ISFP point satisfies the first-order KKT conditions for the ISFP.

Lemma 3.22 *Provided $\gamma \geq \|\nabla_X \sum_{i=1}^N F_i(x, y_i) - A^T\lambda_O\|_\infty$, where*

$$\nabla_X \sum_{i=1}^N F_i(x, y_i) = \begin{pmatrix} \nabla_x F_1(x, y_1) \\ \nabla_x F_2(x, y_2) \\ \vdots \\ \nabla_x F_N(x, y_N) \end{pmatrix},$$

(X_O, λ_O) is a first-order KKT point for the OPGV iff (X_F, λ_F) is a first-order KKT point for the ISFP, where X_F is the ISFP point equivalent to X_O and

$$\lambda_F = \begin{pmatrix} \lambda_O \\ \nabla_X \sum_{i=1}^N F_i(x, y_i) - A^T\lambda_O \\ \lambda_O \\ \gamma e - (\nabla_X \sum_{i=1}^N F_i(x, y_i) - A^T\lambda_O) \\ \gamma e + (\nabla_X \sum_{i=1}^N F_i(x, y_i) - A^T\lambda_O) \end{pmatrix}. \tag{3.41}$$

Proof: We need to show that the first-order KKT conditions A.2–A.6 are satisfied for the OPGV at X_O iff they are satisfied for the ISFP at X_F. From the definition of the equivalent ISFP point, it is obvious that the feasibility conditions A.2 and A.3 are satisfied at X_O iff they are satisfied at X_F. It remains to show that conditions A.4–A.6 are satisfied for the OPGV at X_O iff they are satisfied for the ISFP at X_F.

Assume there exists $\lambda_O \geq 0$ satisfying the complementarity condition A.4 at X_O such that

$$\begin{pmatrix} \sum_{i=1}^N \nabla_x F_i(x, y_i) \\ \nabla_Y \sum_{i=1}^N F_i(x, y_i) \end{pmatrix} = \begin{pmatrix} \tilde{A}^T \\ B^T \end{pmatrix} \lambda_O,$$

where

$$\nabla_Y \sum_{i=1}^N F_i(x, y_i) = \begin{pmatrix} \nabla_{y_1} F_1(x, y_1) \\ \nabla_{y_2} F_2(x, y_2) \\ \vdots \\ \nabla_{y_N} F_N(x, y_N) \end{pmatrix}.$$

Then it is easy to show that condition A.6 holds at X_F with λ_F as defined in (3.41); that is,

$$\begin{pmatrix} 0 \\ \nabla_X \sum_{i=1}^N F_i(x, y_i) \\ \nabla_Y \sum_{i=1}^N F_i(x, y_i) \\ 0 \\ \gamma e \\ \gamma e \end{pmatrix} = \begin{pmatrix} & Z^T & & & \\ A^T & I & & & \\ B^T & & & & \\ -I & & I & & \\ & I & & I & \\ & -I & & & I \end{pmatrix} \lambda_F.$$

Moreover, if $\gamma \geq \|\nabla_X \sum_{i=1}^N F_i(x, y_i) - A^T \lambda_O\|_\infty$, then the non-negativity conditions A.5 hold at (x_F, λ_F). The complementarity conditions A.4 are customarily satisfied for the ISFP at X_F because all of the nonnegativity constraints are active at the equivalent ISFP point.

Conversely, assume conditions A.4–A.6 are satisfied for the ISFP at X_F with $\lambda_F = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$. Then it is easy to prove by arguments identical to those used above that conditions A.4–A.6 are satisfied for the OPGV at X_O with $\lambda_O = \lambda_1$.
■

Definition 3.23 For a given OPGV first-order KKT point (X_O, λ_O), the *equivalent ISFP first-order KKT point* is (X_F, λ_F), where X_F is the ISFP point equivalent to X_O and λ_F is given by (3.41).

The following corollary is a consequence of Lemma 3.22.

Corollary 3.24 *Provided* $\gamma > \|\nabla_X \sum_{i=1}^N F_i(x, y_i) - A^T \lambda_O\|_\infty$, *the SCSC hold at a first-order KKT point for the OPGV iff the SCSC hold at its equivalent ISFP first-order KKT point.*

The following proposition shows the equivalence between the OPGV tangent cone $\mathcal{T}_O$ and the ISFP tangent cone $\mathcal{T}_F$.[2] This result is used later to prove equivalence between OPGV and ISFP in terms of SOSC.

Proposition 3.25 *For a given OPGV first-order KKT point* (X_O, λ_O), *there is a one-to-one correspondence between the set of vectors* $\tau_O \in \mathcal{T}_O(X_O, \lambda_O)$ *and the set of vectors* $\tau_F \in \mathcal{T}_F(X_F, \lambda_F)$, *where* (X_F, λ_F) *is the ISFP first-order KKT point equivalent to* (X_O, λ_O).

Proof: Given $\tau_O = (x, Y) \in \mathcal{T}_O(X_O, \lambda_O)$, we construct the following ISFP vector

$$\tau_F = (x, X, Y, (AX + BY), 0, 0),$$

where $X \in \mathbb{R}^{N \cdot n}$ is $X = (x, x, \dots, x)$. Then

$$J_F \tau_F = \begin{pmatrix} & A & B & -I & & \\ Z & I & & & I & -I \\ & & & I & & \\ & & & & I & \\ & & & & & I \end{pmatrix} \begin{pmatrix} x \\ X \\ Y \\ AX + BY \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ Zx + X \\ AX + BY \\ 0 \\ 0 \end{pmatrix}. \tag{3.42}$$

Moreover, it is obvious from the definitions of A, B, Z, X, and Y that

$$J_F \tau_F = \begin{pmatrix} 0 \\ Zx + X \\ AX + BY \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ J_O \tau_O \\ 0 \\ 0 \end{pmatrix}. \tag{3.43}$$

Then from (3.43) and the definition of λ_F, it follows that $\tau_F \in \mathcal{T}_F(X_F, \lambda_F)$.

[2]See A.12 for a definition of tangent cone.

Conversely, assume $\tau_F = (x, X, Y, R, S, T)$ with $X = (x_1, x_2, \ldots, x_N)$ satisfies $J_F \tau_F = 0$. Then

$$AX + BY - R = 0 \tag{3.44}$$

$$Zx + X + S - T = 0 \tag{3.45}$$

$$R = 0 \tag{3.46}$$

$$S = 0 \tag{3.47}$$

$$T = 0. \tag{3.48}$$

From (3.47)–(3.48), $S = T = 0$. Hence from (3.45) we know that

$$X = (x, x, \ldots, x). \tag{3.49}$$

Consider the OPGV vector $\tau_O = (x, Y)$. We have $J_O^T \tau_O = \tilde{A}x + BY$, and then (3.44), (3.46), (3.49), and the definition of λ_F imply that $\tau_O \in \mathcal{T}_O(X_O, \lambda_O)$. ■

Definition 3.26 For a given OPGV first-order KKT point (X_O, λ_O) and a tangent cone vector $\tau_O = (x, Y) \in \mathcal{T}_O(X_O, \lambda_O)$, the ISFP tangent cone vector equivalent to τ_O is $\tau_F = (x, X, Y, (AX + BY), 0, 0)$, where $X \in \mathbb{R}^{N \cdot n}$ is $X = (x, x, \ldots, x)$.

Finally, the following theorem builds on Lemma 3.22 and Proposition 3.25.

Theorem 3.27 *An OPGV point X_O^* is a minimizer satisfying the SOSC for the OPGV iff its equivalent ISFP point X_F^* is a minimizer satisfying the SOSC for the ISFP.*

Proof: In Lemma 3.22 we showed that a first-order KKT point (X_O^*, λ_O) satisfies the KKT conditions iff its equivalent ISFP first-order KKT point (X_F^*, λ_F) satisfies the KKT conditions for the ISFP. Also, in Proposition 3.25, we showed that there is a one-to-one correspondence between vectors $\tau_O \in \mathcal{T}_O(X_O^*, \lambda_O)$ and vectors $\tau_F \in \mathcal{T}_F(X_F^*, \lambda_F)$. The proof will be complete if we show that for all $\tau_O = (x, Y) \in \mathcal{T}_O(X_O^*, \lambda_O)$ we have

$$\tau_O^T \nabla^2 \mathcal{L}_O(X_O^*, \lambda_O) \tau_O = \tau_F^T \nabla^2 \mathcal{L}_F(X_F^*, \lambda_F) \tau_F,$$

where τ_F is the ISFP tangent cone vector equivalent to τ_O and $\nabla^2\mathcal{L}$ is the Hessian of the Lagrangian. Denote the OPGV Lagrange multiplier vector by $\lambda_O = ((\lambda_O)_1, (\lambda_O)_2, \ldots, (\lambda_O)_N)$, where $(\lambda_O)_i$ are the OPGV Lagrange multipliers corresponding to the ith system active constraints $\hat{c}_i(x, y_i)$. It follows that the quantity $\tau_F^T \nabla^2 \mathcal{L}_F(X_F^*, \lambda_F)\tau_F$ is

$$(x, X, Y, (AX+BY), 0, 0)^T \begin{pmatrix} 0 & & & & & \\ & C & D & & & \\ & D^T & E & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & 0 \end{pmatrix} \begin{pmatrix} x \\ X \\ Y \\ AX+BY \\ 0 \\ 0 \end{pmatrix}, \tag{3.50}$$

where C is the block-diagonal matrix whose ith block is

$$\nabla^2_{xx} F_i(x^*, y_i^*) - (\lambda_O)_i \nabla^2_{xx} c_i(x^*, y_i^*),$$

D is the block-diagonal matrix whose ith block is

$$\nabla^2_{x,y_i} F_i(x^*, y_i^*) - (\lambda_O)_i \nabla^2_{x,y_i} c_i(x^*, y_i^*),$$

and E is the block-diagonal matrix whose ith block is

$$\nabla^2_{y_i,y_i} F_i(x^*, y_i^*) - (\lambda_O)_i \nabla^2_{y_i,y_i} c_i(x^*, y_i^*).$$

From (3.50) we deduce that

$$\begin{aligned} \tau_F^T \nabla^2 \mathcal{L}_F(X_F^*, \lambda_F)\tau_F &= \sum_{i=1}^N x^T \left(\nabla^2_{xx} F_i(x^*, y_i^*) - (\lambda_O)_i \nabla^2_{xx} c_i(x^*, y_i^*)\right) x \\ &+ \sum_{i=1}^N y_i^T \left(\nabla^2_{y_i,y_i} F_i(x^*, y_i^*) - (\lambda_O)_i \nabla^2_{y_i,y_i} c_i(x^*, y_i^*)\right) y_i \\ &+ 2\sum_{i=1}^N y_i^T \left(\nabla^2_{x,y_i} F_i(x_i^*, y_i^*) - (\lambda_O)_i \nabla^2_{x,y_i} c_i(x^*, y_i^*)\right) x \\ &= \tau_O^T \nabla^2 \mathcal{L}_O(X_O^*, \lambda_O)\tau_O. \end{aligned} \tag{3.51}$$

■

Theorem 3.28 *An OPGV point X_O^* is a minimizer satisfying the LICQ, SCSC, and SOSC for the OPGV iff its equivalent ISFP point X_F^* is a minimizer satisfying the LICQ, SCSC, and SOSC for the ISFP.*

Proof. The result is an immediate consequence of Lemma 3.21, Corollary 3.24, and Theorem 3.27. ■

BISFP Nondegeneracy

The following theorem follows from Theorems 14 and 17 in Fiacco and McCormick [FM68] and ensures the existence of a trajectory of nondegenerate BISFP minimizers converging to every nondegenerate ISFP minimizer.

Theorem 3.29 *If X_F^* is an ISFP minimizer satisfying the LICQ, SCSC and SOSC, then there is a positive neighborhood of $\mu = 0$ for which there exists a unique once continuously differentiable trajectory $X_B^*(\mu)$ of BISFP minimizers satisfying the LICQ, SCSC, and SOSC such that $lim_{\mu \to 0} X_B^*(\mu) = X_F^*$.*

EPD Nondegeneracy

We show that for any nondegenerate BISFP minimizer $X_B^*(\mu)$ there exists an equivalent EPD minimizer $X_E^*(\mu)$. Moreover, we show that $X_E^*(\mu)$ is nondegenerate with respect to the EPD master problem and subproblems.

First we introduce notation. Note that BISFP points X_B and EPD points X_E have the same dimension. In particular $X_B, X_E \in \mathbb{R}^{n_E}$, where $n_E = (3N+1)n + \sum_{i=1}^N n_i + \sum_{i=1}^N m_i$. We may write an EPD point as $X_E = (z, X, Y, R, S, T)$, where $X = (x_1, x_2, \ldots, x_N)$ with $x_i \in \mathbb{R}^n$, $Y = (y_1, y_2, \ldots, y_N)$ with $y_i \in \mathbb{R}^{n_i}$, $R = (r_1, r_2, \ldots, r_N)$ with $r_i \in \mathbb{R}^{m_i}$, $S = (s_1, s_2, \ldots, s_N)$ with $s_i \in \mathbb{R}^n$, and $T = (t_1, t_2, \ldots, t_N)$ with $t_i \in \mathbb{R}^n$. An EPD point can also be written as

$$X_E = P(z, (X_E)_1, (X_E)_2, \ldots, (X_E)_N),$$

where $(X_E)_i = (x_i, y_i, r_i, s_i, t_i)$ and $P^T \in \mathbb{R}^{n_E \times n_E}$ is a permutation matrix that rearranges X_E so that the components corresponding to the ith system are contiguous.

Definition 3.30 A point $X_E^* = (z^*, X^*, Y^*, R^*, S^*, T^*)$ is a *semi-local EPD minimizer* if $(X_E^*)_i = (x_i^*, y_i^*, r_i^*, s_i^*, t_i^*)$ is a local minimizer for the ith EPD subproblem with $z = z^*$.

Definition 3.31 A point $X_E^* \in \mathbb{R}^{n_E}$ is a strict local EPD minimizer if: (i) X_E^* is a semi-local EPD minimizer, and (ii) there exists a neighborhood $\mathcal{N}_\epsilon(X_E^*)$ such that if $X_E \in \mathcal{N}_\epsilon(X_E^*)$ is a semi-local minimizer, then

$$\sum_{i=1}^{N} \left[F_i(x_i^*, y_i^*) + \gamma e^T(t_i^* + s_i^*) - \mu\phi(r_i^*, s_i^*, t_i^*)\right] < \sum_{i=1}^{N} \left[F_i(x_i, y_i) + \gamma e^T(t_i + s_i) - \mu\phi(r_i, s_i, t_i)\right]. \quad (3.52)$$

Lemma 3.32 *If X_B^* is a BISFP minimizer satisfying SOSC, then X_B^* is also a strict local EPD minimizer.*

Proof: We need to show that conditions (i) and (ii) in Definition 3.31 are satisfied at X_B^*. Assume X_B^* is a BISFP minimizer satisfying SOSC. Then there exists a neighborhood $\mathcal{N}_\epsilon(X_B^*)$ such that for all feasible points $X_B \in \mathcal{N}_\epsilon(X_B^*)$,

$$\sum_{i=1}^{N} \left[F_i(x_i^*, y_i^*) + \gamma e^T(t_i^* + s_i^*) - \mu\phi(s_i^*, t_i^*, r_i^*)\right] < \sum_{i=1}^{N} \left[F_i(x_i, y_i) + \gamma e^T(t_i + s_i) - \mu\phi(r_i, s_i, t_i)\right]. \quad (3.53)$$

In particular, for all

$$X_B = P(z^*, (X_B)_1^*, (X_B)_2^*, \ldots, (X_B)_i^* + \Delta(X_B)_i^T, \ldots, (X_B)_N^*)$$

such that $\|\Delta(X_B)_i^T\| \le \epsilon$ we know by (3.53) that

$$F_i(x_i^*, y_i^*) + \gamma e^T(t_i^* + s_i^*) - \mu\phi(r_i^*, s_i^*, t_i^*) < F_i(x_i^* + \Delta x_i, y_i^* + \Delta y_i) + \gamma e^T(t_i^* + \Delta t_i + s_i^* + \Delta s_i) - \mu\phi(r_i^* + \Delta r_i, s_i^* + \Delta x_i, t_i^* + \Delta t_i), \quad (3.54)$$

and therefore X_B^* is a semi-local EPD minimizer; that is, condition (i) in Definition 3.31 holds at X_B^*. Also, every semi-local minimum for problem EPD is a feasible

BISFP point. Therefore, by (3.53), we know that X_B^* satisfies condition (ii) in Definition 3.31. ■

Theorem 3.33 *If $X_B^* = (z^*, X^*, Y^*, R^*, S^*, T^*)$ is a minimizer satisfying the LICQ, SCSC, and SOSC for the BISFP, then $(X_B^*)_i = (x_i^*, y_i^*, r_i^*, s_i^*, t_i^*)$ is a minimizer satisfying the LICQ, SCSC, and SOSC for the ith EPD subproblem with $z = z^*$.*

Proof: Let λ_B be the unique Lagrange multiplier vector for the BISFP at X_B^*. Then

$$\begin{pmatrix} 0 \\ \nabla_X \sum_{i=1}^N F_i(x_i^*, y_i^*) \\ \nabla_Y \sum_{i=1}^N F_i(x_i^*, y_i^*) \\ \mu \hat{R}^{-2} \\ \gamma e + \mu \hat{S}^{-2} \\ \gamma e + \mu \hat{T}^{-2} \end{pmatrix} = \begin{pmatrix} & Z^T \\ A^T & I_{N\cdot n} \\ B^T & \\ -I_m & \\ & I_{N\cdot n} \\ & -I_{N\cdot n} \end{pmatrix} \lambda_B, \qquad (3.55)$$

where $\hat{R} = \mathrm{diag}(R^*)$, $\hat{S} = \mathrm{diag}(S^*)$, $\hat{T} = \mathrm{diag}(T^*)$. From (3.55) we know that for $i = 1{:}N$,

$$\begin{pmatrix} \nabla_{x_i} F_i(x_i^*, y_i^*) \\ \nabla_{y_i} F_i(x_i^*, y_i^*) \\ \mu \hat{r}_i^{-2} \\ \gamma e + \mu \hat{s}_i^{-2} \\ \gamma e + \mu \hat{t}_i^{-2} \end{pmatrix} = \begin{pmatrix} A_i^T & I_n \\ B_i^T & \\ -I_{m_i} & \\ & I_n \\ & -I_n \end{pmatrix} (\lambda_B)_i, \qquad (3.56)$$

where $\hat{r}_i = \mathrm{diag}(r_i^*)$, $\hat{s}_i = \mathrm{diag}(s_i^*)$, $\hat{t}_i = \mathrm{diag}(t_i^*)$, and $(\lambda_B)_i$ are those components in the Lagrange multiplier vector λ_B corresponding to the ith system constraints. Therefore, $((X_B^*)_i, (\lambda_B)_i)$ is a first-order KKT point for the ith EPD subproblem. Moreover, the EPD constraints corresponding to the ith EPD subproblem obviously satisfy LICQ at any point. Furthermore, the SCSC always hold because the ith subproblem is an equality constrained problem. It remains to show that SOSC hold at $((X_B^*)_i, (\lambda_B)_i)$ for the ith EPD subproblem. The BISFP Jacobian at X_B^* is

$$J_F = \begin{pmatrix} Z_1 & (J_E)_1 & & & \\ Z_2 & & (J_E)_2 & & \\ Z_3 & & & \ddots & \\ Z_4 & & & & (J_E)_N \end{pmatrix} P^T, \tag{3.57}$$

where P is a permutation matrix that rearranges the columns of the BISFP Jacobian so that columns corresponding to the same system are contiguous, $Z_i = (0_{m_i \times m_i}, I_n)$ and $(J_E)_i$ is the Jacobian of the ith EPD subproblem evaluated at $(X_B^*)_i$. From (3.57) it is clear that for any vector $(\tau_E)_i$ belonging to the ith subproblem tangent cone at $(X_B^*)_i$ we can form a vector

$$\tau_B = P(0, 0, \ldots, 0, (\tau_E)_i, 0, \ldots, 0) \tag{3.58}$$

belonging to the BISFP tangent cone at X_B^*. The BISFP Lagrangian Hessian is

$$\nabla^2 \mathcal{L}_B = P \begin{pmatrix} 0 & & & & \\ & (\nabla^2 \mathcal{L}_E)_1 & & & \\ & & (\nabla^2 \mathcal{L}_E)_2 & & \\ & & & \ddots & \\ & & & & (\nabla^2 \mathcal{L}_E)_N \end{pmatrix} P^T, \tag{3.59}$$

where $(\nabla^2 \mathcal{L}_E)_i$ is the ith subproblem Lagrangian Hessian at $(X_B^*)_i$. Because the SOSC hold at X_B^*, for any tangent cone vector τ_B we have

$$\tau_B^T (\nabla^2 \mathcal{L}_B) \tau_B > 0, \tag{3.60}$$

where τ_B is given by (3.58). Then (3.58)–(3.60) imply

$$(\tau_E)_i^T (\nabla^2 \mathcal{L}_E)_i (\tau_E)_i > 0. \tag{3.61}$$

■

Lemma 3.34 *If the functions F_i and c_i are three times continuously differentiable and the BISFP minimizer $X_B^* = P(z^*, (X_B^*)_1, (X_B^*)_2, \ldots, (X_B^*)_N)$ satisfies the LICQ,* *SCSC, and SOSC, then the EPD master problem objective*

$$F^*(\mu, z) = \sum_{i=1}^{N} F_i^*(\mu, z)$$

can be defined as a twice continuously differentiable function in a neighborhood $\mathcal{N}_\epsilon(z^*)$.

Proof: From Theorem 3.33 we know that if X_B^* is a nondegenerate BISFP minimizer, then $(X_B^*)_i$ is a minimizer satisfying the LICQ, SCSC, and SOSC for the ith EPD subproblem with $z = z^*$. Therefore, by the implicit function theorem and Theorem 6 in [FM68] we know that if F_i and c_i are three times continuously differentiable, then there exists a unique twice continuously differentiable trajectory of minimizers to the ith EPD subproblem $(X_B^*)_i(z)$ defined in a neighborhood $\mathcal{N}_{\epsilon_i}(z^*)$. The $(X_B^*)_i(z)$ define in turn a unique twice continuously differentiable function $F^*(\mu, z) = \sum_{i=1}^N F_i^*(\mu, z)$ on $\mathcal{N}_\epsilon(z^*)$, where $\epsilon = \min(\epsilon_1, \epsilon_1, \ldots, \epsilon_N)$. ■

Definition 3.35 Let $X_B^* = P(z^*, (X_B^*)_1, (X_B^*)_2, \ldots, (X_B^*)_N)$ be a BISFP point such that for $i = 1{:}N$ the vector $(X_B^*)_i$ is a minimizer satisfying the SOSC for the ith EPD subproblem with $z = z^*$. Then z^* is an EPD master problem minimizer if there exists a neighborhood $\mathcal{N}_\epsilon(z^*)$ such that for all $z \in \mathcal{N}_\epsilon(z^*)$ we have $F^*(\mu, z) > F^*(\mu, z^*)$, where F^* is the twice continuously differentiable master problem objective given by Theorem 3.34.

Theorem 3.36 *If* $X_B^* = P(z^*, (X_B^*)_1, (X_B^*)_2, \ldots, (X_B^*)_N)$ *is a minimizer satisfying the LICQ, SCSC, and SOSC for the BISFP, then* z^* *is a strict EPD master problem minimizer satisfying the SOSC.*

Proof: From Theorem 3.33 and Lemma 3.34 we know that there exist twice continuously differentiable trajectories of EPD subproblem minimizers $(X_B^*)_i(z)$ defined in a neighborhood $\mathcal{N}_{\epsilon_1}(z^*)$. Then, by the differentiability of $(X_B^*)_i(z)$ we know that for all $\epsilon_2 > 0$ we can always find $\epsilon_3 > 0$ such that $\epsilon_3 < \epsilon_1$ and for all $z \in \mathcal{N}_{\epsilon_3}(z^*)$,

$$\overline{X}_B(z) = (z, (X_B^*)_1(z), (X_B^*)_2(z), \ldots, (X_B^*)_N(z)) \in \mathcal{N}_{\epsilon_2}(X_B^*). \tag{3.62}$$

Because X_B^* is a strict EPD local minimizer, by Lemma 3.32, (3.62) implies that there exists $\epsilon_4 < \epsilon_3$ such that $F^*(\mu, z) > F^*(\mu, z^*)$ for all $z \in \mathcal{N}_{\epsilon_4}(z^*)$, where F^* is the master problem objective given by Lemma 3.34. Thus z^* is a strict EPD master problem minimizer. It only remains to show that the SOSC hold at z^* for the EPD master problem. It suffices to show that for all $v \neq 0$,

$$\frac{d^2 F^*(\mu, z^* + rv)}{dr^2} > 0.$$

But notice that

$$F^*(\mu, z^* + rv) = F_B(\hat{X}_B(r)),$$

where F_B is the BISFP objective function and $\hat{X}_B(r) = \overline{X}_B(z^* + rv)$. Moreover, because $\hat{X}_B(r)$ is BISFP feasible and the BISFP only has equality constraints, we know that

$$F_B(\hat{X}_B(r)) = \mathcal{L}_B(\hat{X}_B(r), \lambda^*),$$

where $\mathcal{L}_B$ is the Lagrangian function. Therefore,

$$\frac{d^2 F^*(\mu, z^* + rv)}{dr^2} = \frac{d^2 \mathcal{L}_B(\hat{X}_B(r), \lambda^*)}{dr^2}.$$

The first derivative of the BISFP Lagrangian function with respect to r is

$$\frac{d\mathcal{L}_B(\hat{X}_B(r), \lambda^*)}{dr} = \nabla_X \mathcal{L}_B(\hat{X}_B(r), \lambda^*) \frac{d\hat{X}_B(r)}{dr},$$

and the second derivative is

$$\begin{aligned} \frac{d^2 \mathcal{L}_B(\hat{X}_B(r), \lambda^*)}{dr^2} &= \frac{d\hat{X}_B(r)}{dr}^T \nabla^2_{XX} \mathcal{L}_B(\hat{X}_B(r), \lambda^*) \frac{d\hat{X}_B(r)}{dr} \\ &+ \nabla_x \mathcal{L}_B(\hat{X}_B(r), \lambda^*) \frac{d^2 \hat{X}_B(r)}{dr^2}. \end{aligned} \tag{3.63}$$

Evaluating (3.63) at $r = 0$ and because $\hat{X}_B(0)$ is a BISFP stationary point, we get

$$\left. \frac{d^2 F^*(\mu, z^* + rv)}{dr^2} \right|_{r=0} = \frac{d\hat{X}_B(0)}{dr}^T \nabla^2_{XX} \mathcal{L}_B(\hat{X}_B(0), \lambda^*) \frac{d\hat{X}_B(0)}{dr}. \tag{3.64}$$

Because $\hat{X}_B(r)$ is twice continuously differentiable, $\hat{X}_B(r)$ remains feasible for r small, and the LICQ holds for the BISFP at X_B^*, we know that $\frac{d\hat{X}_B(0)}{dr}$ belongs to the BISFP tangent cone. Moreover, because (X_B^*, λ^*) satisfies the SOSC, (3.64) implies that

$$\left. \frac{d^2 F^*(\mu, z^* + rv)}{dr^2} \right|_{r=0} = \frac{d\hat{X}_B(0)}{dr}^T \nabla^2_{XX} \mathcal{L}_B(\hat{X}_B(0), \lambda^*) \frac{d\hat{X}_B(0)}{dr} > 0.$$

■

3.3 Global Convergence Discussion

In Sections 3.1 and 3.2, we showed that IPD and EPD are fast locally convergent for the nonconvex OPGV. Unfortunately, we do not know of any globally convergent decomposition algorithm for the nonconvex OPGV. In this section we discuss the main difficulties encountered when trying to prove global convergence for decomposition algorithms for the nonconvex OPGV.

3.3.1 Multiple Subproblem Minimizers

For the nonconvex OPGV there might exist multiple local minimizers for each of the IPD and EPD subproblems. Consequently, the master problem objective is a set-valued function (see Aubin and Frankowska [AF90] for a reference on set-valued analysis.) Global convergence proofs for most optimization algorithms assume that the problem is defined in terms of single-valued functions [OR70].

Even if we assumed the set-valued function $F^*(z)$ is composed of a set of disjoint smooth functions $\{f_i^*(z)\}_{i=1}^k$ (a situation depicted in Figure 3.1), a line-search optimization algorithm could fail if, when it performs the master problem line search, it finds subproblem minimizers corresponding to different components functions. Likewise, a trust-region optimization algorithm would fail if, when it checks the accuracy of the local model for the master problem objective, it finds subproblem minimizers corresponding to different component functions. This difficulty could be alleviated by using parametric optimization algorithms [GVJ90]. However, global convergence has not been shown for parametric optimization algorithms.

3.3.2 Nonsmoothness

In Sections 3.1 and 3.2, we showed that given a nondegenerate OPGV minimizer, the IPD and EPD master problem objective function F^* is differentiable in a neighborhood of the equivalent IPD and EPD minimizers. However, F^* may be, for the nonconvex OPGV, nondifferentiable or even discontinuous at points located far from the minimizer. A discontinuous master problem objective is depicted in Figure 3.2.

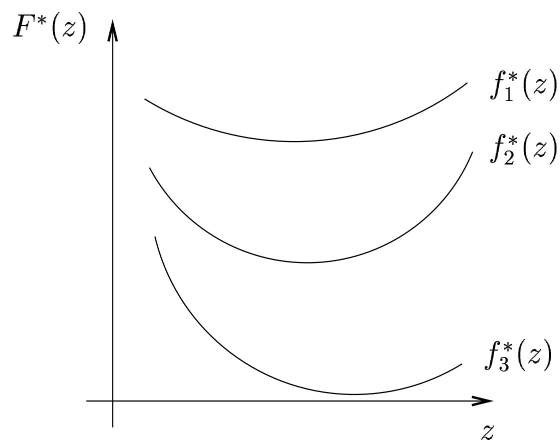

Figure 3.1: Master problem objective set-valued function.

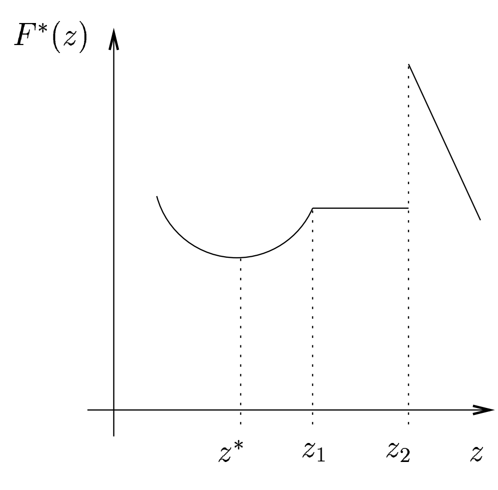

Figure 3.2: Master problem nonsmooth objective function: $F^*(z)$ is differentiable at the minimizer z^* but nondifferentiable at z_1 and discontinuous at z_2.

Chapter 4

A Test-Problem Set

A key factor in the development of any optimization algorithm is the availability of a suitable test-problem set. Unfortunately, there does not exist an appropriate nonconvex OPGV test-problem set. Although several OPGV test-problem sets [BDG+87], [Inf94, p. 47] have been developed in the context of the stochastic programming problem [BL97, Inf94], most of them correspond to linear or convex problems.

A more general test-problem set is the multidisciplinary design optimization test suite [PAG96]. For each test problem, a problem description, a benchmark solution method, sample input and output files, as well as source codes are available from the NASA Langley Research Center internet site. Test problems range from simple synthetic problems to some real engineering design problems. Unfortunately, the user has no control over important problem characteristics such as convexity and degree of degeneracy. Moreover, different test problems are given in different formats, and the implementation requires the modification of complicated FORTRAN source codes.

Easy-to-use nonconvex test-problem sets are available for several types of optimization problems related to the OPGV. Calamai and Vicente [CV94] developed a FORTRAN code to generate quadratic bilevel programs. The user can choose the test-problem size and the number of local and global minimizers. Moreover, all local and global minimizers are known a priori. Jiang and Ralph [JR99] developed a MATLAB code to generate mathematical programs with equilibrium constraints. Their test problems are more general than Calamai and Vicente's (which can be generated as a particular case) and the user can choose test-problem characteristics such as size, convexity, degeneracy, and ill conditioning. A disadvantage is that the minimizers of the test problems are not known in general.

In this chapter, we modify Calamai and Vicente's test problems to create a quadratic programming OPGV test-problem set. The test problem objective and constraint functions can be evaluated using two MATLAB M-files available from the author upon request. The user can choose the test-problem size, convexity, degeneracy, and degree of coupling. We calculate all local and global minimizers of the test-problem set and study their degree of degeneracy. Finally, the quadratic programming character of these test problems is not the limitation it may appear at first, because the master problem resulting from decomposition of a quadratic programming test problem is not, in general, a quadratic program.

4.1 A Convex Separable Test Problem

We propose the following convex quadratic programming OPGV:

$$
\begin{array}{lll}
\min\limits_{x,y_1,y_2} & \frac{1}{2}k_1\|x-a\|^2 + \frac{1}{2}k_2\|y_{11}-x\|^2 + \frac{1}{2}\|y_{12}\|^2 + & \\
& \frac{1}{2}k_1\|x-a\|^2 + \frac{1}{2}k_2\|y_{21}+x\|^2 + \frac{1}{2}\|y_{22}\|^2 & \\
s.t. & e \le \quad x + y_{11} \quad \le 2e, & \\
& \qquad\quad x - y_{11} \quad \le e, & \\
& e \le \; -x + y_{21} \quad \le 2e, & \\
& \qquad -x - y_{21} \quad \le e, &
\end{array}
\tag{4.1}
$$

where $x \in \mathbb{R}^n$ are the global variables, $y_i = (y_{i1}, y_{i2}) \in \mathbb{R}^{n_i}$ are the ith system local variables with $y_{i1} \in \mathbb{R}^n$, $y_{i2} \in \mathbb{R}^{n_i - n}$, $k_1, k_2 \in \mathbb{R}$, and $e \in \mathbb{R}^n$ is the vector whose components are all ones.

For $k_1, k_2 > 0$ the objective function Hessian corresponding to (4.1) is positive definite and therefore the quadratic program is strictly convex. By changing n, n_1, and n_2, we can choose the size of the test problem. Likewise, by changing the ratios n_1/n and n_2/n, the user can control the degree of coupling among the two systems that compose (4.1). Finally, different degrees of degeneracy can be obtained by careful choice of a.

Note that Problem 4.1 can be separated into $n+2$ independent problems. Each of the first n problems is formed by the objective function terms and the constraints that depend on the rth component of the vectors x, y_{11}, and y_{21}, which we denote as x_r, y_{11r}, and y_{21r}. We term these n problems *three-variable convex problems*, namely,

$$\begin{array}{llcccl} \min\limits_{x_r, y_{11r}, y_{21r}} & \multicolumn{5}{l}{\frac{1}{2}k_1(x_r - a)^2 + \frac{1}{2}k_2(y_{11r} - x_r)^2 +} \\ & \multicolumn{5}{l}{\frac{1}{2}k_1(x_r - a)^2 + \frac{1}{2}k_2(y_{21r} + x_r)^2} \\ s.t. & 1 & \leq & x_r + y_{11r} & \leq & 2, \\ & & & x_r - y_{11r} & \leq & 1, \\ & 1 & \leq & -x_r + y_{21r} & \leq & 2, \\ & & & -x_r - y_{21r} & \leq & 1. \end{array} \tag{4.2}$$

The last two problems that compose (4.1) are unconstrained quadratic programs formed by the objective function terms that depend only on y_{12} or y_{22}, namely,

$$\min_{y_{r2}} \frac{1}{2}\|y_{r2}\|^2, \qquad r = 1, 2. \tag{4.3}$$

Although these unconstrained problems may seem trivial at first glance, in Section 4.3 we explain how a change of variables can be used to intertwine Problems 4.2 and 4.3 into a nonseparable test problem. Moreover, these unconstrained problems allow us to control the degree of coupling among systems by changing the dimension of y_{12} and y_{22}.

4.1.1 Minimizers

To find the minimizer of the convex test problem it suffices to find the minimizer of the $n+2$ problems that compose it. Since the minimizers of the two unconstrained problems (4.3) are obviously $y_{12}^* = 0$ and $y_{22}^* = 0$, it only remains to calculate the minimizers of the three-variable convex problem.

Provided $k_1, k_2 > 0$, (4.2) is a strictly convex quadratic program. Moreover, its objective function is nonnegative and hence bounded below. Therefore, for each a, there exists a unique minimizer of (4.2). This unique minimizer can be found by solving the KKT conditions. Here, we give the minimizer for $a \geq 0$. Because of the symmetry of the problem, the minimizer for $a < 0$ is just $(-x_r^*, y_{11r}^*, y_{21r}^*)$, where $(x_r^*, y_{11r}^*, y_{21r}^*)$ is the minimizer corresponding to $|a|$. We distinguish four cases:

Case 1 ($0 \leq a \leq 1/2 + 2k_2/k_1$): The active set is formed by the constraints $x_r + y_{11r} = 1$ and $-x_r + y_{21r} = 1$. The minimizer is

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} \frac{k_1}{k_1+4k_2}a \\ 1 - x_r^* \\ 1 + x_r^* \end{pmatrix}.$$

Case 2 ($1/2+2k_2/k_1 \leq a \leq 1+3k_2/k_1$): The active set is formed by the constraint $-x_r + y_{21r} = 1$. The minimizer is

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} \frac{k_1 a - k_2}{k_1+2k_2} \\ x_r^* \\ 1 + x_r^* \end{pmatrix}.$$

Case 3 ($1+3k_2/k_1 \leq a \leq 3/2+5k_2/k_1$): The active set is formed by the constraints $x_r + y_{11r} = 2$ and $-x_r + y_{21r} = 1$. The minimizer is

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} \frac{k_1 a + k_2}{k_1+4k_2} \\ 2 - x_r^* \\ 1 + x_r^* \end{pmatrix}.$$

Case 4 ($3/2+5k_2/k_1 \leq a$): The active set is formed by the constraints $x_r + y_{11r} = 2$, $x_r - y_{11r} = 1$ and $-x_r + y_{21r} = 1$. The minimizer is

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0.5 \\ 2.5 \end{pmatrix}.$$

The set of minimizers of the three-variable convex problem corresponding to $a \in (-\infty, \infty)$ is depicted in Figure 4.1. The graph at the top represents y_{11r}^* as a function of x_r^* and the graph at the bottom represents y_{21r}^* as a function of x_r^*.

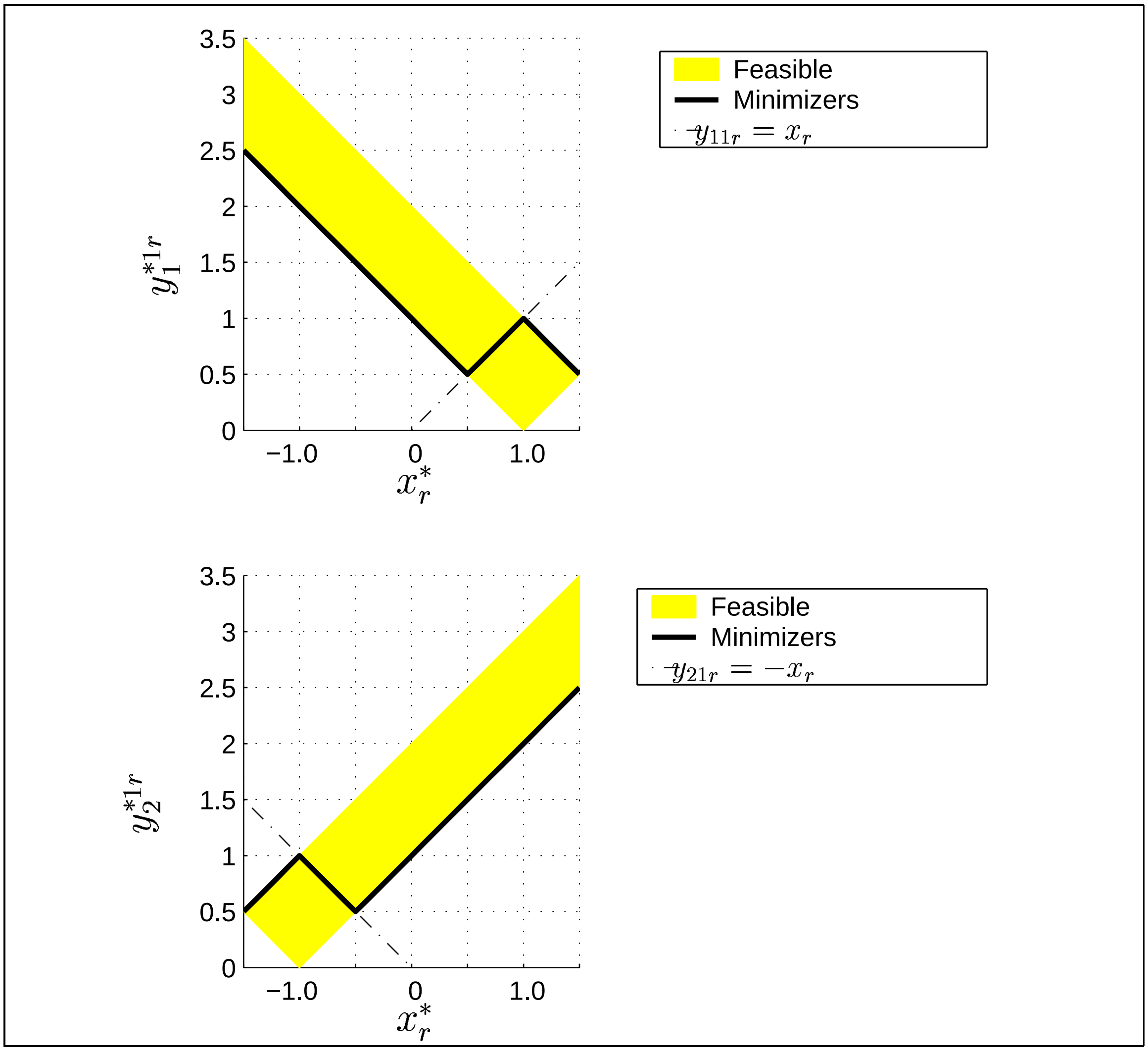

Figure 4.1: Minimizers to the three variable convex problem for $a \in (-\infty, \infty)$.

4.1.2 Degeneracy

The degree of degeneracy of the minimizer of (4.1) depends on the value of a. Provided $n \geq 1$ and $k_1, k_2 > 0$, the following propositions give the set of values of a for which the LICQ, SCSC, and SOSC hold at the minimizer.

Proposition 4.1 *The LICQ and SOSC hold at the unique minimizer of (4.1) for all a.*

Proof: It is easy to show from the structure of the active set at the minimizer of (4.2) that the LICQ holds. Also, if $k_1, k_2 > 0$, then the Hessian of the Lagrangian for Problem 4.1 is positive definite for all a and therefore the SOSC hold. ■

Proposition 4.2 *The SCSC holds at the minimizer of (4.1) iff for $i = 1{:}n$, a_i is not in the set $\{1/2 + 2k_2/k_1, 1 + 3k_2/k_1, 3/2 + 5k_2/k_1\}$.*

Proof: This follows immediately from the KKT conditions of (4.2). ■

Proposition 4.3 *The SLICQ holds at the minimizer of (4.1) iff for $i = 1{:}n$, $a_i < 3/2 + 5k_2/k_1$.*

Proof: This is obvious from the active set at the minimizer of (4.2). ■

4.2 A Nonconvex Separable Test Problem

We propose the following nonconvex quadratic programming OPGV:

$$\begin{array}{lrcll} \min\limits_{x,y_1,y_2} & \multicolumn{4}{l}{\frac{1}{2}k_1\|x-a\|^2 - \frac{1}{2}k_2\|y_{11} - (-x+be)\|^2 + \frac{1}{2}\|y_{12}\|^2 +} \\ & \multicolumn{4}{l}{\frac{1}{2}k_1\|x-a\|^2 - \frac{1}{2}k_2\|y_{21} - (x+be)\|^2 + \frac{1}{2}\|y_{22}\|^2} \\ s.t. & e \;\leq & x + y_{11} & \leq & 2e, \\ & & x - y_{11} & \leq & e, \\ & e \;\leq & -x + y_{21} & \leq & 2e, \\ & & -x - y_{21} & \leq & e. \end{array} \tag{4.4}$$

Note that the feasible regions of the convex and nonconvex test problems are identical. The nonconvex test problem is obtained from the convex test problem by replacing the objective function terms $\frac{1}{2}k_2\|y_{11}-x\|^2+\frac{1}{2}k_2\|y_{21}+x\|^2$ by the terms $-\frac{1}{2}k_2\|y_{11}-(-x+be)\|^2-\frac{1}{2}k_2\|y_{21}-(x+be)\|^2$.

As in the convex case, the nonconvex test problem can be separated into $n+2$ independent problems. The first n problems are termed *three-variable nonconvex problems* and can be written as follows:

$$\begin{array}{lrcll} \min\limits_{x_r, y_{11r}, y_{21r}} & \multicolumn{4}{l}{\frac{1}{2}k_1(x_r-a)^2+\frac{1}{2}k_2(y_{11r}-(-x_r+b))^2+} \\ & \multicolumn{4}{l}{\frac{1}{2}k_1(x_r-a)^2+\frac{1}{2}k_2(y_{21r}-(x_r+b))^2} \\ s.t. & 1 \leq & x_r+y_{11r} & \leq & 2, \\ & & x_r-y_{11r} & \leq & 1, \\ & 1 \leq & -x_r+y_{21r} & \leq & 2, \\ & & -x_r-y_{21r} & \leq & 1. \end{array} \tag{4.5}$$

The last two problems that compose the nonconvex test problem are the following unconstrained optimization problems:

$$\min_{y_{r2}} \frac{1}{2}\|y_{r2}\|^2, \qquad r=1,2. \tag{4.6}$$

4.2.1 Minimizers

Since the minimizers of the two unconstrained problems (4.3) are obviously $y_{12}^*=0$ and $y_{22}^*=0$, we only need to calculate the minimizers of the three-variable nonconvex problem. Because of the symmetry of the problem it suffices to compute the local minimizers for $a>0$. Here, we give the local minimizers for $k_1>2k_2>0$ and $b=1.5$. We distinguish five cases:

Case 1 $(0\leq a\leq 1)$: There exist four local minimizers that are also global:

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} a \\ 1-x_r^* \\ 1+x_r^* \end{pmatrix}, \begin{pmatrix} a \\ 2-x_r^* \\ 1+x_r^* \end{pmatrix}, \begin{pmatrix} a \\ 1-x_r^* \\ 2+x_r^* \end{pmatrix}, \begin{pmatrix} a \\ 2-x_r^* \\ 2+x_r^* \end{pmatrix}.$$

Case 2 $(1<a\leq 1+(b-1)k_2/k_1)$: There exist two global minimizers:

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} a \\ 2-x_r^* \\ 1+x_r^* \end{pmatrix}, \begin{pmatrix} a \\ 2-x_r^* \\ 2+x_r^* \end{pmatrix},$$

and two local minimizers:

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}.$$

Case 3 $(1 + (b-1)k_2/k_1 \leq a < 1.25)$: There exist two global minimizers:

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} a \\ 2 - x_r^* \\ 1 + x_r^* \end{pmatrix}, \begin{pmatrix} a \\ 2 - x_r^* \\ 2 + x_r^* \end{pmatrix},$$

and two local minimizers:

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} \frac{k_1 a - (1+b)k_2}{k_1 - 2k_2} \\ x_r^* - 1 \\ 1 + x_r^* \end{pmatrix}, \begin{pmatrix} \frac{k_1 a - (1+b)k_2}{k_1 - 2k_2} \\ x_r^* - 1 \\ 2 + x_r^* \end{pmatrix}.$$

Case 4 $(1.25 \leq a \leq 1.5)$: There exist two global minimizers:

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} a \\ 2 - x_r^* \\ 1 + x_r^* \end{pmatrix}, \begin{pmatrix} a \\ 2 - x_r^* \\ 2 + x_r^* \end{pmatrix}.$$

Case 5 $(1.5 \leq a)$: There exist two global minimizers:

$$\begin{pmatrix} x_r^* \\ y_{11r}^* \\ y_{21r}^* \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0.5 \\ 2.5 \end{pmatrix}, \begin{pmatrix} 1.5 \\ 0.5 \\ 3.5 \end{pmatrix}.$$

The set of minimizers of the three-variable nonconvex problem for $k_1 > 2k_2 > 0$ and $b = 1.5$ is depicted in Figure 4.2 for $a \in (-\infty, \infty)$. The graph at the top represents y_{11r}^* as a function of x_r^* and the graph at the bottom represents y_{21r}^* as a function of x_r^*.

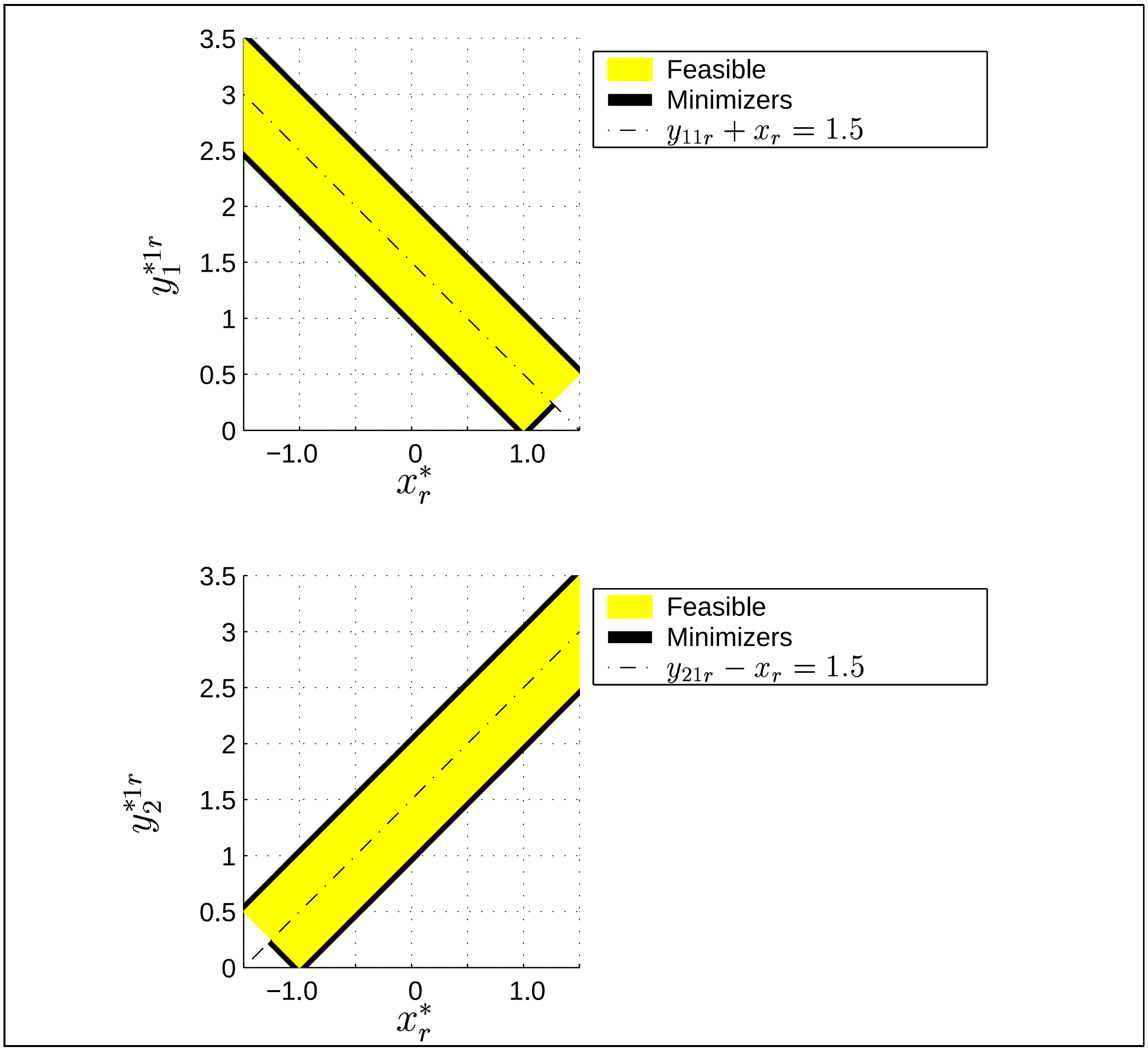

Figure 4.2: Minimizers to the three variable nonconvex problem for $a \in (-\infty, \infty)$.

4.2.2 Degeneracy

Provided $n \geq 1$, $k_1 > 2k_2 > 0$ and $b = 1.5$ the following propositions give the particular values of a for which the LICQ, SCSC, and SOSC hold at the minimizer of (4.4).

Proposition 4.4 *The LICQ and SOSC hold at all local minimizers of (4.4) for all* a.

Proof: It is easy to show from the structure of the active set at the minimizer of (4.5) that the LICQ holds. Likewise, if $k_1 > 2k_2 > 0$ and $b = 1.5$ it is easy to show that the SOSC are satisfied at all local minimizers. ■

Proposition 4.5 *The SCSC holds at all local minimizers of (4.4) iff for* $i = 1{:}n$, a_i *is not in the set* $\{1, 1.5\}$.

Proof: This follows immediately from the KKT conditions of (4.5). ■

Proposition 4.6 *The SLICQ holds at the minimizer of (4.1) iff for* $i = 1{:}n$, $a_i < 3/2$.

Proof: This is obvious from the active set at the minimizer. ■

4.3 A Nonseparable Test Problem

The convex and nonconvex test problems can be separated into $n + 2$ independent problems. The iterative procedure required to solve these separable test problems is numerically equivalent, for most algorithms, to the one needed to solve the $n + 2$ problems independently. Therefore, to analyze how the performance of a decomposition algorithm depends on problem size, we need to modify our test problems so that they are not separable.

Vicente and Calamai used a transformation matrix to obtain nonseparable test problems from their separable bilevel quadratic test problems. Here, we need to ensure that the test problems maintain the OPGV structure. The transformation we propose is

$$\begin{pmatrix} \hat{x} \\ \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} = \begin{pmatrix} Q_x & & \\ & Q_{y_1} & \\ & & Q_{y_2} \end{pmatrix} \begin{pmatrix} x \\ y_1 \\ y_2 \end{pmatrix},$$

where Q_x, Q_{y_1}, and Q_{y_2} are randomly generated orthogonal matrices. It is easy to show that the test problems in the variables $(\hat{x}, \hat{y}_1, \hat{y}_2)$ are OPGVs. Moreover, the transformed test problems are not separable.

Chapter 5

Computational Results

Both IPD and EPD have been implemented and applied to the OPGV test-problem set introduced in Chapter 4. A variety of convex and nonconvex test problems corresponding to different sizes, degrees of degeneracy, and intensity of coupling among systems were solved. The results show that both decomposition algorithms performed satisfactorily on the test-problem set.

In Section 5.1, we give the details of the IPD and EPD algorithms. In Section 5.2, we compute minimizers of the IPD and EPD subproblems and analyze the characteristics of the computed solution. Finally, in Section 5.3, we analyze the numerical performance of IPD and EPD on the test-problem set.

5.1 Algorithm Statement

5.1.1 Master Problem Algorithm

We use the MATLAB M-file QNSOL, which is a modification of a previous code obtained from Philip Gill, to solve the IPD and EPD master problems. QNSOL is a BFGS quasi-Newton unconstrained optimization algorithm [GMW81, Chapter 4]. QNSOL may be stated as follows:

Algorithm QNSOL

Step 0 (Initialization) Initialize the penalty parameter $\gamma := \gamma_0$, the quasi-Newton Hessian approximation $B := I$, and the optimality tolerance $\epsilon := \epsilon_0$. If EPD then initialize the barrier parameter $\mu := \mu_0$.

Step 1 (Starting point) Set $z := z_0$; call SUBSOL with γ (and μ if EPD) to evaluate the objective function $F^*(z)$ and its gradient $\nabla F^*(z)$.

Step 2 (Solve master problem with current γ, ϵ (and μ if EPD))
while $(\|\nabla F^*(z)\|/(1 + F^*(z)) < \epsilon)$

Step 2.1 (Search direction) Solve $B\Delta z = -\nabla F^*(z)$.

Step 2.3 (Line search) Set $\alpha = 1$.

while $(F^*(z + \alpha\Delta z) - F^*(z) > \sigma\nabla F^*(z)\Delta z)$
Set $\alpha := \alpha/2$; call SUBSOL with γ (and μ if EPD) to evaluate $F^*(z + \alpha\Delta z)$ and $\nabla F^*(z + \alpha\Delta z)$.
endwhile
Set $s := \alpha\Delta z$, $y := \nabla F^*(z + \alpha\Delta z) - \nabla F^*(z)$
Set $z := z + s$; update $F^*(z)$ and $\nabla F^*(z)$

Step 2.4 (BFGS update)

$$B := B - \frac{Bss^TB}{s^TBs} + \frac{yy^T}{y^Ts}$$

endwhile

Step 3 (Convergence check and parameter update)

if IPD
if $(\epsilon < \epsilon_1$ AND $\sum_{i=1}^{N} \|x_i^* - z\|_2^2 < \epsilon)$ then stop,
else update γ and ϵ; call SUBSOL with γ to evaluate $F^*(z)$ and $\nabla F^*(z)$; go to Step 2.
elseif EPD
if $(\epsilon < \epsilon_1$ AND $\sum_{i=1}^{N} \|x_i^* - z\|_1 < \epsilon$ AND $\mu < \epsilon)$ then stop,
else update γ, ϵ and μ; call SUBSOL with γ and μ to evaluate $F^*(z)$ and $\nabla F^*(z)$; go to Step 2.
endif

Note that, although second-order derivatives are known for the OPGV test-problem set, we assume only first-order derivatives can be computed because this is the case for many real-world OPGVs. As a consequence, the exact master problem Hessian cannot be computed from the subproblem solution. However, a BFGS quasi-Newton update is used to build an approximation to the master problem Hessian.

The algorithm parameters are updated in Step 2. The penalty parameter γ is increased until $\max_i \|x_i^* - z\|_\infty$ is smaller than the optimality tolerance ϵ. In addition, for EPD, the barrier parameter μ is also driven below ϵ. Meanwhile, ϵ is driven below the small number ϵ_1. In our implementation we set $\epsilon_1 = 10^{-5}$.

5.1.2 Subproblem Algorithm

To evaluate the master problem objective function, QNSOL calls the MATLAB M-file SUBSOL. This M-file uses the sequential quadratic programming algorithm NPSOL [GMSW86] to solve the IPD subproblems (2.12) and the rudimentary primal-dual algorithm PDSOL (coded in MATLAB) to solve the EPD subproblems (2.15). SUBSOL may be stated as follows:

Algorithm SUBSOL

if IPD then

for i=1:N call NPSOL to solve the ith IPD subproblem (2.12) at z.

Compute $F^*(z) = \sum_{i=1}^N F_i^*(z)$ and $\nabla F^*(z) = -\sum_{i=1}^N (x_i^* - z)$,

elseif EPD then

for i=1:N call PDSOL to solve the ith EPD subproblem (2.15) at z.

Compute $F^*(z) = \sum_{i=1}^N F_i^*(z)$ and $\nabla F^*(z) = \sum_{i=1}^N \lambda_{zi}$, where λ_{zi} are the multipliers corresponding to the constraints $x_i - s_i + t_i = z$.

endif

To compute the master problem objective function with the precision necessary to ensure fast local convergence for QNSOL, a tight optimality tolerance must be used to solve the subproblems in SUBSOL. In our implementation, we use a subproblem optimality tolerance of ϵ^2, where ϵ is the current master problem optimality tolerance.

5.2 Solving the Subproblems

NPSOL and PDSOL are used to compute minimizers of the IPD and EPD subproblems corresponding to the three-variable convex and nonconvex test problems. We analyze the characteristics of the computed minimizers and the corresponding optimal-value functions, that is, the subproblem minimum objectives as a function of the target variables. These characteristics will in turn influence the behavior of the optimization algorithms used to solve the master problem.

5.2.1 Convex Subproblems

To study the smoothness properties of the IPD and EPD subproblem optimal-value functions we solve the first of the two subproblems that result from the decomposition of the three-variable convex test problem (4.2). We set $k_1 = 1$, $k_2 = 1$, and $a = 0$ and find the unique subproblem minimizer for $z \in (-1.6, 1.6)$. We plot the IPD subproblem minimizer and optimal-value function for $\gamma = 20, 50, 10^5$ and the EPD subproblem minimizer and optimal-value function for $\gamma = 20$ and $\mu = 0.3, 0.1, 0.06, 0.01, 0.001$.

Minimizer

Figure 5.1 depicts the trajectories of IPD and EPD subproblem minimizers corresponding to $z \in (0, 1.6)$. The IPD minimizer trajectory (thick line) is nonsmooth around the points $x_1^* = 0.5$ and $x_1^* = 1.0$. On the other hand, the EPD minimizer trajectory (thin line) *is* smooth around these points. However, as μ is driven to zero, the EPD trajectory converges to the IPD trajectory and in the limit is nonsmooth also.

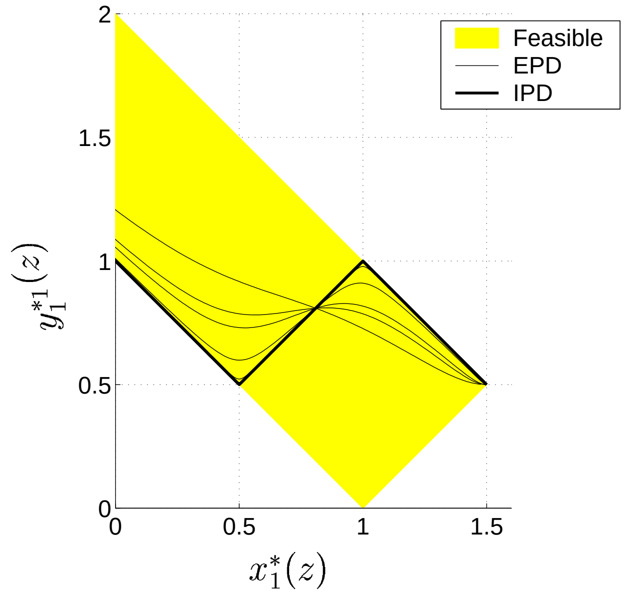

Figure 5.1: IPD and EPD convex subproblem minimizers (1/2).

Figure 5.2 depicts the optimal global variable x_1^* as a function of the target variable z, that is, $x_1^*(z)$. Notice that the EPD subproblem minimizer (solid line) satisfies $x_1^*(z) = z$ for $\gamma = 20$. In other words, the exact penalty function recovers the exact minimizer for finite values of γ. On the other hand, when using IPD (dashed line), we need to drive γ to infinity in order to enforce $x_1^*(z) = z$. In particular, note that for small γ, $x_1^*(z)$ is considerably greater than z.

Optimal-Value Function

Figure 5.3 depicts the optimal-value function $F_1^*(z)$. Both the IPD and EPD optimal-value functions are differentiable for all z. However, when γ is large the IPD optimal-value function has large second derivatives around $z = 1.5$. Likewise, when μ is small the EPD optimal-value function has large second derivatives. In both cases, the minimizer of $F_1^*(z)$ is not close to $z = 1.5$ and therefore numerical difficulty is not expected when solving either the IPD or the EPD master problem.

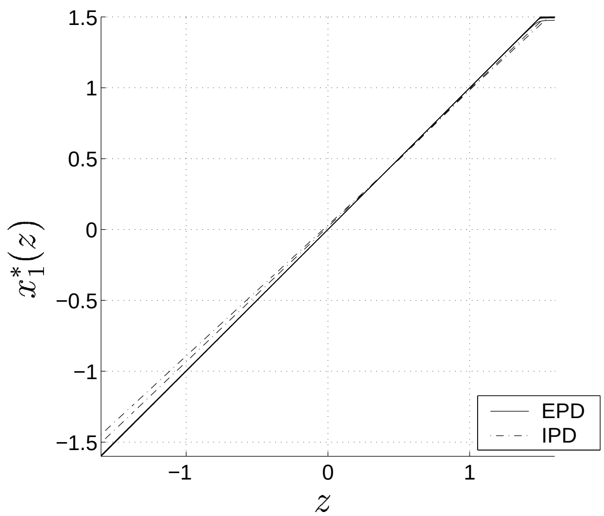

Figure 5.2: IPD and EPD convex subproblem minimizers (2/2).

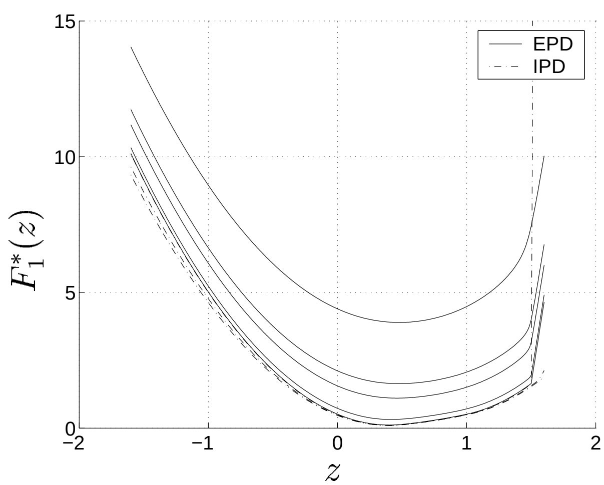

Figure 5.3: IPD and EPD convex subproblem optimal-value functions for $a = 0$.

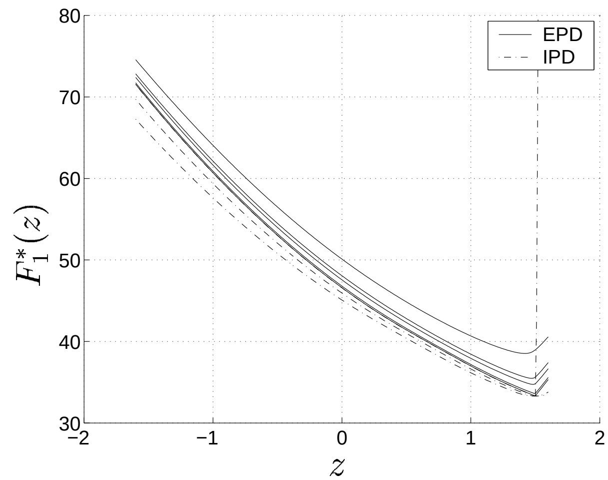

Figure 5.4: IPD and EPD convex subproblem optimal-value functions for $a = 9.6$.

Figure 5.4 depicts $F_1^*(z)$ for $a = 9.6$. In this case, the minimizer of $F_1^*(z)$ is $z = 1.5$. Notice that the IPD and EPD optimal-value functions have large second derivatives around their minimizers. Hence, numerical difficulty can be expected when we try to solve the IPD and EPD master problems near the point of convergence.

5.2.2 Nonconvex Subproblems

When decomposing the nonconvex OPGV, we might find that there exist multiple local minimizers of the subproblems for each value of the target variables. Consequently, there exist several trajectories of subproblem minimizers defined as a function of the target variables. Unfortunately, the algorithm solving the master problem will break if, during the line search, the subproblem solver finds minimizers corresponding to different trajectories.

This difficulty can be alleviated by using the subproblem minimizer found at the current iterate for the target variables as a starting point to solve the subproblem at other points along the search direction. For small step lengths, all subproblem minimizers are likely to belong to the same trajectory.

To illustrate this, we solve the first of the two IPD and EPD subproblems that

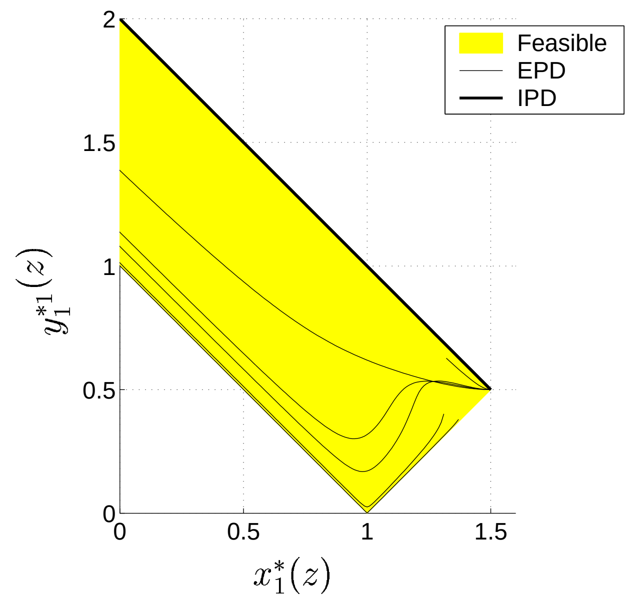

Figure 5.5: IPD and EPD minimizers of a nonconvex subproblem using a good starting point.

result from the decomposition of the three-variable nonconvex test problem (4.5). We set $k_1 = 1$, $k_2 = 1$, $a = 0$, and $b = 1.8$ and find a local minimizer for $z = -1.6{:}0.01{:}1.6$. At each point z, we use the subproblem minimizer found for $z - 0.01$ as a starting point. We compute the IPD subproblem minimizer trajectory and optimal-value function (dashed lines) for $\gamma = 20, 50, 10^5$ and the EPD subproblem minimizer and optimal-value function (solid lines) for $\gamma = 20$ and $\mu = 0.3$, 0.1, 0.06, 0.01, 0.001.

Figure 5.5 depicts the minimizer trajectory $(x_1^*(z), y_{11}^*(z))$. Note that by using a good starting point we manage to find a set of connected local minimizers with $x_1^* \in [-1.6, 1.4)$. At $x_1^* = 1.4$ the lower trajectory of minimizers ends and as a consequence we find minimizers corresponding to the upper trajectory of minimizers with $x_1^* \geq 1.4$. Figure 5.6 depicts $F_1^*(z)$. Note that jumps do *not* occur in the optimal-value function for $z \in [-1.6, 1.4)$. Also, note that the optimal-value function is nonconvex for $z \in [1.0, 1.4)$.

Finally, we solve the EPD nonconvex subproblem for $k_1 = 1$, $k_2 = 10$, $a = 0$, $b = 1.8$, $\gamma = 20$, and $\mu = 0.01$ and manipulate the starting point so that we find

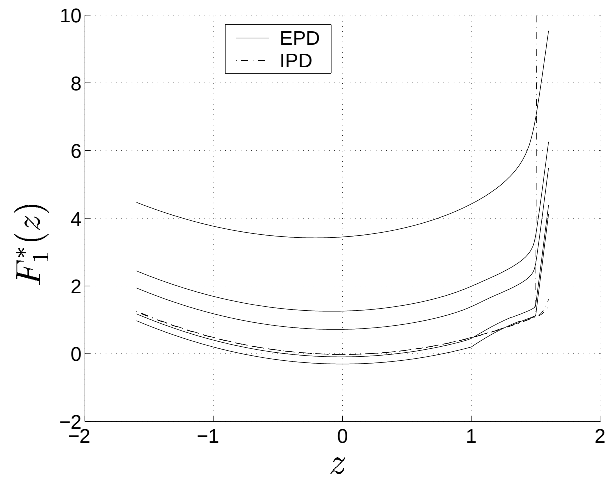

Figure 5.6: IPD and EPD nonconvex subproblem optimal value-function using a good starting point.

minimizers belonging to the lower trajectory of minimizers for $z \in (0.5, 1.0)$ and minimizers belonging to the upper trajectory at all other points. Figure 5.7 depicts $(x_1^*(z), y_{11}^*(z))$ and Figure 5.8 depicts $F_1^*(z)$. Note that discontinuities occur in the optimal-value function when we jump between different trajectories of minimizers.

5.3 Solving the Master Problem

QNSOL is used to solve the IPD and EPD master problems corresponding to a number of OPGV test problems. In Section 5.3.1, we discuss results corresponding to convex test problems. Section 5.3.2 deals with nonconvex test problems. Finally, Section 5.3.3 studies the influence of the degree of coupling among subproblems on the behavior of IPD and EPD.

5.3.1 Convex Test Problems

The convex test problems corresponding to $k_1 = 1$, $k_2 = 1$, $n_1/n = 5$, $n_2/n = 5$ and $n = 2, 3, 4, 5, 6, 7, 8, 9, 10$ were solved. We distinguish two cases: (i) test problems

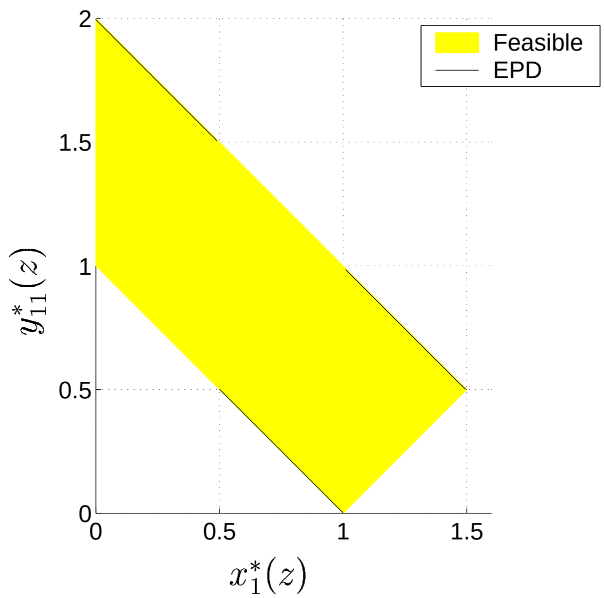

Figure 5.7: EPD minimizers of a nonconvex subproblem using a bad starting point.

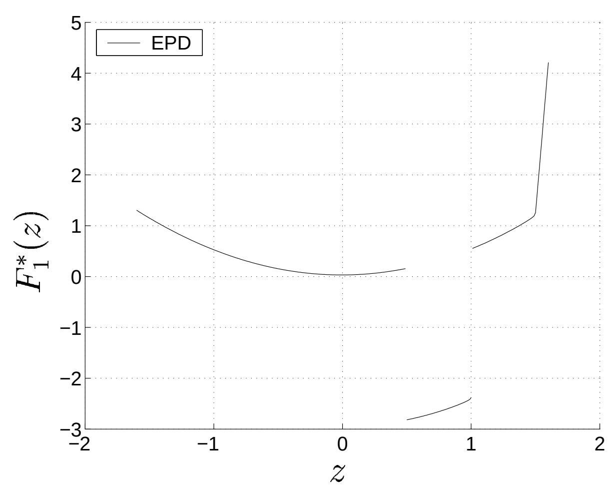

Figure 5.8: EPD nonconvex subproblem optimal value-function using a bad starting point.

satisfying the SLICQ ($a = 0$) and (ii) test problems satisfying only the LICQ ($a = 9.6e$).

Convex Test Problems satisfying SLICQ

The results for $a = 0$ are given in Table 5.1. The first column is the number of variables in the test problem. The next three columns give information regarding the computational effort required to solve the test problem. In particular, the second column is the number of QNSOL iterations required to solve the master problem. The third column is the number of subproblems whose minimizer was found over the number of subproblems tried. The number of subproblems tried reflects the amount of communication required between the master problem and the subproblems. The fourth column is the total number of test-problem function evaluations, which represents the total computational effort needed to solve the test problem. The fifth column gives an exit code equal to 0 if a stationary point for the master problem was found, 6 if the master problem line search failed, and 3 if the iteration limit for QNSOL was exceeded. The sixth column is the final value of the penalty parameter γ. The seventh column is $\max(\|x_1^* - z\|_\infty, \|x_2^* - z\|_\infty)$, which is a measure of the feasibility of the final iterate. The eighth column gives the master problem objective function at the final iterate. Finally, the ninth column gives the norm of the objective function gradient scaled by the value of the objective function $\|\nabla F^*(z)\|/|1 + F^*(z)|$, where the master problem objective $F^*(z) = F_1^*(z) + F_2^*(z)$.

Table 5.1 shows that both IPD and EPD find stationary points of all convex test problems tried. Note that IPD is slightly more efficient than EPD in terms of number of subproblems tried and number of function evaluations. This could be explained by the quadratic nature of the OPGV test-problem set. In particular, a sequential quadratic programming algorithm such as NPSOL is more efficient than a primal-dual method such as PDSOL on quadratic programming test problems.

Note that in IPD, γ has to be driven to 10^5 in order to achieve feasibility, whereas $\gamma = 10$ suffices to achieve feasibility for EPD. Finally, the master problem minimum values computed by IPD and EPD are slightly different because of the different penalty terms used in the subproblem objective functions.

IPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
22	3	13/13	152	0	1.0e+07	1.4e-07	2.0000e+00	4.7e-07
33	4	16/16	187	0	1.0e+07	1.5e-07	3.0000e+00	2.6e-06
44	1	11/11	141	0	1.0e+07	1.7e-07	4.0000e+00	1.0e-08
55	3	13/13	168	0	1.0e+07	2.0e-07	5.0000e+00	3.6e-06
66	3	13/13	171	0	1.0e+07	2.1e-07	6.0000e+00	1.4e-06
77	5	18/18	313	0	1.0e+07	2.4e-07	7.0000e+00	7.6e-08
88	5	19/19	313	0	1.0e+07	2.5e-07	8.0000e+00	3.6e-07
99	2	12/12	164	0	1.0e+07	2.6e-07	9.0000e+00	3.9e-06
110	4	17/17	286	0	1.0e+07	2.7e-07	1.0000e+01	3.3e-06

EPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
22	6	19/19	265	0	1.0e+01	6.0e-08	2.0002e+00	3.7e-06
33	7	21/21	303	0	1.0e+01	6.4e-07	3.0025e+00	1.1e-06
44	7	20/20	290	0	1.0e+01	7.9e-07	4.0033e+00	2.1e-06
55	8	24/24	358	0	1.0e+01	9.5e-07	5.0041e+00	8.4e-06
66	8	25/25	378	0	1.0e+01	1.0e-06	6.0050e+00	2.9e-06
77	9	24/24	372	0	1.0e+01	1.2e-06	7.0058e+00	2.8e-06
88	8	23/23	354	0	1.0e+01	1.4e-06	8.0066e+00	1.9e-06
99	9	25/25	409	0	1.0e+01	1.5e-06	9.0075e+00	3.7e-06
110	8	22/22	396	0	1.0e+01	1.6e-06	1.0008e+01	2.9e-06

Table 5.1: Convex test problems satisfying SLICQ.

IPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
22	16	57/57	405	0	1.0e+05	7.0e-07	1.4822e+02	1.9e-07
33	22	79/79	615	0	1.0e+05	7.3e-07	2.2233e+02	5.1e-08
44	14	77/77	733	0	1.0e+05	8.8e-07	2.9644e+02	6.8e-06
55	19	107/107	1004	0	1.0e+05	1.0e-06	3.7055e+02	1.5e-06
66	36	149/149	1282	0	1.0e+06	1.1e-07	4.4466e+02	1.9e-07
77	33	161/161	1762	0	1.0e+06	1.2e-07	5.1877e+02	1.7e-06
88	35	189/189	2613	0	1.0e+06	1.3e-07	5.9288e+02	2.6e-06
99	37	205/205	3107	0	1.0e+06	1.3e-07	6.6699e+02	6.7e-07
110	47	221/221	3303	0	1.0e+06	1.4e-07	7.4110e+02	3.8e-07

EPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
22	43	113/113	1532	0	1.0e+01	2.8e-09	1.4822e+02	2.1e-07
33	49	126/126	1806	0	1.0e+01	2.9e-09	2.2233e+02	4.4e-08
44	53	145/145	2159	0	1.0e+01	3.5e-09	2.9644e+02	9.5e-07
55	62	182/182	2831	0	1.0e+01	4.0e-09	3.7055e+02	1.1e-06
66	73	204/204	3247	0	1.0e+01	4.3e-09	4.4466e+02	2.5e-07
77	80	233/233	3636	0	1.0e+01	4.9e-09	5.1877e+02	2.9e-06
88	89	254/254	4062	0	1.0e+01	5.1e-09	5.9288e+02	3.0e-07
99	95	269/269	4436	0	1.0e+01	5.3e-09	6.6699e+02	2.4e-06
110	104	297/297	5035	0	1.0e+01	5.5e-09	7.4110e+02	8.8e-07

Table 5.2: Convex test problems satisfying only LICQ.

Convex Test Problems satisfying LICQ

The results for $a = 9.6e$ are given in Table 5.2. Both IPD and EPD solve all test problems tried. Notice that the computational effort required to solve test problems satisfying the LICQ is an order of magnitude higher than that required to solve test problems satisfying the SLICQ. Also, for test problems satisfying only the LICQ the computational effort grows considerably with problem dimension.

IPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
22	2	7/7	105	0	1.0e+02	5.2e-12	-9.8000e-01	1.0e-09
33	2	6/6	100	0	1.0e+02	7.5e-12	-7.2000e-01	2.6e-09
44	1	6/6	116	0	1.0e+02	2.6e-11	-1.6600e+00	4.0e-09
55	2	8/8	174	0	1.0e+02	3.2e-09	-2.3000e+00	6.4e-07
66	2	8/8	168	0	1.0e+02	2.2e-09	-3.2400e+00	2.3e-07
77	2	7/7	143	0	1.0e+02	1.2e-09	-2.6800e+00	1.8e-07
88	2	7/7	141	0	1.0e+02	1.7e-08	-1.8200e+00	3.2e-06
99	2	8/8	185	0	1.0e+02	7.2e-09	-2.7600e+00	1.3e-06
110	2	7/7	158	0	1.0e+02	5.7e-09	-3.7000e+00	5.9e-07

EPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
22	5	15/15	204	0	1.0e+01	3.1e-12	-1.2784e+00	1.4e-06
33	4	14/14	203	0	1.0e+01	4.5e-12	-1.9175e+00	6.0e-06
44	6	18/18	266	0	1.0e+01	5.4e-12	-2.2567e+00	9.3e-06
55	7	20/20	323	0	1.0e+01	3.6e-12	-2.2959e+00	3.1e-07
66	4	15/15	236	0	1.0e+01	5.4e-12	-3.8351e+00	8.6e-06
77	5	17/17	259	0	1.0e+01	4.8e-12	-4.4742e+00	2.1e-07
88	5	17/17	259	0	1.0e+01	5.2e-12	-5.1134e+00	1.6e-07
99	5	17/17	262	0	1.0e+01	5.4e-12	-5.7526e+00	2.2e-07
110	5	17/17	261	0	1.0e+01	5.6e-12	-6.3918e+00	1.5e-07

Table 5.3: Nonconvex test-problems satisfying SLICQ.

5.3.2 Nonconvex Test Problems

The nonconvex test problems corresponding to $k_1 = 1$, $k_2 = 1$, $b = 1.8$, $n_1/n = 5$, $n_2/n = 5$ and $n = 2, 3, 4, 5, 6, 7, 8, 9, 10$ were solved using IPD and EPD. The results for test problems satisfying SLICQ ($a = 0.0$) are given in Table 5.3 and the results for test problems satisfying LICQ ($a = 9.6$) are given in Table 5.4. The behavior of IPD and EPD on the nonconvex test problems is similar to their behavior on convex test problems. However, for some of the nonconvex problems, IPD and EPD find different local minimizers. In particular, for all of the test problems satisfying the SLICQ, IPD and EPD find different local minimizers.

IPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
22	36	100/100	961	0	2.5e+07	2.3e-07	1.3114e+02	2.9e-07
33	41	116/116	1298	0	1.0e+07	5.9e-07	1.9641e+02	2.5e-08
44	36	151/151	1874	0	5.0e+07	1.4e-07	2.6228e+02	9.9e-07
55	52	175/175	2312	0	1.0e+07	8.1e-07	3.2785e+02	7.0e-07
66	65	211/211	3288	0	5.0e+07	1.7e-07	3.9342e+02	1.6e-07
77	65	213/213	3822	0	1.0e+07	9.5e-07	4.5869e+02	5.3e-06
88	87	259/259	5160	0	5.0e+07	2.1e-07	5.2396e+02	5.5e-06
99	77	303/303	6044	0	1.0e+08	1.1e-07	5.8983e+02	4.4e-06
110	93	307/307	7194	0	1.0e+08	1.1e-07	6.5360e+02	5.5e-07

EPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
22	40	133/133	1848	0	1.0e+02	1.1e-08	1.3054e+02	1.7e-06
33	69	200/200	2895	0	5.0e+01	5.0e-09	1.9581e+02	4.7e-07
44	69	196/196	2933	0	5.0e+01	1.2e-08	2.6108e+02	2.1e-06
55	76	227/227	3503	0	1.0e+02	3.3e-09	3.2635e+02	3.7e-07
66	129	451/451	7688	0	1.0e+02	3.6e-09	3.9192e+02	6.9e-07
77	102	304/304	4876	0	5.0e+01	1.8e-08	4.5719e+02	4.7e-06
88	107	331/331	5265	0	1.0e+02	4.3e-09	5.2216e+02	4.2e-06
99	131	416/416	7293	0	1.0e+02	4.5e-09	5.8743e+02	4.4e-06
110	144	424/424	7341	0	1.0e+02	4.6e-09	6.5270e+02	4.8e-06

Table 5.4: Nonconvex test problems satisfying only LICQ.

IPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
18	47	145/145	2266	0	1.0e+06	2.0e-07	4.4466e+02	7.1e-007
54	48	153/153	2375	0	1.0e+06	2.0e-07	4.4466e+02	2.3e-006
90	43	144/144	2543	0	1.0e+06	2.0e-07	4.4466e+02	3.3e-006
126	43	136/136	2265	0	1.0e+06	2.0e-07	4.4466e+02	1.5e-006

EPD

n	it	nsub	feval	ifl	gamma	feasib	objective	grad
18	71	198/198	3178	0	1.0e+01	8.1e-09	4.4466e+02	3.1e-006
54	71	198/198	3695	0	1.0e+01	8.1e-09	4.4466e+02	3.1e-006
90	71	198/198	5899	0	1.0e+01	8.1e-09	4.4466e+02	3.1e-006
126	71	198/198	13532	0	1.0e+01	8.1e-09	4.4466e+02	3.1e-006

Table 5.5: Effect of coupling among subproblems.

5.3.3 Coupling Among Subproblems

The convex test problems corresponding to $k_1 = 1$, $k_2 = 1$, $n = 6$, $a = 9.6$, and $n_1/n = n_2/n = 1, 4, 7, 10$ were solved by IPD and EPD. In other words, the number of global variables was held constant while the number of local variables was increased. The results are given in Table 5.5.

Note that the number of subproblems needed to find a master problem stationary point is roughly the same for all test problems. This seems to imply that the communication between the master problem and the subproblems required to solve the IPD and EPD master problems is entirely determined by the number of *global* variables. Therefore, IPD and EPD will be efficient decomposition algorithms for OPGVs that have only a few global variables and a possibly large number of local variables.

5.3.4 Observed Convergence Rate

In Chapter 3, we proved that the IPD and EPD master problem minimizers satisfy the SOSC and therefore a superlinear convergence rate can be expected from QNSOL. However, in some cases only a linear convergence rate is achieved. This is explained by the ill conditioning of the IPD and EPD subproblems near the minimizer. This ill conditioning adversely affects the accuracy of the computed subproblem minimizers and hence the precision of the master problem objective function. In the absence of a precise objective function, the master problem iterates might never get inside the region of superlinear convergence. Moreover, the master problem region of fast convergence is likely to be small because the IPD and EPD master problems become ill conditioned as the iterates approach the minimizer.

5.4 Summary

Despite the difficulties encountered in our efforts to prove global convergence for decomposition algorithms for the nonconvex OPGV, our implementation of IPD and EPD solved all test problems tried from random starting points. This seems to imply that the theoretical global convergence difficulties discussed in Chapter 2 may not affect the performance of IPD and EPD in practice.

In Chapter 3, we proved that a superlinear convergence rate can be expected from QNSOL when applied to the IPD and EPD master problems. However, in our experiments we observe that, in some cases, the actual convergence rate achieved is only linear because of the ill conditioning of the IPD and EPD master problems and subproblems.

The communication required between the master problem and the subproblems seems to depend only on the number of global variables. Also, those test problems satisfying only LICQ are much more difficult to solve than the test problems satisfying SLICQ. Finally, our experiments show that our IPD implementation is slightly more efficient on the OPGV test-problem set than our EPD implementation (which could be explained by the quadratic nature of the test-problem set).

Note that CO cannot be run on the OPGV test-problem set because it is designed for OPGVs whose objective function depends exclusively on the global variables. However, IPD and EPD successfully solved the two sample test problems proposed by Alexandrov and Lewis [AL00] on which CO failed because of the singularity of the CO master problem Jacobian. IPD and EPD avoided this difficulty by solving a sequence of *unconstrained* master problems.

Chapter 6

Conclusions and Future Research

6.1 Local Convergence

In Chapter 3, we prove fast local convergence for conventional optimization algorithms for smooth problems when applied to the IPD (EPD) master problem for each value of the penalty (barrier) parameter. Hence we overcome the degeneracy difficulties associated with CO on the one hand, and, on the other hand, we relax the assumptions made by Tammer [Tam87] because we just assume the LICQ instead of the restrictive SLICQ.

A natural extension of our fast local convergence result would be proving that a fast convergence rate can still be achieved while simultaneously updating the penalty (barrier) parameter. The result would follow if we could prove that once the iterates are close to the minimizer, the penalty (barrier) parameter can be updated at a superlinear or quadratic rate and only one iteration suffices to solve the master problem to the degree of accuracy required. For EPD the result would be an extension of previous theory developed for primal-dual methods [GOST00, VTZ99].

6.2 Global Convergence

Despite the fact that there does not exist any global convergence proof for decomposition algorithms for the nonconvex OPGV, our numerical results indicate that this might not be a difficulty in practice. In either case, we believe that the application of parametric optimization algorithms [GVJ90] to solve the subproblems could improve the global convergence behavior of IPD and EPD.

6.3 Computational Results

In Chapter 5, we described our implementation of IPD and EPD and applied both algorithms to solve the new OPGV test-problem set introduced in Chapter 4. Both IPD and EPD successfully solved a variety of convex and nonconvex test problems corresponding to different degrees of degeneracy and coupling among systems. IPD was found to be more efficient than EPD and we conjectured that this could be attributed to the quadratic nature of the OPGV test-problem set. We plan to confirm our hypothesis by replacing the linear constraints in the OPGV test-problem set by nonlinear constraints and running the experiments again.

We noticed that, in some instances, the ill conditioning of the IPD and EPD master problems prevents QNSOL from achieving superlinear convergence. Another interesting topic of future research is the development of specialized algorithms capable of dealing with this ill conditioning. The work developed for primal-dual methods could be vital in developing such algorithms for the EPD master problem.

Appendix A

Optimality Conditions

We consider a general nonlinear optimization problem of the form

$$\begin{array}{lrcl} \min\limits_{x} & F(x) & & \\ s.t. & c(x) & \geq & 0, \\ & d(x) & = & 0, \end{array} \tag{A.1}$$

where $x \in \mathbb{R}^n$, $c(x) : \mathbb{R}^n \to \mathbb{R}^m$, and $d(x) : \mathbb{R}^n \to \mathbb{R}^p$.

Definition A.1 A neighborhood of a point x^* is $\mathcal{N}_\epsilon(x) = \{x : \|x - x^*\| < \epsilon\}$, where $\epsilon > 0$.

Definition A.2 The *feasible set* is $\Omega = \{x : c(x) \geq 0, d(x) = 0\}$.

Definition A.3 Any nonempty set of the form $L(F_0) = \{x : c(x) \geq 0, d(x) = 0$ and $F(x) \leq F_0\}$ is a *level set* of problem A.1.

Definition A.4 A point x^* is a *local minimizer* of A.1 if $x^* \in \Omega$ and there exists $\epsilon > 0$ such that if $x \in \mathcal{N}_\epsilon(x^*) \cap \Omega$ then $F(x^*) \leq F(x)$.

Definition A.5 A point x^* is an *isolated local minimizer* of A.1 if $x^* \in \Omega$ and there exists $\epsilon > 0$ such that x^* is the only local minimizer in $\mathcal{N}_\epsilon(x^*) \cap \Omega$.

Definition A.6 For a given a feasible point x, the set of *active inequality constraints* is $B = \{i : c_i(x) = 0\}$.

Definition A.7 The *Lagrangian function* $\mathcal{L}(x, \lambda, \nu)$ is

$$\mathcal{L}(x, \lambda, \nu) = F(x) - c(x)^T \lambda + d(x)^T \nu.$$

Definition A.8 The *Linear Independence Constraint Qualification* (LICQ) holds at a feasible point x if the gradients of all active inequality constraints ($\nabla c_i(x)$, all $i \in B$) and the gradients of all equality constraints ($\nabla d_i(x)$, $i = 1 : p$) are linearly independent.

Theorem A.9 (First-order necessary conditions) *Provided F, c and d are differentiable at x^* and the LICQ holds at x^*, if x^* is a local minimizer of problem A.1 then there exist vectors λ^* and ν^* such that*

$$c(x^*) \geq 0, \tag{A.2}$$

$$d(x^*) = 0, \tag{A.3}$$

$$c(x^*)^T \lambda^* = 0, \tag{A.4}$$

$$\lambda^* \geq 0, \tag{A.5}$$

$$\nabla \mathcal{L}(x^*, \lambda^*, \nu^*) = 0. \tag{A.6}$$

Definition A.10 The triple (x, λ, ν) is a *first-order KKT point* for problem A.1 if it satisfies conditions A.2–A.6.

Definition A.11 The *Strict Complementarity Slackness Conditions* (SCSC) hold at a first-order KKT point (x^*, λ^*, ν^*) for problem A.1 if exactly one of λ_i^* and $c_i(x^*)$ is zero for each $i = 1 : m$.

Definition A.12 For a given first-order KKT point (x^*, λ^*, ν^*) for problem A.1, the *tangent cone* $\mathcal{T}(x^*, \lambda^*, \nu^*)$ is the set of vectors τ such that

$$\begin{aligned} \tau^T \nabla c_i(x^*) &= 0, \qquad \forall i \in D \equiv \{i : \lambda_i^* > 0\}, \\ \tau^T \nabla c_i(x^*) &\geq 0, \qquad \forall i \in B - D, \\ \tau^T \nabla d_i(x^*) &= 0, \qquad \forall i = 1{:}p. \end{aligned}$$

Definition A.13 For a given first-order KKT point (x^*, λ^*, ν^*) for problem A.1, the *strict tangent cone* $\hat{\mathcal{T}}(x^*, \lambda^*, \nu^*)$ is the set of vectors τ such that

$$\begin{aligned} \tau^T \nabla c_i(x^*) = 0, &\quad \forall i \in D \equiv \{i : \lambda_i^* > 0\}, \\ \tau^T \nabla d_i(x^*) = 0, &\quad \forall i = 1{:}p. \end{aligned}$$

Theorem A.14 (Second-order sufficient conditions (SOSC)) *Second-order sufficient conditions that a point x^* be an isolated local minimizer when F, c and d are twice differentiable at x^*, are that there exist vectors λ^* and ν^* satisfying conditions A.2–A.6 and for every τ in the* tangent cone $\mathcal{T}(x^*, \lambda^*, \nu^*)$,

$$\tau^T \nabla^2 \mathcal{L}(x^*, \lambda^*, \nu^*) \tau > 0.$$

Theorem A.15 (Strong second-order sufficient conditions (SSOSC)) *Strong second-order sufficient conditions that a point x^* be an isolated local minimizer when F, c and d are twice differentiable at x^*, are that there exist vectors λ^* and ν^* satisfying conditions A.2–A.6 and for every τ in the* strict tangent cone $\hat{\mathcal{T}}(x^*, \lambda^*, \nu^*)$,

$$\tau^T \nabla^2 \mathcal{L}(x^*, \lambda^*, \nu^*) \tau > 0.$$

Note that the Optimization Problem with Global Variables (OPGV) (1.1) is just a particular case of problem A.1. Thus, all definitions and theorems given above may be applied to the OPGV. In addition, we make use of the following constraint qualification for the OPGV.

Definition A.16 The *Strong Linear Independence Constraint Qualification* (SLICQ) holds for the OPGV at a feasible point $(x, y_1, \ldots, y_N)$ if for $i = 1{:}N$ the matrix $B_i = \nabla_{y_i} \hat{c}_i(x, y_i)$, where $\hat{c}_i$ are the active constraints, has full rank.

Bibliography

[AF90] J.P. Aubin and H. Frankowska. *Set-Valued Analysis*. Birkhauser, Boston, 1990.

[AK98] N.M. Alexandrov and S. Kodiyalam. Initial results of an MDO method evaluation study. In *Seventh AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 1998. AIAA Paper 98-4884.

[AL99] N.M. Alexandrov and R.M. Lewis. Comparative properties of collaborative optimization and other approaches to MDO. In *Proc., First ASMO UK/ISSMO conference on Engineering Design Optimization*. MCB Press, Jul 1999.

[AL00] N.M. Alexandrov and R.M. Lewis. Analytical and computational aspects of collaborative optimization. Technical Report TM-2000-210104, NASA, 2000.

[Bar98] J.F. Bard. *Practical Bilevel Optimization: Algorithms and Applications*. Kluwer Academic Publishers, Boston, 1998.

[BDG+87] J.R. Birge, M.A.H. Dempster, H.I. Gassman, E.A. Gunn, A.J. King, and S.W. Wallace. A standard input format for multiperiod stochastic linear programs. *COAL Newsletter*, 17:1–20, 1987.

[Ben62] J.F. Benders. Partitioning procedures for solving mixed variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.

[BL97] J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming.* Springer-Verlag, New York, 1997.

[Bra96] R.D. Braun. *Collaborative Optimization: An Architecture for Large-Scale Distributed Design.* PhD thesis, Stanford University, 1996.

[CDF+94] E.J. Cramer, J.E. Dennis, P.D. Frank, R.M. Lewis, and G.R. Shubin. Problem formulation for multidisciplinary optimization. *SIAM J. Optimization*, 4(4):754–776, November 1994.

[CV94] P.H. Calamai and L.N. Vicente. Generating quadratic bilevel programming test problems. *ACM Transactions on Mathematical Software*, 20(1):103–119, March 1994.

[Dan63] G.B. Dantzig. *Linear Programming and Extensions.* Princeton University Press, Princeton, 1963.

[dM00] A.V. de Miguel and W. Murray. An analysis of collaborative optimization methods. In *Eight AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 2000. AIAA Paper 00-4720.

[DT] G.B. Dantzig and M.N. Thapa. Linear programming 3: Extensions. Manuscript in progress.

[DW60] G.B. Dantzig and P. Wolfe. Decomposition principle for linear programs. *Operations Research*, 8(1):101–111, 1960.

[EdGP96] L.F. Escudero, J.L. de la Fuente, C. Garcia, and F.J. Prieto. Hydropower generation management under uncertainty via scenario analysis and parallel computation. *IEEE Transactions on Power Systems*, 11(2):683–689, 1996.

[FK93] O.E. Flippo and A.H.G. Rinnooy Kan. Decomposition in general mathematical programming. *Mathematical Programming*, 60:361–382, 1993.

[FL95] J.E. Falk and J. Liu. On bilevel programming, part I: general nonlinear cases. *Mathematical Programming*, 70:47–72, 1995.

[FM68] A.V. Fiacco and G.P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley & Sons, New York, 1968.

[Geo72] A.M. Geoffrion. Generalized Benders decomposition. *JOTA*, 10(4):237–260, 1972.

[GMS97] P.E. Gill, W. Murray, and M.A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. Technical Report SOL 97-3, Dept. of EESOR, Stanford University, 1997. To appear in *SIAM J. Optimization*.

[GMSW86] P.E. Gill, W. Murray, M.A. Saunders, and M.H. Wright. User's guide for NPSOL (version 4.0). Technical Report SOL 86-2, Dept. of Operations Research, Stanford University, 1986.

[GMW81] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Academic Press, New York, 1981.

[Gol56] A.J. Goldman. Resolution and separation theorems for polyhedral convex sets. In H.W. Kuhn and A.W. Tucker, editors, *Linear Inequalities and Related Systems*, volume 38 of *Anals of Mathematics*, pages 41–51. Princeton University Press, Princeton, 1956.

[GOST00] N.I.M. Gould, D. Orban, A. Sartenaer, and P.L. Toint. Superlinear convergence of primal-dual interior point algorithms for nonlinear programming. Technical Report TR/PA/00/20, CERFACS, 2000.

[GVJ90] J. Guddat, F. Guerra Vazquez, and H.Th. Jongen. *Parametric Optimization: Singularities, Pathfollowing, and Jumps*. John Wiley & Sons, New York, 1990.

[HSS97] R.T. Haftka and J. Sobieszczanski-Sobieski. Multidisciplinary aerospace design optimization: Survey of recent developments. *Structural Optimization*, 14:1–23, 1997.

[HUL93] J. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms*. Springer-Verlag, New York, 1993.

[Inf94] G. Infanger. *Planning Under Uncertainty: Solving Large-Scale Stochastic Linear Programs*. Boyd and Fraser, Danvers, Mass., 1994.

[JR99] H. Jiang and D. Ralph. QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints. *Computational Optimization and Applications*, 13:25–59, 1999.

[Kro97] I.M. Kroo. MDO for large-scale design. In N.M. Alexandrov and M.Y. Hussaini, editors, *Multidisciplinary Design Optimization: State of the Art*, 1997.

[Man99] V.M. Manning. *Large-Scale Design of Supersonic Aircraft Via Collaborative Optimization*. PhD thesis, Stanford University, 1999.

[Mif77] R. Mifflin. An algorithm for constrained optimization with semismooth functions. *Mathematics of Operations Research*, 2(2):191–207, 1977.

[MP95] W. Murray and F.J. Prieto. A sequential quadratic programming algorithm using an incomplete solution of the subproblem. *SIAM J. Optimization*, 5(3):590–640, 1995.

[Mur71] W. Murray. Analytical expressions for the eigenvalues and eigenvectors of the Hessian matrices of barrier and penalty functions. *JOTA*, 7:189–196, 1971.

[Mur97] W. Murray. Sequential quadratic programming methods for large-scale problems. *Computational Optimization and Applications*, 7:127–142, 1997.

[NW99] J. Nocedal and S.J. Wright. *Numerical Optimization.* Springer-Verlag, New York, 1999.

[OR70] J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, 1970.

[PAG96] S.L. Padula, N.M. Alexandrov, and L.L. Green. MDO test suite at NASA Langley Research Center. In *Sixth AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 1996. AIAA Paper 96-4028.

[SIB97] K. Shimizu, Y. Ishizuka, and J.F. Bard. *Nondifferentiable and Two-Level Mathematical Programming.* Kluwer Academic Publishers, Boston, 1997.

[Sob98] I.P. Sobieski. *Multidisciplinary Design Using Collaborative Optimization.* PhD thesis, Stanford University, 1998.

[SZ92] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM J. Optimization*, 2:121–152, 1992.

[Tam87] K. Tammer. The application of parametric optimization and imbedding for the foundation and realization of a generalized primal decomposition approach. In J. Guddat, H. Jongen, B. Kummer, and F. Nozicka, editors, *Parametric Optimization and Related Topics*, volume 35 of *Mathematical Research.* Akademie-Verlag, Berlin, 1987.

[TW81] J. Tind and L.A. Wolsey. An elementary survey of general duality theory in mathematical programming. *Mathematical Programming*, 21(3):241–261, 1981.

[VTZ99] M.C. Villalobos, R.A. Tapia, and Y. Zhang. The sphere of convergence of Newton's method on two equivalent systems from nonlinear programming. Technical Report TR-99-15, Department of Computational and Applied Mathematics, Rice University, 1999.

[Wri99] M.H. Wright. Ill-conditioning and computational error in interior methods for nonlinear programming. *SIAM J. Optimization*, 9:84–111, 1999.