A Local Convergence Analysis of Bilevel Decomposition Algorithms

Angel-Victor DeMiguel
Decision Sciences
London Business School
avmiguel@london.edu

Walter Murray
Management Science and Engineering
Stanford University
walter@stanford.edu

4th December 2003

Decomposition algorithms exploit the structure of large-scale optimization problems by breaking them into a set of smaller subproblems and a coordinating master problem. Cutting-plane methods have been extensively used to decompose *convex* problems. In this paper, however, we focus on certain *nonconvex* problems arising in engineering. Engineers have been using *bilevel decomposition algorithms* to tackle these problems. These algorithms use nonlinear optimization techniques to solve both the master problem and the subproblems. In this paper, we propose two new bilevel decomposition algorithms and show that they converge locally at a superlinear rate. Our computational experiments illustrate the numerical performance of the algorithms.

Key words. Decomposition algorithms, bilevel programming, nonlinear programming, multidisciplinary design optimization.

1 Introduction

Decomposition algorithms exploit the structure of loosely-coupled large-scale optimization problems by breaking them into a set of smaller independent *subproblems*. Then, a so-called *master problem* is used to coordinate the subproblem solutions and find an overall problem minimizer.

A class of large-scale optimization problems that has been extensively analyzed is the *stochastic programming problem* [24, 5]. These problems arise, for instance, when a discrete number of *scenarios* is used to model the uncertainty of some of the variables. Usually, a large number of scenarios must be considered in order to represent the uncertainty appropriately. Decomposition algorithms exploit the structure of the stochastic program by breaking it into a set of smaller (and tractable) subproblems, one per scenario.

Two main types of decomposition algorithms have been proposed for the stochastic programming problem: cutting-plane algorithms and augmented Lagrangian algorithms. Cutting-plane methods use convex duality theory to build a linear approximation to the master problem [4, 35, 20, 30]. Augmented Lagrangian approaches, on the other hand, use an estimate of the Lagrange multipliers to decompose the stochastic program into a set of subproblems. Then, the subproblem minimizers are used to update the current estimate of the Lagrange multipliers [29, 31].

In this paper we focus on a different class of large-scale problems that remains largely unexplored by the operations research community: the *Multidisciplinary Design Optimization* (MDO) problem [1, 10]. These problems usually arise in engineering design projects that require the collaboration of several departments within a company. For example, when designing an airplane, one department may be in charge of the analysis of the aerodynamic system and another in charge of the structural system. Typically 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 may be impractical. In this context, decomposition algorithms are needed to allow the different departments to find an overall optimal design while keeping the amount of communication required between them limited.

At first sight, it may seem reasonable to apply cutting-plane or augmented Lagrangian decomposition methods, available for stochastic programming, to decompose the MDO problem. Unfortunately, this may not be a good idea. While most real-world stochastic programs are linear, or, at least, convex, most real-world MDO problems are *nonconvex* nonlinear problems. This precludes the use of cutting-plane methods, which rely heavily on convex duality theory, to solve MDO problems. One may feel tempted to use augmented Lagrangian methods to decompose the MDO problem. Unfortunately, it is well-known that these algorithms may converge slowly in practice (see [23, 9] and the numerical results in Section 6).

Pressed by the need to solve MDO problems, engineers have turned to *bilevel decomposition algorithms*. Once decomposed into a master problem and a set of subproblems, the MDO is a particular type of *bilevel program* [25, 32, 16]. Bilevel decomposition algorithms apply nonlinear optimization techniques to solve both the master problem and the subproblems. At each iteration of the algorithm solving the master problem, each of the subproblems is solved, and their minimizers used to compute the master problem derivatives and their associated Newton direction.

Braun [6] proposed a bilevel decomposition algorithm known as collaborative optimization (CO). The algorithm uses an inexact (quadratic) penalty function to decompose the MDO problem. Unfortunately, it is easy to show that both the CO master problem and subproblems are degenerate [12]. Nondegeneracy is a common assumption when proving local convergence for most optimization algorithms. Not surprisingly, CO fails to converge on some simple test problems even when the starting point is very close to the minimizer [2]. Despite these difficulties, the algorithm has been applied to some real-world problems [7, 33, 26].

In this paper, we propose two bilevel decomposition algorithms closely related to CO and show that their master problem and subproblems are nondegenerate. As a result, we show that the proposed algorithms have better local convergence properties than CO. This is of great importance in the context of the MDO problem. In particular, engineers often know good starting points (from previous designs) and thus their decomposition algorithms usually achieve global convergence. But, due to the high computational expense usually associated with MDO function evaluations, it is crucial that the iterates generated by the algorithm converge at a fast rate in the vicinity of the minimizer.

The first algorithm is termed Inexact Penalty Decomposition (IPD) and uses an inexact penalty function. The second algorithm is termed Exact Penalty Decomposition (EPD) and employs an exact penalty function in conjunction with a barrier term. Our main contribution is to show that the master problem and subproblems corresponding to the proposed decomposition algorithms are nondegenerate. Thus, we can apply standard convergence results to show that the decomposition algorithms converge locally at a superlinear rate for each value of the penalty and barrier parameters. The question remains whether the ill-conditioning usually associated with the penalty and barrier functions may hinder the practical performance of IPD and EPD. Our computational experiments, however, show that the numerical performance of both algorithms is good.

This paper is organized as follows. In Section 2, we give the mathematical statement of the MDO problem. In Section 3, we describe collaborative optimization and point out its difficulties. In Section 4, we introduce IPD and analyze its local convergence properties. Section 5 is devoted to EPD. Finally, we show the results of our computational experiments in Section 6.

2 Problem Statement

Multidisciplinary Design Optimization (MDO) problems arise in engineering design projects that require the participation of several departments or groups within a corporation. Each of the groups is usually in charge of the design of one of the systems that compose the overall product. The number of design parameters that affect all of the systems is often very small. Thus, the MDO problem can be seen as the aggregation of the objective and constraint functions corresponding to a number of systems that are coupled only through a few of the variables known as *global variables*. Mathematically, the MDO may be formulated as

$$\begin{array}{lrcl} \min\limits_{x,y_1,\ldots,y_N} & F_1(x,y_1)+F_2(x,y_2)+\ldots+F_N(x,y_N) & & \\ s.t. & c_1(x,y_1) & \geq & 0 \\ & c_2(x,y_2) & \geq & 0 \\ & & \vdots & \\ & c_N(x,y_N) & \geq & 0, \end{array} \tag{1}$$

where $x \in \mathbb{R}^n$ are the *global* variables, $y_i \in \mathbb{R}^{n_i}$ are the ith system *local* variables, $c_i : \mathbb{R}^n \times \mathbb{R}^{n_i} \to \mathbb{R}^{m_i}$ are the ith system constraints, and $F_i : \mathbb{R}^n \times \mathbb{R}^{n_i} \to \mathbb{R}$ is the objective function term corresponding to the ith system. Note that while global variables appear in all of the constraints and objective function terms, local variables only appear in the objective function term and constraints corresponding to one of the systems.

3 Decomposition Algorithms

The structure of the MDO problem suggests it may be advantageously broken into N independent subproblems, one per system. In this section, we explain how bilevel decomposition algorithms in general, and collaborative optimization in particular, decompose problem (1).

3.1 Bilevel Decomposition Algorithms

A bilevel decomposition algorithm divides the job of finding a minimizer to problem (1) into two different tasks: (i) finding an optimal value of the local variables $(y_1^*(x), y_2^*(x), \dots, y_N^*(x))$ for a given value of the global variables x, and (ii) finding an overall optimal value of the global variables x^*. The first task is performed by solving N subproblems. Then the subproblem solutions are used to define a master problem whose solution accomplishes the second task. A general bilevel decomposition algorithm may be described as follows. Solve the following *master problem*:

$$\min_{x} \quad F_1^*(x) + F_2^*(x) + \cdots + F_N^*(x),$$

where $F_i^*(x) = F_i(x, y_i^*(x))$ for $i = 1 : N$ are the subproblem optimal value functions

$$F_1^*(x) = \begin{array}{ll} \min\limits_{y_1} & F_1(x, y_1) \\ s.t. & c_1(x, y_1) \geq 0 \end{array} \qquad \cdots \qquad F_N^*(x) = \begin{array}{ll} \min\limits_{y_N} & F_N(x, y_N) \\ s.t. & c_N(x, y_N) \geq 0 \end{array}$$

Note that the global variables are just a parameter within each of the subproblems defined above. Consequently, the subproblems are independent from each other and may be solved in parallel. Moreover, the master problem only depends on the global variables. Thus, the above formulation allows the different systems to be dealt with almost independently, and only a limited amount of information regarding the global variables is exchanged between the master problem and the subproblems.

After breaking the MDO into a master problem and a set of subproblems, a bilevel decomposition algorithm applies a nonlinear optimization method to solve the master problem. At each iteration, a new estimate of the global variables x_k is generated and each of the subproblems is solved using x_k as a parameter. Then, sensitivity analysis formulae [18] are used to compute the master problem objective and its derivatives from the subproblem minimizers. Using this information, a new estimate of the global variables x_{k+1} is computed. This procedure is repeated until a master problem minimizer is found.

Tammer [34] shows that the general bilevel programming approach stated above converges locally at a superlinear rate assuming that the MDO minimizer satisfies the so-called strong linear independence constraint qualification; that is, assuming that at the minimizer $(x, y_1, \ldots, y_N)$, the vectors $\nabla_{y_i} c_i(x, y_i)$ corresponding to the active constraints are linearly independent. Roughly speaking, the SLICQ implies that for *any* value of the global variables in a neighborhood of x, there exist values of the local variables for which all of the constraints are satisfied. As the name implies this is a restrictive assumption and is in a sense equivalent to assuming (at least locally) that the problem is an unconstrained one in the global variables only. DeMiguel and Nogales [15] propose a bilevel decomposition algorithm based on interior-point techniques that also converges superlinearly under the SLICQ.

But in this paper we focus on the case where only the weaker (and more realistic) linear independence constraint qualification (LICQ) holds; that is, the case where the gradients $\nabla_{x,y_1,\ldots,y_N} c_i(x, y_i)$ corresponding to the active constraints are linearly independent. In this case, a major difficulty in the general bilevel programming approach stated above is that the algorithm breaks down whenever one of the subproblems is infeasible at one of the master problem iterates. In particular, the algorithm will fail if for a given master problem iterate x_k and for at least one of the systems $j \in \{1, 2, \ldots, N\}$, there does not exist y_j such that $c_j(x_k, y_j) \geq 0$. Unlike when SLICQ holds, when only LICQ holds this difficulty may arise even when the iterates are in the vicinity of the minimizer.

3.2 Collaborative Optimization (CO)

To overcome this difficulty, Braun [6] proposed a bilevel decomposition algorithm for which the subproblems are always feasible, provided the original MDO problem is feasible. To do so, Braun allows the global variables to take a different value x_i within each of the subproblems. Then, the global variables for all systems (x_i for $i = 1 : N$) are enforced to converge to the same value (given by the *target* variable vector z) by using quadratic penalty functions. The resulting algorithm may be stated as follows. Solve the following master problem

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

where $p_i^*(z)$ for $i = 1 : N$ are the ith subproblem optimal-value functions:

$$\begin{array}{lll} p_1^*(z) = & \min\limits_{x_1, y_1} & \|x_1 - z\|_2^2 \\ & s.t. & c_1(x_1, y_1) \geq 0. \end{array} \quad \cdots \quad \begin{array}{lll} p_N^*(z) = & \min\limits_{x_N, y_N} & \|x_N - z\|_2^2 \\ & s.t. & c_N(x_N, y_N) \geq 0, \end{array}$$

where y_i and c_i are as defined in (1), the global variables $x_1, x_2, \ldots, x_N \in \mathbb{R}^n$, the target variables $z \in \mathbb{R}^n$. Finally, it is assumed that the objective function $\sum_i F_i(x, y_i)$ takes the specific form $\sum_i F_i(x)$. Note that this assumption is crucial because otherwise the master problem would also depend on the local variables. Note also that, at a solution z^* to the master problem, $p_i^*(z^*) = 0$ and therefore $x_i^* = z^*$ for $i = 1{:}N$.

The main advantage of CO is that, provided problem (1) is feasible, the subproblems are feasible for any value of the target variables. Unfortunately, it is easy to show that its master problem and the subproblems are *degenerate*; that is, their minimizer does not satisfy the linear independence constraint qualification, the strict complementarity slackness conditions, and the second order sufficient conditions (see Appendix A for a definition of these nondegeneracy conditions). The degeneracy of the subproblems is easily deduced from the form of the subproblem objective gradient:

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

At the solution, $x_i^* = z^*$ and therefore the subproblem objective gradient is zero. Given that the original problem (1) satisfies the linear independence constraint qualification, this in turn implies that the subproblem Lagrange multipliers are zero and therefore the strict complementarity slackness conditions do not hold at the subproblem minimizer. Thus, the subproblems are degenerate at the solution.

Moreover, the degeneracy of the subproblems implies that the optimal-value functions $p_i^*(z)$ are not smooth in general [18, Theorem 12] and thus the master problem is a nonsmooth optimization problem. Nondegeneracy and smoothness are common assumptions when proving local convergence for most optimization algorithms [27]. In fact, Alexandrov and Lewis [2] give a linear and a quadratic programming problem on which CO failed to converge to a minimizer even from starting points close to a minimizer.

4 Inexact Penalty Decomposition

In this section, we propose a decomposition algorithm closely related to collaborative optimization. Like CO, our algorithm makes use of an inexact (quadratic) penalty function. Our formulation, however, overcomes the principle difficulties associated with CO. We term the algorithm Inexact Penalty Decomposition (IPD).

The rest of this section is organized as follows. In Section 4.1, we state the proposed master problem and subproblems. In Section 4.2, we describe the optimization algorithms used to solve the master problem and the subproblems. In Section 4.3, we show that, under standard nondegeneracy assumptions on the MDO minimizer, the proposed master problem and subproblems are also nondegenerate. Using these nondegeneracy results, we show in Section 4.4 that the optimization algorithms used to solve the master problem and the subproblems converge locally at a superlinear rate for each value of the penalty parameter.

4.1 Problem Formulation

Our formulation shares two features with collaborative optimization: (i) it allows the global variables to take a different value within each of the subproblems, and (ii) it uses quadratic penalty functions to enforce the global variables to asymptotically converge to the target variables. However, unlike CO, our formulation explicitly includes a penalty parameter γ that allows control over the speed at which the global variables converge to the target variables. In particular, we propose solving the following master problem for a sequence of penalty parameters $\{\gamma_k\}$ such that $\gamma_k \to \infty$:

$$\min_{z} \quad \sum_{i=1}^{N} F_i^*(\gamma_k, z), \tag{2}$$

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

$$\begin{aligned} F_i^*(\gamma_k, z) \quad = \quad & \min_{x_i, y_i} \quad F_i(x_i, y_i) + \gamma_k \|x_i - z\|_2^2 \\ & s.t. \quad c_i(x_i, y_i) \geq 0, \end{aligned} \tag{3}$$

and F_i, c_i, and y_i are as in problem (1), $x_i \in \mathbb{R}^n$ are the ith system global variables, and $z \in \mathbb{R}^n$ are the target variables. Note that, because quadratic penalty functions are inexact for any finite value of γ_k, we have to solve the above master problem for a sequence of penalty parameters $\{\gamma_k\}$ such that $\gamma_k \to \infty$ in order to recover the exact solution $x_i^* = z^*$. Finally, to simplify notation, herein we omit the dependence of the subproblem optimal value function on the penalty parameter; thus we refer to $F_i^*(\gamma_k, z)$ as $F_i^*(z)$.

Note that, unlike CO, IPD uses the subproblem optimal-value functions $F_i^*(z)$ as penalty terms within the objective function of an *unconstrained* master problem. In CO, constraints are included in the master problem setting the penalty functions to zero. In essence, this is equivalent to an IPD approach in which the penalty parameter is set to a very large value from the very beginning. In contrast, IPD allows the user to drive the penalty parameter to infinity gradually. This is essential to the development of a suitable convergence theory.

4.2 Algorithm Statement

In the previous section, we proposed solving the sequence of master problems corresponding to a sequence of increasing penalty parameters $\{\gamma_k\}$. Consequently, we need an optimization algorithm capable of solving the proposed master problem for each value of the penalty parameter γ_k. In addition, to evaluate the master problem objective function, we need an algorithm capable of solving the subproblems at a given value of the target variables z.

To solve the master problem, we use a BFGS quasi-Newton method [22, 19]. These methods use the objective function gradient to progressively build a better approximation of the Hessian matrix. At each iteration, the quasi-Newton method generates a new estimate of the target variables z_k. Then, all of the subproblems are solved with $z = z_k$ and the master problem objective and its gradient are computed from the subproblem minimizers. Using this information, the master problem Hessian is updated and a new estimate of the target variables z_{k+1} is generated. This procedure is repeated until a master problem minimizer is found.

To evaluate the master problem objective function and its gradient, we need to solve the subproblems. To do this, we use the sequential quadratic programming algorithm NPSOL [21]. Then, one can compute the master problem objective and its gradient from the subproblem minimizers. In particular, the master problem objective $F^*(z)$ can be evaluated at the subproblem solution $(x_i(z), y_i(z))$ as

$$F^*(z) = \sum_{i=1}^{N} F_i^*(z) = \sum_{i=1}^{N} [F_i(x_i(z), y_i(z)) + \gamma \|x_i(z) - z\|_2^2]. \tag{4}$$

Moreover, in Section 4.3 we will show that, provided z is close to the master problem minimizer z^*, the subproblem minimizers satisfy the nondegeneracy conditions A.1-A.3. This implies [18, Theorem 6] that the master problem objective is differentiable. Furthermore, it is easy to compute the master problem gradient $\nabla_z F^*(z)$ from the subproblem minimizer. To see this, note that by the complementarity conditions we have that

$$F_i^*(z) = F_i(x_i(z), y_i(z)) + \gamma \|x_i(z) - z\|_2^2 = L_i(x_i(z), y_i(z), \lambda_i(z), z)$$

where L_i is the ith subproblem Lagrangian function, $(x_i(z), y_i(z))$ is the minimizer to the ith subproblem as a function of z, and $\lambda_i(z)$ are the corresponding Lagrange multipliers. Then applying the chain rule we have that

$$\frac{dL_i(x_i(z), y_i(z), \lambda_i(z), z)}{dz} = \nabla_{x_i} L_i(x_i(z), y_i(z), \lambda_i(z), z)\ x_i'(z) \tag{5}$$

$$+ \nabla_{y_i} L_i(x_i(z), y_i(z), \lambda_i(z), z)\ y_i'(z) \tag{6}$$

$$+ \nabla_{\lambda_i} L_i(x_i(z), y_i(z), \lambda_i(z), z)\ \lambda_i'(z), \tag{7}$$

$$+ \nabla_z L_i(x_i(z), y_i(z), \lambda_i(z), z), \tag{8}$$

where $x_i'(z)$, $y_i'(z)$, and $\lambda_i'(z)$ denote the Jacobian matrices of $x_i(z)$, $y_i(z)$, and $\lambda_i(z)$ evaluated at z. Note that (5) and (6) are zero because of the optimality of $(x_i(z), y_i(z))$, (7) is zero by the feasibility and strict complementarity slackness of $(x_i(z), y_i(z))$, and $\nabla_z L_i(x_i(z), y_i(z), \lambda_i(z), z) = -2\gamma(x_i(z) - z)$. Thus, we can write the master problem objective gradient as

$$\nabla F^*(z_k) = \sum_{i=1}^{N} \nabla F_i^*(z_k) = -2\gamma \sum_{i=1}^{N} (x_{ik} - z_k). \tag{9}$$

The resulting decomposition algorithm is outlined in Figure 1.

Initialization: Initialize γ and set the iteration counter $k := 0$. Choose a starting point z_0. For $i = 1 : N$, call NPSOL to find the subproblem minimizers (x_{i0}, y_{i0}). Set the initial Hessian approximation B_0 equal to the identity. Choose $\sigma \in (0, 1)$.

while $(\sum_{i=1}^{N} \|x_{ik} - z_k\|_2^2/(1 + \|z_k\|) > \epsilon)$

1. **Penalty parameter update:** increase γ.
2. **Solve master problem** with current γ:
   repeat
   (a) Function evaluation: for $i = 1 : N$ call NPSOL with γ and compute $F^*(z_k)$ and $\nabla F^*(z_k)$ from (4) and (9).
   (b) Search direction: Solve $B_k \Delta z = -\nabla F^*(z_k)$.
   (c) Backtracking line search:
   $\alpha := 1$.
   while $(F^*(z_k + \alpha \Delta z) - F^*(z_k) > \sigma \nabla F^*(z_k)^T \Delta z)$
   $\alpha := \alpha/2$. Call NPSOL to evaluate $F^*(z_k + \alpha \Delta z)$.
   endwhile
   $s := \alpha \Delta z$, $y := \nabla F^*(z_k + s) - \nabla F^*(z_k)$.
   $z_{k+1} := z_k + s$.
   (d) BFGS Hessian update: $B_{k+1} := B_k - \frac{B_k s s^T B_k}{s^T B_k s} + \frac{y y^T}{y^T s}$
   (e) k = k+1.
   until $(\|\nabla F^*(z_k)\|/|1 + F^*(z_k)| < \epsilon)$

endwhile

Figure 1: Inexact penalty decomposition

Note that the penalty parameter is increased in Step 1 until the global variables are close enough to the target variables; that is, until $\sum_{i=1}^{N} \|x_{ik} - z_k\|_2^2/(1 + \|z_k\|) < \epsilon$, where ϵ is the optimality tolerance. Note also that, for each value of the penalty parameter, the master problem is solved using a BFGS quasi-Newton method (Step 2). Finally, in order to compute the master problem objective function with the precision necessary to ensure fast local convergence, a tight optimality tolerance must be used to solve the subproblems with NPSOL. In our implementation, we use a subproblem optimality tolerance of ϵ^2, where ϵ is the optimality tolerance used to solve the master problem.

4.3 Nondegeneracy Results

In the previous sections, we stated the IPD master problem and subproblems, and proposed optimization algorithms to solve them. In this section we show that, given a nondegenerate MDO minimizer (that is, given an MDO minimizer satisfying the linear independence constraint qualification (LICQ), the strict complementarity slackness conditions, and the second order sufficient conditions; see Appendix A), there exists an equivalent minimizer to the IPD master problem and subproblems that is also nondegenerate. Note that we only assume the LICQ and not the strong linear independence constraint qualification (SLICQ). As mentioned in Section 3, the LICQ is a weaker (and more realistic) assumption than the SLICQ assumed in [34] and [15]. In Section 4.4, we use this result to prove that the algorithms proposed to solve the master problem and the subproblems converge superlinearly for each value of the penalty parameter.

To prove the nondegeneracy result, it is essential to realize that the proposed master problem and subproblems can be derived from the MDO problem through a sequence of three manipulations: (i) introduction of the target variables, (ii) introduction of an inexact penalty function, and (iii) decomposition. The nondegeneracy result follows by showing that the nondegeneracy of an MDO minimizer is preserved by each of these transformations.

The remainder of this section is organized as follows. In Section 4.3.1 we discuss the assumptions and some useful notation. In Sections 4.3.2, 4.3.3, and 4.3.4 we show that the nondegeneracy of an MDO minimizer is preserved by each of the three transformations listed above.

4.3.1 Notation and assumptions

To simplify notation and without loss of generality, herein we consider the MDO problem composed of only two systems. The interested reader is referred to [11, 14] for an exposition with N systems. The MDO problem with two systems is:

$$\begin{array}{lrcl} \min\limits_{x,y_1,y_2} & F_1(x,y_1)+F_2(x,y_2) & & \\ s.t. & c_1(x,y_1) & \geq & 0, \\ & c_2(x,y_2) & \geq & 0, \end{array} \tag{10}$$

where all functions and variables are as defined in (1).

We assume there exists a minimizer (x^*, y_1^*, y_2^*) and a Lagrange multiplier vector $(\lambda_1^*, \lambda_2^*)$ satisfying the KKT conditions for problem (10). In addition, we assume that the functions in problem (10) are three times continuously-differentiable in an open convex set containing w^*. Moreover, we assume that the KKT point $w^* = (x^*, y_1^*, y_2^*, \lambda_1^*, \lambda_2^*)$ is nondegenerate; that is, the linear independence constraint qualification A.1, the strict complementarity slackness conditions A.2, and the second order sufficient conditions A.3 hold at w^*; see Appendix A for a definition of the nondegeneracy conditions A.1-A.3. Note that unlike in the analysis in [34, 15] we only assume the LICQ and not the SLICQ holds at the MDO minimizer.

4.3.2 Introducing target variables

The first manipulation applied to the MDO problem on our way to the proposed master problem is the introduction of the target variable vector $z \in \mathbb{R}^n$. The result is the following problem:

$$\begin{array}{rrcl} \min_{x_1,x_2,y_1,y_2,z} & \sum_{i=1}^{2} F_i(x_i,y_i) & & \\ s.t. & c_i(x_i,y_i) & \geq & 0, \quad i=1:2, \\ & x_i - z & = & 0, \quad i=1:2. \end{array} \tag{11}$$

The following theorem shows that, given a nondegenerate MDO minimizer, there is an equivalent minimizer to problem (11) that is also nondegenerate.

Theorem 4.1 *A point (x, y_1, y_2) is a minimizer to problem (10) satisfying the nondegeneracy conditions A.1-A.3 with the Lagrange multiplier vector (λ_1, λ_2) iff the point (x_1, x_2, y_1, y_2, z) with x_1, x_2, and z equal to x is a minimizer to problem (11) satisfying the nondegeneracy conditions A.1-A.3 with the Lagrange multiplier vector*

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \nabla_x F_1(x,y_1) - (\nabla_x c_1(x,y_1))^T \lambda_1 \\ \nabla_x F_2(x,y_2) - (\nabla_x c_2(x,y_2))^T \lambda_2 \end{pmatrix}. \tag{12}$$

Proof: First we prove that the linear independence constraint qualification A.1 holds at a point (x, y_1, y_2) for problem (10) iff it holds at (x_1, x_2, y_1, y_2, z) with x_1, x_2, and z equal to x for problem (11). To see this, suppose that condition A.1 does not hold for problem (10). Thus, there exists $(\lambda_1, \lambda_2) \neq 0$ such that

$$\begin{pmatrix} (\nabla_x \hat{c}_1(x,y_1))^T & (\nabla_x \hat{c}_2(x,y_2))^T \\ (\nabla_{y_1} \hat{c}_1(x,y_1))^T & \\ & (\nabla_{y_2} \hat{c}_2(x,y_2))^T \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = 0, \tag{13}$$

where $\hat{c}$ are the constraints active at (x, y_1, y_2). Then it is easy to see that

$$\begin{pmatrix} (\nabla_x \hat{c}_1(x,y_1))^T & & I & \\ & (\nabla_x \hat{c}_2(x,y_2))^T & & I \\ (\nabla_{y_1} \hat{c}_1(x,y_1))^T & & & \\ & (\nabla_{y_2} \hat{c}_2(x,y_2))^T & & \\ & & -I & -I \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix} = 0, \tag{14}$$

with $\lambda_3 = -(\nabla_x c_1(x,y_1))^T\lambda_1$ and $\lambda_4 = -(\nabla_x c_2(x,y_2))^T\lambda_2$. Thus, condition A.1 does not hold for problem (11). Conversely, assume there exists $(\lambda_1,\lambda_2,\lambda_3,\lambda_4) \neq 0$ such that equation (14) holds. Then, by the first and second rows in (14) we know that $\lambda_3 = -(\nabla_x c_1(x,y_1))^T\lambda_1$ and $\lambda_4 = -(\nabla_x c_2(x,y_2))^T\lambda_2$. This, by the third, fourth, and fifth rows in (14) shows that equation (13) holds for (λ_1,λ_2).

Now we turn to proving that (x,y_1,y_2) satisfies the KKT conditions with the Lagrange multiplier vector (λ_1,λ_2) for problem (10) iff (x_1,x_2,y_1,y_2,z) with x_1, x_2, and z equal to x satisfies the KKT conditions with the Lagrange multiplier vector given by (12) for problem (11). First, note that obviously the feasibility conditions (45)–(46) are satisfied at (x,y_1,y_2) for problem (10) iff they are satisfied at (x_1,x_2,y_1,y_2,z) with x_1, x_2, and z equal to x for problem (11). Moreover, assume there exists $(\lambda_1,\lambda_2) \geq 0$ satisfying the complementarity condition (47) at (x,y_1,y_2) such that

$$\begin{pmatrix} \sum_{i=1}^2 \nabla_x F_i(x,y_i) \\ \nabla_{y_1} F_1(x,y_1) \\ \nabla_{y_2} F_2(x,y_2) \end{pmatrix} = \begin{pmatrix} (\nabla_x c_1(x,y_1))^T & (\nabla_x c_2(x,y_2))^T \\ (\nabla_{y_1} c_1(x,y_1))^T & \\ & (\nabla_{y_2} c_2(x,y_2))^T \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}. \tag{15}$$

Then we obviously have that

$$\begin{pmatrix} \nabla_x F_1(x,y_1) \\ \nabla_x F_2(x,y_2) \\ \nabla_{y_1} F_1(x,y_1) \\ \nabla_{y_2} F_2(x,y_2) \\ 0 \end{pmatrix} = \begin{pmatrix} (\nabla_x c_1(x,y_1))^T & & I & \\ & (\nabla_x c_2(x,y_2))^T & & I \\ (\nabla_{y_1} c_1(x,y_1))^T & & & \\ & (\nabla_{y_2} c_2(x,y_2))^T & & \\ & & -I & -I \end{pmatrix} \lambda, \tag{16}$$

with λ as given by (12). Conversely, if there exists $(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$ satisfying equation (16), then by the first and second rows in (16) we know that $\lambda_3 = \nabla_x F_1(x,y_1) - (\nabla_x c_1(x,y_1))^T\lambda_1$ and $\lambda_4 = \nabla_x F_2(x,y_2) - (\nabla_x c_2(x,y_2))^T\lambda_2$. Then, by the third, fourth, and fifth rows in (16) we know that equation (15) is satisfied at (x,y_1,y_2) with Lagrange multipliers (λ_1,λ_2). Furthermore, it is clear that the complementarity and nonnegativity conditions (47) and (48) hold at (x,y_1,y_2) for problem (10) with (λ_1,λ_2) iff they hold at (x_1,x_2,y_1,y_2,z) with x_1, x_2, and z equal to x for problem (11) with the Lagrange multiplier given by (12).

To see that the strict complementarity slackness conditions A.2 hold at (x,y_1,y_2) for problem (10) with (λ_1,λ_2) iff they hold at (x_1,x_2,y_1,y_2,z) with x_1, x_2, and z equal to x for problem (11) with the Lagrange multiplier given by (12), simply note that the Lagrange multipliers associated with the inequality constraints are the same for both problems (λ_1,λ_2).

It only remains to show that the second order sufficient conditions for optimality A.3 hold at (x,y_1,y_2) for problem (10) with (λ_1,λ_2) iff they hold at (x_1,x_2,y_1,y_2,z) with x_1, x_2, and z equal to x for problem (11) with the Lagrange multiplier given by (12). To prove this, we first note that there is a one-to-one correspondence between the vectors tangent to the inequality constraints active at (x,y_1,y_2) for problem (10) and the vectors tangent to the equality and inequality constraints active at (x_1,x_2,y_1,y_2,z) with

x_1, x_2, and z equal to x for problem (11). In particular, given $\tau_1 = (\tau_x, \tau_{y_1}, \tau_{y_2})$ such that

$$\begin{pmatrix} \nabla_x \hat{c}_1(x, y_1) & \nabla_{y_1} \hat{c}_1(x, y_1) & \\ \nabla_x \hat{c}_2(x, y_2) & & \nabla_{y_2} \hat{c}_2(x, y_2) \end{pmatrix} \tau_1 = 0, \tag{17}$$

where $\hat{c}$ are the inequalities active at (x, y_1, y_2), we know that $\tau_2 = (\tau_{x_1}, \tau_{x_2}, \tau_{y_1}, \tau_{y_2}, \tau_z)$ with τ_{x_1}, τ_{x_2}, and τ_z equal to τ_x satisfies

$$\begin{pmatrix} \nabla_x \hat{c}_1(x, y_1) & & \nabla_{y_1} \hat{c}_1(x, y_1) & & \\ & \nabla_x \hat{c}_2(x, y_2) & & \nabla_{y_2} \hat{c}_2(x, y_2) & \\ I & & & & -I \\ & I & & & -I \end{pmatrix} \tau_2 = 0. \tag{18}$$

Conversely, if $\tau_2 = (\tau_{x_1}, \tau_{x_2}, \tau_{y_1}, \tau_{y_2}, \tau_z)$ satisfies (18), then we know that τ_{x_1} and τ_{x_2} are equal to τ_z and that $\tau_1 = (\tau_x, \tau_{y_1}, \tau_{y_2})$ with $\tau_x = \tau_z$ satisfies (17). Finally, given $\tau_1 = (\tau_x, \tau_{y_1}, \tau_{y_2})$ and $\tau_2 = (\tau_{x_1}, \tau_{x_2}, \tau_{y_1}, \tau_{y_2}, \tau_z)$ with τ_{x_1}, τ_{x_2}, and τ_z equal to τ_x, it is easy to see from the form of problems (10) and (11) that $\tau_1^T \nabla^2 \mathcal{L}_1 \tau_1 = \tau_2^T \nabla^2 \mathcal{L}_2 \tau_2$, where $\nabla^2 \mathcal{L}_1$ is the Hessian of the Lagrangian for problem (10) evaluated at $(x, y_1, y_2, \lambda_1, \lambda_2)$, and $\nabla^2 \mathcal{L}_2$ is the Hessian of the Lagrangian for problem (11) evaluated at $(x_1, x_2, y_1, y_2, z, \lambda)$ with λ as given by (12) and x_1, x_2 and z equal to x. ■

4.3.3 Introducing an inexact penalty function

The second transformation operated on problem (10) in order to obtain the desired master problem is the introduction of an inexact penalty function. We use a quadratic penalty function to remove the equality constraints in problem (11) to derive the following problem:

$$\begin{array}{lll} \min\limits_{x_1, x_2, y_1, y_2, z} & \sum\limits_{i=1}^{2} [F_i(x_i, y_i) + \gamma \|x_i - z\|_2^2] & \\ s.t. & c_i(x_i, y_i) \;\; \geq \;\; 0 \quad i = 1:2. & \end{array} \tag{19}$$

The following theorem follows from Theorems 14 and 17 in [18] and shows that the nondegeneracy of a minimizer to problem (11) is preserved by the transformation introduced above.

Theorem 4.2 *If* (x_1, x_2, y_1, y_2, z) *is a minimizer satisfying the nondegeneracy conditions A.1-A.3 for problem* (11)*, then for* γ *sufficiently large, there exists a locally unique once continuously-differentiable trajectory of points* $(x_1(\gamma), x_2(\gamma), y_1(\gamma), y_2(\gamma), z(\gamma))$ *satisfying the nondegeneracy conditions A.1-A.3 for problem* (19) *such that* $\lim_{\gamma \to \infty} (x_1(\gamma), x_2(\gamma), y_1(\gamma), y_2(\gamma), z(\gamma)) = (x_1, x_2, y_1, y_2, z)$.

4.3.4 Decomposition

Note that problem (19) can be *decomposed* into N independent subproblems by simply setting the target variables to a fixed value. The subproblem optimal-value functions can then be used to define a master problem. The result is the IPD decomposition algorithm. Namely, solve the following master problem

$$\min_{z} \quad \sum_{i=1}^{2} F_i^*(z), \tag{20}$$

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

$$\begin{aligned} F_i^*(z) \quad = \quad & \min_{x_i,y_i} \quad F_i(x_i,y_i) + \gamma\|x_i - z\|_2^2 \\ & s.t. \quad c_i(x_i,y_i) \geq 0. \end{aligned} \tag{21}$$

In this section, we show that the nondegeneracy of a minimizer to problem (19) is preserved by the decomposition described above. First, we introduce some useful definitions.

Definition 4.3 A vector $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ is a *semi-local minimizer* to problem (20) if for $i = 1, 2$ the vector (x_i^*, y_i^*) is a local minimizer for the ith subproblem (21) with $z = z^*$.

Definition 4.4 A vector $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ is a *strict local minimizer* to problem (20) if: (i) $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ is a semi-local minimizer to problem (20), and (ii) there exists a neighborhood $\mathcal{N}_\epsilon(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ such that if (x_1, x_2, y_1, y_2, z) lying in $\mathcal{N}_\epsilon$ is a semi-local minimizer then $\sum_{i=1}^{2}[F_i(x_i^*, y_i^*) + \gamma\|x_i^* - z^*\|_2^2] < \sum_{i=1}^{2}[F_i(x_i, y_i) + \gamma\|x_i - z\|_2^2]$.

The following lemma shows that a nondegenerate minimizer to problem (19) is also a strict local minimizer to problem (20).

Lemma 4.5 *If* $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ *is a minimizer to problem* (19) *satisfying the second order sufficient conditions A.3, then* $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ *is also a strict local minimizer to problem* (20).

Proof: We need to show that conditions (i) and (ii) in Definition 4.4 hold at $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$. Assume $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ is a minimizer to problem (19) satisfying condition A.3. Then, there exists a neighborhood $\mathcal{N}_\epsilon(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ such that for all points (x_1, x_2, y_1, y_2, z) feasible with respect to problem (19) and lying in $\mathcal{N}_\epsilon(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ we have that,

$$\sum_{i=1}^{2}[F_i(x_i^*, y_i^*) + \gamma\|x_i^* - z^*\|_2^2] < \sum_{i=1}^{2}[F_i(x_i, y_i) + \gamma\|x_i - z\|_2^2]. \tag{22}$$

In particular, for all (x_1, y_1) such that $\|(x_1^*, y_1^*) - (x_1, y_1)\|_2 < \epsilon$ we know by (22) that $F_1(x_1^*, y_1^*) + \gamma\|x_1^* - z^*\|_2^2 \leq F_1(x_1, y_1) + \gamma\|x_1 - z^*\|_2^2$. Likewise, for all (x_2, y_2) such that $\|(x_2^*, y_2^*) - (x_2, y_2)\|_2 < \epsilon$ we know that $F_2(x_2^*, y_2^*) + \gamma\|x_2^* - z^*\|_2^2 \leq F_2(x_2, y_2) + \gamma\|x_2 - z^*\|_2^2$. Thus, $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ is a semi-local minimizer to problem (20); i.e., condition (i) in Definition 4.4 holds. Also, every semi-local minimizer is feasible with respect to problem (19). Therefore, by (22), we know that $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ satisfies condition (ii) in Definition 4.4. ■

Finally, in the following two theorems we show that the proposed subproblems and master problem are nondegenerate.

Theorem 4.6 *If $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ is a minimizer satisfying the nondegeneracy conditions A.1-A.3 for problem* (19)*, then (x_i^*, y_i^*) is a minimizer satisfying the nondegeneracy conditions A.1-A.3 for the ith subproblem* (21) *with $z = z^*$.*

Proof: Note that the constraints in the ith subproblem (21) are a subset of those in problem (19). Moreover, none of these constraints depends on z. Therefore, if the the linear independence constraint qualification A.1 holds at $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ for problem (19) then it also holds at (x_i^*, y_i^*) for the ith subproblem (21) with $z = z^*$.

Let (λ_1, λ_2) be the unique Lagrange multiplier vector for problem (19) at $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$. By the KKT conditions we have

$$\begin{pmatrix} \nabla_{x_1} F_1(x_1^*, y_1^*) + 2\gamma(x_1^* - z^*) \\ \nabla_{x_2} F_2(x_2^*, y_2^*) + 2\gamma(x_2^* - z^*) \\ \nabla_{y_1} F_1(x_1^*, y_1^*) \\ \nabla_{y_2} F_2(x_2^*, y_2^*) \\ -2\sum_{i=1}^2 \gamma(x_i^* - z^*) \end{pmatrix} = \begin{pmatrix} \nabla_{x_1} c_1(x_1^*, y_1^*) & 0 \\ 0 & \nabla_{x_2} c_2(x_2^*, y_2^*) \\ \nabla_{y_1} c_1(x_1^*, y_1^*) & 0 \\ 0 & \nabla_{y_2} c_2(x_2^*, y_2^*) \\ 0 & 0 \end{pmatrix} \lambda, \tag{23}$$

where $\lambda = (\lambda_1, \lambda_2)$. The first four rows in (23) show that the KKT conditions hold for the ith subproblem (21) at (x_i^*, y_i^*) with the Lagrange multiplier λ_i. Moreover, if the strict complementarity slackness conditions A.2 hold at $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ with (λ_1, λ_2), then they obviously also hold at (x_i^*, y_i^*) with λ_i for the ith subproblem (21).

It only remains to show that the second order sufficient conditions for optimality A.3 hold at (x_i^*, y_i^*) with λ_i for the ith subproblem (21). Let (τ_{x_1}, τ_{y_1}) be tangent to the inequality constraints active at (x_1^*, y_1^*) for the first subproblem (21). Then we have that, because the constraints in the first subproblem (21) are a subset of the constraints in problem (19), the vector $(\tau_{x_1}, \tau_{x_2}, \tau_{y_1}, \tau_{y_2}, \tau_z)$ with τ_{x_2}, τ_{y_2}, and τ_z equal to zero is tangent to the inequality constraints active at $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ for problem (19).

Moreover, note that

$$
(\tau_{x_1}^T \;\; \tau_{y_1}^T)\; \nabla^2 \mathcal{L}_1 \begin{pmatrix} \tau_{x_1} \\ \tau_{y_1} \end{pmatrix} = (\tau_{x_1}^T \; \tau_{x_2}^T \; \tau_{y_1}^T \; \tau_{y_2}^T \; \tau_z^T)\; \nabla^2 \mathcal{L}_2 \begin{pmatrix} \tau_{x_1} \\ \tau_{x_2} \\ \tau_{y_1} \\ \tau_{y_2} \\ \tau_z \end{pmatrix},
$$

where τ_{x_2}, τ_{y_2}, and τ_z are equal to zero, $\nabla^2 \mathcal{L}_1$ is the Hessian of the Lagrangian for the first subproblem (21) evaluated at (x_1^*, y_1^*) with λ_1, and $\nabla^2 \mathcal{L}_2$ is the Hessian of the Lagrangian for problem (19) evaluated at $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ with (λ_1, λ_2). Thus the second order sufficient conditions hold for the first subproblem. The same argument can be used to show that the second order sufficient conditions hold for the second subproblem. ■

Theorem 4.7 *If the functions F_i and c_i are three times continuously differentiable and $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ is a minimizer to problem* (19) *satisfying the nondegeneracy conditions A.1-A.3 with Lagrange multipliers $(\lambda_1^*, \lambda_2^*)$, then:*

1. *the objective function of the master problem* (20) *can be defined locally as a twice continuously-differentiable function $F^*(z) = \sum_{i=1}^2 F_i^*(z)$ in a neighborhood $\mathcal{N}_\epsilon(z^*)$,*

2. *z^* is a minimizer to $F^*(z)$ satisfying the second order sufficient conditions.*

Proof: To see Part 1 note that from Theorem 4.6, we know that the vector (x_i^*, y_i^*) is a minimizer satisfying the nondegeneracy conditions A.1-A.3 for the ith subproblem (21) with $z = z^*$. Therefore, by the implicit function theorem and Theorem 6 in [18], we know that if F_i and c_i are three times continuously-differentiable, then there exists a locally unique twice continuously-differentiable trajectory $(x_i(z), y_i(z))$ of minimizers to the ith subproblem (21) defined for z in a neighborhood $\mathcal{N}_\epsilon(z^*)$ and such that $(x_i^*(z^*), y_i^*(z^*)) = (x_i^*, y_i^*)$. These trajectories define in turn a unique twice continuously-differentiable function $F^*(z)$ on $\mathcal{N}_\epsilon(z^*)$.

To prove Part 2, first we show that z^* is a minimizer to $F^*(z)$ and then we prove that it satisfies the second order sufficient conditions. From Part 1, we know that there exists a twice continuously-differentiable trajectory of minimizers to the ith subproblem $(x_i(z), y_i(z))$ defined in a neighborhood $\mathcal{N}_{\epsilon_1}(z^*)$. Then, by the differentiability of these trajectories, we know that for all $\epsilon_2 > 0$ there exists $\epsilon_3 > 0$ such that $\epsilon_3 < \epsilon_1$ and for all $z \in \mathcal{N}_{\epsilon_3}(z^*)$,

$$
(x_1(z), x_2(z), y_1(z), y_2(z), z) \in \mathcal{N}_{\epsilon_2}(x_1^*, x_2^*, y_1^*, y_2^*, z^*). \tag{24}
$$

Moreover, by Lemma 4.5 $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ is a strict local minimizer to problem (20), and therefore (24) implies that there exists $\epsilon_3 > 0$ such that $\epsilon_3 < \epsilon_1$ and $F^*(z) > F^*(z^*)$ for all $z \in \mathcal{N}_{\epsilon_3}(z^*)$. Thus z^* is a strict minimizer to $F^*(z)$.

It only remains to show that the second order sufficient conditions hold at z^* for the master problem. It suffices to show that for all nonzero $v \in \mathbb{R}^n$,

$$\left.\frac{d^2 F^*(z^* + rv)}{dr^2}\right|_{r=0} > 0.$$

But note that

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

Moreover, because $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ with $\lambda^* = (\lambda_1^*, \lambda_2^*)$ satisfies the strict complementarity slackness conditions A.2 for problem (19), and the implicit function theorem guarantees that the active set remains fixed for small r we know that

$$F^*(z^* + rv) = \mathcal{L}(X(r), \lambda^*),$$

where $\mathcal{L}$ is the Lagrangian function for problem (19) and $X(r) = (x_1(z^* + rv), x_2(z^* + rv), y_1(z^* + rv), y_2(z^* + rv), z^* + rv)$. Thus we have,

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

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

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

and the second derivative is

$$\begin{aligned} \frac{d^2 \mathcal{L}(X(r), \lambda^*)}{dr^2} =& \frac{dX(r)}{dr}^T \nabla_{XX}^2 \mathcal{L}(X(r), \lambda^*) \frac{dX(r)}{dr} \\ &+ \nabla_x \mathcal{L}(X(r), \lambda^*) \frac{d^2 X(r)}{dr^2}. \end{aligned} \tag{25}$$

Because $X(0)$ satisfies the KKT conditions for problem (19), (25) at $r = 0$ gives

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

The differentiability and feasibility with respect to problem (19) of $X(r)$ for r small, together with the fact that the linear independence constraint qualification holds at X^* for problem (19), imply that $dX(0)/dr$

is tangent to the inequality constraints active at $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ for problem (19). This, together with the fact that the second order sufficient conditions hold at $(x_1^*, x_2^*, y_1^*, y_2^*, z^*, \lambda^*)$ for problem (19), implies that

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

■

4.4 Local Convergence

In this section we use the nondegeneracy results of the previous section to analyze the local convergence properties of the proposed decomposition algorithm. In particular, we show in the following theorem that, for each value of the penalty parameter, the BFGS quasi-Newton method proposed to solve the master problem (Step 2 in Figure 1) converges locally at a superlinear rate. Moreover, we show that the sequential quadratic programming algorithm NPSOL converges locally at a superlinear rate when applied to the subproblem (21).

Theorem 4.8 *Let (x^*, y_1^*, y_2^*) be a minimizer to problem (10) satisfying the nondegeneracy conditions A.1-A.3. Then:*

1. *There exists a locally unique twice continuously-differentiable trajectory $z^*(\gamma)$ of minimizers to $F^*(z)$ defined for $\gamma \in (1/\epsilon_1, \infty)$ for some $\epsilon_1 > 0$, such that $\lim_{\gamma\to\infty} z^*(\gamma) = x^*$. Moreover, for each $\gamma \in (1/\epsilon_1, \infty)$ the minimizer $z^*(\gamma)$ satisfies the second order sufficient conditions A.3.*

2. *For each $\gamma \in (1/\epsilon_1, \infty)$, there exists $\epsilon_2 > 0$ such that if $\|z_0 - z^*(\gamma)\| < \epsilon_2$, then the iterates z_k generated by the BFGS quasi-Newton method (Step 2 in Figure 1) converge locally and superlinearly to $z^*(\gamma)$.*

3. *For each $\gamma \in (1/\epsilon_1, \infty)$ there exists $\epsilon_2 > 0$ such that there exists a locally unique trajectory $(x_i(z_k), y_i(z_k))$ of minimizers to the ith subproblem (21) satisfying the nondegeneracy conditions A.1-A.3 defined for $\|z_k - z^*(\gamma)\| < \epsilon_2$.*

4. *For each $\gamma \in (1/\epsilon_1, \infty)$, there exists $\epsilon_2 > 0$ such that if $\|z_k - z^*(\gamma)\| < \epsilon_2$, the iterates generated by NPSOL converge locally and superlinearly to $(x_i(z_k), y_i(z_k))$ when applied to the ith subproblem (21) with $z = z_k$.*

Proof: Part 1 follows from Theorems 4.1, 4.2, and 4.7. To prove Part 2, note that from Theorem 4.7 we know that $F^*(z)$ may be defined locally as a twice continuously-differentiable function and its Hessian matrix is positive definite at $z^*(\gamma)$. The local and superlinear convergence of the iterates generated by the master problem algorithm follows from standard theory for quasi-Newton unconstrained optimization algorithms [19].

From Theorems 4.1, 4.2, and 4.6 we know that there exists a minimizer to the ith subproblem $(x_i(z^*(\gamma)), y_i(z^*(\gamma)))$ for $z = z^*(\gamma)$ satisfying the nondegeneracy conditions A.1-A.3. Then, Part 3 follows from the nondegeneracy of $(x_i(z^*(\gamma), y_i(z^*(\gamma))$ and the implicit function theorem [17]. Finally, the local and superlinear convergence of NPSOL (Part 4) when applied to the subproblem (21) follows from the nondegeneracy of the subproblem minimizer $(x_i(z_k), y_i(z_k))$ given by Part 3 and standard convergence theory for sequential quadratic programming algorithms [27]. ■

5 Exact Penalty Decomposition

In the previous section, we proposed a decomposition algorithm based on the use of an inexact penalty function. In this section, we propose a decomposition algorithm based on an exact penalty function. The advantage is that, with an exact penalty function, the exact solution ($x_i = z$) is computed for finite values of the penalty parameter and thus we avoid the ill-conditioning introduced by large penalty parameters [28, Chapter 17]. The difficulty is that exact penalty functions are nonsmooth. We show that this difficulty can be overcome by employing barrier terms in conjunction with the exact penalty function. We term the algorithm Exact Penalty Decomposition (EPD).

Despite the obvious differences between the IPD and EPD formulations, the convergence analysis and results for both algorithms are strikingly similar. For the sake of brevity, we will not include those proofs that can be easily inferred from their counterparts for IPD.

The rest of this section is organized as follows. In Section 5.1, we formulate the proposed master problem and subproblems. In Section 5.2, we describe the optimization algorithms used to solve the master problem and the subproblems. In Section 5.3, we show that, under standard nondegeneracy assumptions on the MDO minimizer, the proposed master problem and subproblems are nondegenerate as well. Using these nondegeneracy results, we show in Section 5.4 that the optimization algorithms used to solve the master problem and the subproblems converge locally at a superlinear rate.

5.1 Formulation

As in CO and IPD, in EPD we allow the global variables to take a different value x_i within each of the subproblems. However, instead of using a quadratic penalty function, we use the l_1 exact penalty function $\|x_i - z\|_1 = \sum_{j=1}^{n} |x_{ij} - z_j|$ to enforce the global variables x_i to converge to the target variables z. But in order to avoid the nonsmoothness of the absolute value function, rather than using the l_1 exact penalty function explicitly, we introduce the *elastic variables* s_i and t_i. The result of introducing the elastic variables is the following master problem:

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

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

$$\begin{array}{rcll} F_i^*(z) & = & \min\limits_{x_i,y_i,s_i,t_i} & F_i(x_i,y_i) + \gamma e^T(s_i+t_i) \\ & & s.t. & c_i(x_i,y_i) \geq 0 \\ & & & x_i + s_i - t_i = z \\ & & & s_i, t_i \geq 0. \end{array} \tag{26}$$

It is obvious that, at a minimizer to subproblem (26), $\|x_i - z\|_1 = e^T(s_i + t_i)$. Moreover, when using the exact penalty function, we have that for γ sufficiently large but finite $x_i = z$ for $i = 1 : N$. Unfortunately, it is easy to show that, even if the linear independence constraint qualification holds at a minimizer to problem (10), the gradients of the active constraints at the minimizer to subproblem (26) are not linearly independent in general. This, by standard sensitivity theory [18], implies that the optimal-value function $F_i^*(z)$ in the above master problem is not smooth in general.

Fortunately, as we will show in the rest of this section, this difficulty can be overcome by introducing barrier terms to remove the nonnegativity constraints from the subproblems. The result is the exact penalty decomposition algorithm: solve the following master problem for a decreasing sequence of barrier parameters $\{\mu_k\}$ such that $\mu_k \to 0$:

$$\min_{z} \quad \sum_{i=1}^{N} F_i^*(z), \tag{27}$$

where $F_i^*(z)$ [1] is the optimal-value function of the i-th subproblem,

$$\begin{array}{rl} \min\limits_{x_i,y_i,s_i,t_i} & F_i(x_i,y_i) + \gamma e^T(s_i+t_i) - \mu \sum_{j=1}^{n} (\log s_{ij} + \log t_{ij}) \\ s.t. & c_i(x_i,y_i) \geq 0, \\ & x_i + s_i - t_i = z. \end{array} \tag{28}$$

5.2 Algorithm Statement

To solve the EPD master problem (27), we use the same BFGS quasi-Newton method we proposed for the IPD master problem. However, we use a primal-dual interior-point method [36] to solve the subproblems (28). These methods are especially designed to deal with problems that include barrier functions. We coded the method in the MATLAB file PDSOL.

[1] Although the subproblem optimal-value functions $F_i^*(z)$ depend on γ and μ, we do not include this dependence explicitly to simplify notation.

The master problem objective $F^*(z_k)$ can be computed from the subproblem minimizers $(x_{ik}, y_{ik}, s_{ik}, t_{ik})$ as

$$F^*(z_k) = \sum_{i=1}^{N} F_i(x_{ik}, y_{ik}) + \gamma e^T(s_{ik} + t_{ik}) - \mu \sum_{j=1}^{n} (\log(s_{ik})_j + \log(t_{ik})_j). \tag{29}$$

Moreover, as we show in Section 5.3, the computed subproblem minimizers are nondegenerate provided z_k is close to the master problem minimizer z^*. Thus, standard sensitivity results [18] can be used to compute the master problem objective gradient as:

$$\nabla F^*(z_k) = \sum_{i=1}^{N} \nabla F_i^*(z_k) = -\sum_{i=1}^{N} \lambda_{zi}, \tag{30}$$

where λ_{zi} are the Lagrange multipliers corresponding to the equality constraints $x_i + s_i - t_i = z$.

The proposed decomposition algorithm is outlined in Figure 2.

Initialization: Initialize μ and choose γ sufficiently large. Set $k := 0$. Choose a starting point z_0. For $i = 1 : N$, call PDSOL to find the subproblem minimizers $(x_{i0}, y_{i0}, s_{i0}, t_{i0})$. Set the initial Hessian approximation B_0 equal to the identity. Choose $\sigma \in (0, 1)$.

while $(\mu > \epsilon)$

1. **Barrier parameter update:** decrease μ.
2. **Solve master problem** with current μ:
   repeat
   (a) Function evaluation: for $i = 1 : N$ call PDSOL with μ and compute $F^*(z_k)$ and $\nabla F^*(z_k)$ from (29) and (30).
   (b) Search direction: Solve $B_k \Delta z = -\nabla F^*(z_k)$.
   (c) Backtracking line search: $\alpha := 1$.
   while $(F^*(z_k + \alpha\Delta z) - F^*(z_k) > \sigma \nabla F^*(z_k)^T \Delta z)$
   $\alpha := \alpha/2$. Call PDSOL to evaluate $F^*(z_k + \alpha\Delta z)$.
   endwhile
   $s := \alpha\Delta z$, $y := \nabla F^*(z_k + s) - \nabla F^*(z_k)$.
   $z_{k+1} := z_k + s$.
   (d) BFGS Hessian update: $B_{k+1} := B_k - \frac{B_k s s^T B_k}{s^T B_k s} + \frac{y y^T}{y^T s}$
   (e) k = k+1.
   until $(\|\nabla F^*(z_k)\| / |1 + F^*(z_k)| < \epsilon)$

endwhile

Figure 2: Exact penalty decomposition

Note that, assuming we choose the initial penalty parameter sufficiently large, there is no need to update it. On the other hand, we need to progressively drive the barrier parameter to zero by decreasing it in Step 1. The only other difference with the IPD algorithm stated in Figure 1 is that we use PDSOL and not NPSOL to solve the subproblems.

5.3 Nondegeneracy Results

In this section, we show that, given a nondegenerate MDO minimizer, there exists an equivalent minimizer to the EPD master problem and subproblems that is nondegenerate as well. To prove the result, it is essential to note that the EPD master problem (27) can be derived from the MDO problem through a sequence of three manipulations: (i) introduction of target and elastic variables, (ii) introduction of barrier terms, and (iii) decomposition. The nondegeneracy result follows by showing that the nondegeneracy of an MDO minimizer is preserved by each of the transformations listed above. The analysis in this section parallels that in Section 4.3. For the sake of brevity, we will not include those proofs that can be easily inferred from their counterparts for IPD.

5.3.1 Notation and assumptions

As in IPD, to facilitate the exposition and without loss of generality, herein we consider the MDO problem composed of two systems (10). The interested reader is referred to [11] for an exposition with more than one system. As in IPD, we assume there exists a KKT point $w^* = (x^*, y_1^*, y_2^*, \lambda_1^*, \lambda_2^*)$ for problem (10) satisfying the nondegeneracy conditions A.1-A.3. We also assume that the functions in problem (10) are three times continuously-differentiable in an open convex set containing w^*.

5.3.2 Introducing target and elastic variables

The first manipulation applied to the MDO problem on our way to the proposed master problem is the introduction of the elastic variable vectors $s_i, t_i \in \mathbb{R}^n$ and the target variable vector $z \in \mathbb{R}^n$. The result is the following problem:

$$\begin{array}{lrcll} \min\limits_{x_i,y_i,s_i,t_i,z} & \multicolumn{4}{l}{\sum_{i=1}^{2} F_i(x_i,y_i) + \gamma e^T(s_i+t_i)} \\ s.t. & c_i(x_i,y_i) & \geq & 0, & i=1,2, \\ & x_i+s_i-t_i & = & z, & i=1,2, \\ & s_i,t_i & \geq & 0, & i=1,2, \end{array} \tag{31}$$

where γ is the penalty parameter and $e \in \mathbb{R}^n$ is the vector of ones.

The following theorem shows that, given a nondegenerate MDO minimizer, there exists an equivalent minimizer to problem (31) that is nondegenerate as well.

Theorem 5.1 *Provided*

$$\gamma > \left\| \begin{pmatrix} \nabla_x F_1(x,y_1) - (\nabla_x c_1(x,y_1))^T \lambda_1 \\ \nabla_x F_2(x,y_2) - (\nabla_x c_2(x,y_2))^T \lambda_2 \end{pmatrix} \right\|_\infty, \tag{32}$$

a point (x, y_1, y_2) *is a minimizer to problem* (10) *satisfying the nondegeneracy conditions A.1-A.3 with Lagrange multipliers* (λ_1, λ_2) *if and only if the point* $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ *with* $s_1, s_2, t_1, t_2 = 0$ *and* $x_1, x_2, z = x$ *is a minimizer to problem* (31) *satisfying the nondegeneracy conditions A.1-A.3 with Lagrange multipliers*

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \nabla_x F_1(x, y_1) - (\nabla_x c_1(x, y_1))^T \lambda_1 \\ \nabla_x F_2(x, y_2) - (\nabla_x c_2(x, y_2))^T \lambda_2 \\ \gamma e - \nabla_x F_1(x, y_1) + (\nabla_x c_1(x, y_1))^T \lambda_1 \\ \gamma e - \nabla_x F_2(x, y_2) + (\nabla_x c_2(x, y_2))^T \lambda_2 \\ \gamma e + \nabla_x F_1(x, y_1) - (\nabla_x c_1(x, y_1))^T \lambda_1 \\ \gamma e + \nabla_x F_2(x, y_2) - (\nabla_x c_2(x, y_2))^T \lambda_2 \end{pmatrix} \tag{33}$$

Proof: First, we introduce some notation. Let

$$J_1^T = \begin{pmatrix} (\nabla_x c_1(x, y_1))^T & (\nabla_x c_2(x, y_2))^T \\ (\nabla_{y_1} c_1(x, y_1))^T & \\ & (\nabla_{y_2} c_2(x, y_2))^T \end{pmatrix}$$

be the transposed Jacobian matrix for problem (10) and let $\hat{J}_1^T$ be the submatrix of J_1^T composed only by the gradients of those inequality constraints $\hat{c}_1$ and $\hat{c}_2$ that are active at (x, y_1, y_2). Likewise, let the transposed Jacobian matrix for problem (31) be

$$J_2^T = \begin{pmatrix} (\nabla_x c_1(x, y_1))^T & & I & & & & & \\ & (\nabla_x c_2(x, y_2))^T & & I & & & & \\ (\nabla_{y_1} c_1(x, y_1))^T & & & & & & & \\ & (\nabla_{y_2} c_2(x, y_2))^T & & & & & & \\ & & I & & I & & & \\ & & & I & & I & & \\ & & -I & & & & I & \\ & & & -I & & & & I \\ & & -I & -I & & & & \end{pmatrix},$$

where I is the identity matrix of dimension n, and let $\hat{J}_2^T$ be the submatrix of J_2^T composed only by the gradients of the constraints active at the point $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ with $s_1, s_2, t_1, t_2 = 0$ and $x_1, x_2, z = x$ for problem (31).

Now we prove that the linear independence constraint qualification A.1 holds at (x, y_1, y_2) for problem (10) iff it holds at $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ with $s_1, s_2, t_1, t_2 = 0$ and $x_1, x_2, z = x$ for problem (31). To see this, suppose that condition A.1 does not hold for problem (10). Thus, suppose there exists

$(\lambda_1, \lambda_2) \neq 0$ such that

$$\hat{J}_1^T \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = 0. \tag{34}$$

Then it is easy to see that

$$\hat{J}_2^T \lambda = 0, \tag{35}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_8)$ with $\lambda_3 = -\lambda_5 = \lambda_7 = -(\nabla_x \hat{c}_1(x, y_1))^T \lambda_1$, and $\lambda_4 = -\lambda_6 = \lambda_8 = -(\nabla_x \hat{c}_2(x, y_2))^T \lambda_2$. Thus, condition A.1 does not hold for problem (31). Conversely, it is easy to show by similar arguments that if there exists $\lambda_1, \ldots, \lambda_8 \neq 0$ such that equation (35) holds, then (34) holds for $(\lambda_1, \lambda_2) \neq 0$.

Now we turn to the KKT conditions. First, it is obvious that the feasibility conditions (45)–(46) are satisfied at (x, y_1, y_2) for problem (10) iff they are satisfied at $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ with $s_1, s_2, t_1, t_2 = 0$ and $x_1, x_2, z = x$ for problem (31). Moreover, assume there exists $(\lambda_1, \lambda_2) \geq 0$ satisfying the complementarity condition (47) at (x, y_1, y_2) such that

$$\begin{pmatrix} \sum_{i=1}^2 \nabla_x F_i(x, y_i) \\ \nabla_{y_1} F_1(x, y_1) \\ \nabla_{y_2} F_2(x, y_2) \end{pmatrix} = J_1^T \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}. \tag{36}$$

Then we obviously have that

$$\begin{pmatrix} \nabla_x F_1(x, y_1) \\ \nabla_x F_2(x, y_2) \\ \nabla_{y_1} F_1(x, y_1) \\ \nabla_{y_2} F_2(x, y_2) \\ \gamma e \\ \gamma e \\ \gamma e \\ \gamma e \\ 0 \end{pmatrix} = J_2^T \lambda, \tag{37}$$

with λ as given by (33). Conversely, if there exists $(\lambda_1, \lambda_2, \ldots, \lambda_8)$ satisfying equation (37), then $(\lambda_1, \lambda_2, \ldots, \lambda_8)$ must be as given in (33) and (36) is satisfied at (x, y_1, y_2) for problem (10) with (λ_1, λ_2). Furthermore, it is clear that, provided condition (32) holds, the complementarity, nonnegativity, and strict complementarity conditions [(47), (48), and A.2] hold at (x, y_1, y_2) for problem (10) with (λ_1, λ_2) iff they hold at $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ with $s_1, s_2, t_1, t_2 = 0$ and $x_1, x_2, z = x$ for problem (31) with the Lagrange multiplier given by (33).

Finally, we turn to the second order sufficient conditions for optimality A.3. First, note that there is a one-to-one correspondence between the vectors tangent to the inequality constraints active at

(x, y_1, y_2) for problem (10) and the vectors tangent to the equality and inequality constraints active at $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ with with $s_1, s_2, t_1, t_2 = 0$ and $x_1, x_2, z = x$ for problem (31). In particular, given $\tau_1 = (\tau_x, \tau_{y_1}, \tau_{y_2})$ such that

$$\hat{J}_1 \tau_1 = 0, \tag{38}$$

we know that $\tau_2 = (\tau_{x_1}, \tau_{x_2}, \tau_{y_1}, \tau_{y_2}, \tau_{s_1}, \tau_{s_2}, \tau_{t_1}, \tau_{t_2}, \tau_z)$ with $\tau_{x_1}, \tau_{x_2}, \tau_z = \tau_x$ and $\tau_{s_1}, \tau_{s_2}, \tau_{t_1}, \tau_{t_2} = 0$ satisfies

$$\hat{J}_2 \tau_2 = 0. \tag{39}$$

Conversely, if τ_2 satisfies (39), then we know that $\tau_{x_1}, \tau_{x_2} = \tau_z$, and that $\tau_{s_1}, \tau_{s_2}, \tau_{t_1}, \tau_{t_2} = 0$ and $(\tau_z, \tau_{y_1}, \tau_{y_2})$ satisfies (38). Finally, given τ_1 and τ_2 with $\tau_{x_1}, \tau_{x_2}, \tau_z = \tau_x$ and $\tau_{s_1}, \tau_{s_2}, \tau_{t_1}, \tau_{t_2} = 0$, it is easy to see from the form of problems (10) and (31) that $\tau_1^T \nabla^2 \mathcal{L}_1 \, \tau_1 = \tau_2^T \nabla^2 \mathcal{L}_2 \, \tau_2$ where $\nabla^2 \mathcal{L}_1$ is the Hessian of the Lagrangian for problem (10) and $\nabla^2 \mathcal{L}_2$ is the Hessian of the Lagrangian for problem (31). ■

5.3.3 Introducing barrier terms

The second transformation operated on problem (10) in order to obtain the desired master problem is the introduction of barrier terms to remove the nonnegativity bounds. The result is the following problem:

$$\begin{array}{lrcl} \min\limits_{x_i, y_i, s_i, t_i, z} & \multicolumn{3}{l}{\sum\limits_{i=1}^{2} [F_i(x_i, y_i) + \gamma e^T (s_i + t_i) - \mu \sum\limits_{j=1}^{n} (\log s_{ij} + \log t_{ij})]} \\ s.t. & c_i(x_i, y_i) & \geq & 0, \quad i = 1, 2, \\ & x_i + s_i - t_i & = & z, \quad i = 1, 2, \end{array} \tag{40}$$

where μ is the barrier parameter.

The following theorem follows from Theorems 14 and 17 in [18] and shows that the nondegeneracy of a minimizer to problem (31) is preserved through the transformation introduced above.

Theorem 5.2 *If $X = (x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ is a minimizer satisfying the nondegeneracy conditions A.1-A.3 for problem* (31)*, then there exists $\epsilon > 0$ such that for $\mu \in (0, \epsilon)$ there exists a unique once continuously differentiable trajectory of minimizers to problem* (40) *$X(\mu)$ satisfying the the nondegeneracy conditions A.1-A.3 such that $lim_{\mu \to 0} X(\mu) = X$.*

5.3.4 Decomposition

Note that problem (40) can be *decomposed* into N independent subproblems by simply setting the target variables to a fixed value. The subproblem optimal-value functions can then be used to define a master problem. The result is the EPD master problem

$$\min_{z} \quad \sum_{i=1}^{2} F_i^*(z), \tag{41}$$

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

$$\begin{array}{ll} \min\limits_{x_i,y_i,s_i,t_i} & F_i(x_i,y_i) + \gamma e^T(s_i+t_i) - \mu \sum_{j=1}^{n}(\log s_{ij} + \log t_{ij}) \\ s.t. & c_i(x_i,y_i) \geq 0, \\ & x_i + s_i - t_i = z. \end{array} \tag{42}$$

In this section, we show that the nondegeneracy of a minimizer to (40) is preserved by the decomposition described above.

The proof to the following lemma parallels that of Lemma 4.5.

Lemma 5.3 *If* $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ *is a minimizer to problem* (40) *satisfying the second order sufficient conditions A.3, then it is also a strict local minimizer to problem* (41)*.*

Finally, in the following two theorems we show that the proposed subproblem and master problem are nondegenerate.

Theorem 5.4 *If* $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ *is a minimizer satisfying the nondegeneracy conditions A.1-A.3 for problem* (40)*, then* (x_i, y_i, s_i, t_i) *is a minimizer satisfying the nondegeneracy conditions A.1-A.3 for the* i*th subproblem* (42) *with* z*.*

Proof: Note that, because of the elastic variables, the gradients of the equality constraints in the ith subproblem (42) are always linearly independent from the gradients of the active inequality constraints. Moreover, if the linear independence constraint qualification holds for problem (40) then the gradients of the inequality constraints $c_i(x_i, y_i)$ active at (x_i, y_i, s_i, t_i) must be linearly independent. Therefore, the linear independence constraint qualification A.1 holds for the ith subproblem.

Now we turn to proving that the KKT conditions hold for the ith subproblem. Let $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ be the unique Lagrange multiplier vector for problem (40) at $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$. By the KKT conditions we have

$$\begin{pmatrix} \nabla_{x_1} F_1(x_1, y_1) \\ \nabla_{x_2} F_2(x_2, y_2) \\ \nabla_{y_1} F_1(x_1, y_1) \\ \nabla_{y_2} F_2(x_2, y_2) \\ \gamma e + \mu S_1^{-2} e \\ \gamma e + \mu S_2^{-2} e \\ \gamma e + \mu T_1^{-2} e \\ \gamma e + \mu T_2^{-2} e \\ 0 \end{pmatrix} = \begin{pmatrix} (\nabla_{x_1} c_1(x_1^*, y_1^*))^T & 0 & I & 0 \\ 0 & (\nabla_{x_2} c_2(x_2^*, y_2^*))^T & 0 & I \\ (\nabla_{y_1} c_1(x_1^*, y_1^*))^T & 0 & 0 & 0 \\ 0 & (\nabla_{y_2} c_2(x_2^*, y_2^*))^T & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ 0 & 0 & -I & 0 \\ 0 & 0 & 0 & -I \\ 0 & 0 & -I & -I \end{pmatrix} \lambda, \tag{43}$$

where $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. But notice that by the first, third, fifth, and seventh rows in (43) we know that the KKT conditions hold for the first subproblem. Likewise, the second, fourth, sixth, and eighth rows imply that the KKT conditions hold for the second subproblem. Moreover, if the strict complementarity slackness conditions A.2 hold for problem (40) with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, then they obviously also hold for the first subproblem (42) with (λ_1, λ_3) and for the second subproblem with (λ_2, λ_4).

It only remains to show that the second order sufficient conditions for optimality A.3 hold for the ith subproblem (42). Let $\tau_1 = (\tau_{x_1}, \tau_{y_1}, \tau_{s_1}, \tau_{t_1})$ be tangent to the equality and active inequality constraints at (x_1, y_1, s_1, t_1) for the first subproblem (42). Then the vector $\tau_2 = (\tau_{x_1}, \tau_{x_2}, \tau_{y_1}, \tau_{y_2}, \tau_{s_1}, \tau_{s_2}, \tau_{t_1}, \tau_{t_2}, \tau_z)$ with $\tau_{x_2}, \tau_{y_2}, \tau_{s_2}, \tau_{t_2}, \tau_z = 0$ is tangent to the equality and active inequality constraints for problem (40). Moreover, it is easy to see that $\tau_1^T \mathcal{L}_1 \tau_1 = \tau_2^T \nabla^2 \mathcal{L}_2 \tau_2$, where $\nabla^2 \mathcal{L}_1$ is the Hessian of the Lagrangian for the first subproblem (42), and $\nabla^2 \mathcal{L}_2$ is the Hessian of the Lagrangian for problem (40). The same argument can be applied for the second subproblem. ■

The proof to the following theorem parallels that of Theorem 4.7

Theorem 5.5 *If F_i and c_i are three times continuously-differentiable and the point $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ is a minimizer to problem (40) satisfying the nondegeneracy conditions A.1-A.3, then:*

1. *the objective function of the master problem (41) can be defined locally as a twice continuously-differentiable function $F^*(z)$ in a neighborhood $\mathcal{N}_\epsilon(z)$,*

2. *z is a minimizer to $F^*(z)$ satisfying the second order sufficient conditions.*

5.4 Local Convergence

In the following theorem we show that, for each value of the barrier parameter, the BFGS quasi-Newton method proposed to solve the master problem (Step 2 in Figure 2) converges locally at a superlinear rate. Moreover, we show that the primal-dual interior-point method PDSOL converges locally at a superlinear rate when applied to the subproblem (42).

Theorem 5.6 *Let (x^*, y_1^*, y_2^*) be a minimizer to problem (10) satisfying the nondegeneracy conditions A.1-A.3. Then:*

1. *There exists a locally unique twice continuously-differentiable trajectory $z^*(\mu)$ of minimizers to $F^*(z)$ defined for $\mu \in (0, \epsilon_1)$ for some $\epsilon_1 > 0$, such that $\lim_{\mu \to 0} z^*(\mu) = x^*$. Moreover, for each $\mu \in (0, \epsilon_1)$ the minimizer $z^*(\mu)$ satisfies the second order sufficient conditions A.3.*

2. *For each $\mu \in (0, \epsilon_1)$, there exists $\epsilon_2 > 0$ such that if $\|z_0 - z^*(\mu)\| < \epsilon_2$, then the iterates z_k generated by the BFGS quasi-Newton method (Step 2 in Figure 2) converge locally and superlinearly to $z^*(\mu)$.*

3. *For each $\mu \in (0, \epsilon_1)$, there exists $\epsilon_2 > 0$ such that there exists a locally unique trajectory $(x_i(z_k), y_i(z_k), s_i(z_k), t_i(s_k))$ of minimizers to the ith subproblem (42) satisfying the nondegeneracy conditions A.1-A.3 defined for $\|z_k - z^*(\mu)\| < \epsilon_2$.*

4. *For each $\mu \in (0, \epsilon_1)$, there exists $\epsilon_2 > 0$ such that if $\|z_k - z^*(\mu)\| < \epsilon_2$, the iterates generated by PDSOL converge locally and superlinearly to $(x_i(z_k), y_i(z_k), s_i(z_k), t_i(s_k))$ when applied to the subproblem (42) with $z = z_k$.*

Proof: Part 1 follows from Theorems 5.1, 5.2, and 5.5. To prove Part 2, note that from Theorem 5.5 we know that $F^*(z)$ may be defined locally as a twice continuously-differentiable function and its Hessian matrix is positive definite at $z^*(\mu)$. The local and superlinear convergence of the iterates generated by the master problem algorithm follows from standard theory for quasi-Newton unconstrained optimization algorithms [19].

From Theorems 5.1, 5.2, and 5.4 we know that there exists a minimizer to the ith subproblem (42) $(x_i(z^*(\mu)), y_i(z^*(\mu)), s_i(z^*(\mu)), t_i(z^*(\mu)))$ for $z = z^*(\mu)$ satisfying the nondegeneracy conditions A.1-A.3. Then, Part 3 follows from the nondegeneracy of $(x_i(z^*(\mu)), y_i(z^*(\mu)), s_i(z^*(\mu)), t_i(z^*(\mu)))$ and the implicit function theorem [17]. Finally, the local and superlinear convergence of PDSOL (Part 4) when applied to the subproblem (42) follows from the nondegeneracy of the ith subproblem minimizer $(x_i(z_k), y_i(z_k), s_i(z_k), t_i(z_k))$ given by Part 3 and standard convergence theory for primal-dual interior-point methods [37]. ■

6 Numerical Results

In the previous sections, we have proved that IPD and EPD converge locally at a superlinear rate for each value of the penalty and barrier parameter. The question remains, however, whether the ill-conditioning usually associated with large values of the penalty parameter or small values of the barrier parameter will hinder the numerical performance of IPD and EPD.

To address this question, in this section we give a preliminary analysis the numerical performance of IPD and EPD. A thorough computational test is out of the scope of this paper and will be done elsewhere. Our preliminary results show that the numerical performance of our algorithms is not seriously impacted by the ill-conditioning of the penalty and barrier functions.

Although there exist several stochastic programming test-problem sets (e.g. [24, p. 47], and [3]), most of them correspond to linear or convex test problems. To overcome this difficulty, DeMiguel [13] modified the bilevel programming test problems by Calamai and Vicente [8] to create a nonlinear (nonconvex) MDO test-problem set. The nature of these test problems is such that the user can choose the test-problem size, convexity, and the type of constraint qualification satisfied at the minimizer. Moreover, all local and global minimizers of the test problems are known a priori.

We compare our algorithms with the Progressive Hedging Algorithm (PHA) [29]. The PHA uses an augmented Lagrangian function to break the MDO problem into a set of subproblems. We ran IPD, EPD, and PHA on 30 test problems from [13]. For PHA, we use three different values of the penalty parameter $\gamma = 0.01, 1, 100$. All results were obtained using MATLAB 5.3 on a Dell laptop with an 800 MHz Pentium III, 256MB RAM and running under WINDOWS 2000. The total time needed to solve the 30 test problems was less than two hours.

The results are given in Table 1. The first four columns give information about the test problem. The first column indicates whether the test problem is a quadratic program (Q) or a nonlinear program (N). The second column indicates whether the test problem is convex (C) or nonconvex (N). The third column indicates whether the test problem satisfies the linear independence constraint qualification (L) or the strong linear independence constraint qualification (S); the strong linear independence constraint qualification holds if for $i = 1 : N$ the matrix $\nabla_{y_i} c_i(x, y_i)$ has full row rank. Finally, the fourth column is the number of variables of the test problem.

The numerical performance of each decomposition algorithm is described by two columns. The first column gives the number of subproblems that had to be solved in order to find the overall minimizer. This quantity measures the amount of communication required between the master problem and the subproblems. In situations were this communication is expensive, such as in the context of the multidisciplinary design optimization problem, this is the main measure of the effectiveness of a decomposition algorithm. We interrupted the execution of the decomposition algorithms when a maximum of 400 subproblems were solved; that is, a 400 means the decomposition algorithm failed to find a stationary point after solving 400 subproblems. The second column is the average number of function evaluations needed to solve each subproblem. Finally, note that IPD could not be applied to solve nonlinear test problems because the subproblem solver PDSOL is capable of solving only quadratic programs.

The main insight from the numerical results is that whereas IPD and EPD found stationary points for all test problems, PHA reached the maximum number of subproblems (400) without finding a stationary

Test Problem				IPD		EPD		PHA $\gamma : 0.01$		PHA $\gamma : 1$		PHA $\gamma : 10^2$	
				sub	fun	sub	fun	sub	fun	sub	fun	sub	fun
Q	C	S	42	12	47	13	9	400	39	103	29	400	31
Q	C	S	84	27	80	24	11	400	77	400	72	400	62
Q	C	S	126	33	109	23	12	400	90	400	80	400	92
Q	C	S	168	34	149	41	12	400	113	400	81	400	158
Q	C	S	210	66	165	111	12	400	141	400	159	400	126
Q	C	L	42	68	24	57	7	400	21	39	17	159	37
Q	C	L	84	133	49	91	8	400	35	193	42	400	63
Q	C	L	126	170	47	248	8	400	15	342	23	400	69
Q	C	L	168	220	73	299	10	400	26	400	54	400	107
Q	C	L	210	348	75	351	10	400	24	400	133	400	104
Q	N	S	42	10	47	11	7	36	12	173	18	400	43
Q	N	S	84	20	79	23	10	400	26	400	71	400	67
Q	N	S	126	17	119	33	12	211	25	400	90	400	113
Q	N	S	168	17	154	30	25	145	27	400	105	400	149
Q	N	S	210	33	186	99	14	400	127	400	170	400	162
Q	N	L	42	83	28	103	7	400	23	39	15	122	35
Q	N	L	84	147	54	221	8	400	41	108	56	400	62
Q	N	L	126	224	62	297	9	400	54	119	44	400	63
Q	N	L	168	281	86	307	10	400	73	293	47	400	110
Q	N	L	210	331	109	397	11	400	86	400	146	400	121
N	N	S	42	9	52	-	-	400	13	120	38	400	44
N	N	S	84	18	89	-	-	400	52	400	73	400	69
N	N	S	126	17	134	-	-	400	36	400	123	400	109
N	N	S	168	16	174	-	-	186	69	400	140	400	153
N	N	S	210	28	210	-	-	400	182	400	152	400	153
N	N	L	42	142	31	-	-	400	24	46	33	400	45
N	N	L	84	285	58	-	-	400	70	400	77	400	61
N	N	L	126	272	84	-	-	400	87	386	83	400	100
N	N	L	168	279	120	-	-	400	98	140	98	400	149
N	N	L	210	356	141	-	-	400	123	400	148	400	155

Table 1: **Numerical results:** the first four columns describe the test problem. In the first column, Q means quadratic program and N means nonlinear program. In the second column, C means convex, N means nonconvex. In the third column, S means strong linear independence constraint qualification and L means linear independence constraint qualification. The fourth column is the number of variables. For each algorithm we give the number of subproblems solved (sub) and the average number of function evaluations per subproblem (fun). We interrupt the execution of an algorithm when it reaches 400 subproblems. EPD can not solve nonlinear programs because the subproblem solver PDSOL is a quadratic programming solver.

point for most of the test problems and for all three values of the penalty parameter. The reason is that while IPD and EPD converge at a superlinear rate for each value of the penalty and barrier parameters, the PHA achieves only a linear convergence rate and, for the test problems tried, with a convergence constant very close to one. Table 2 illustrates the typical convergence rate of the three algorithms.

Table 2: Typical convergence rate for IPD, EPD, and PHA.

IPD	EPD	PHA
1.71e-002	3.93e-002	1.57e-005
1.95e-003	1.74e-003	1.54e-005
1.09e-004	3.29e-004	1.50e-005
6.57e-006	4.54e-006	1.46e-005

Note that the number of subproblems that had to be solved in order to find the overall minimizer is quite similar for IPD and EPD. However, the average number of function evaluations needed to solve each subproblem is substantially larger for IPD. The reason for this is that the IPD subproblem solver (NPSOL) makes use only of first derivatives, whereas the EPD subproblem solver (PDSOL) uses also second derivatives. Notice also that the computational effort required to solve test problems satisfying only the linear independence constraint qualification (LICQ) is an order of magnitude larger than that required to solve test problems satisfying the strong linear independence constraint qualification (SLICQ). This confirms that the distinction between LICQ and SLICQ is also important in practice.

To analyze the effect that the number of global variables has on the computational effort, we solved a number of test problems with the same number of global variables but with an increasing number of local variables. The number of subproblems needed to find a stationary point remained roughly constant for all test problems. This seems to imply that the computational effort will depend mostly on the number of *global* variables.

Finally, note that the CO algorithm cannot be run on our 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 in [2] on which CO failed.

7 Conclusion

We propose two bilevel decomposition algorithms and show that their master problem and subproblems are nondegenerate. As a consequence, we show that they converge locally at a superlinear rate for each value of the penalty and barrier parameters. Our preliminary computational experiments show that the performance of our algorithms is not seriously impacted by the ill-conditioning associated with the penalty and barrier functions. In fact, our algorithms perform better on the test problems tried than the progressive hedging algorithm [29]. We believe IPD and EPD will be especially efficient when applied to problems having only a few global variables and a large number of local variables.

Acknowledgements. We are grateful to R.W. Cottle, M.P. Friedlander, M.A. Saunders, S. Scholtes for many helpful comments. We would also like to acknowledge comments from seminar participants at Stanford University, London Business School, Cambridge University, Universidad Carlos III de Madrid, INFORMS International Conference (Hawaii, 2001), Argonne National Laboratory, and Northwestern University. Financial support for this research was provided by the La Caixa Fellowship Program, the National Science Foundation Grant CCR-9988205, the Department of Naval Research Grant 00014-96-1-0274, and General Motors under the GM/Stanford Collaborative Research Laboratory in Work Systems.

A Optimality Conditions

Consider a general nonlinear optimization problem of the form

$$\begin{array}{lrcl} \min\limits_{x} & F(x) & & \\ s.t. & c(x) & \geq & 0, \\ & d(x) & = & 0, \end{array} \tag{44}$$

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

Definition A.1 The *linear independence constraint qualification* holds at a feasible point x if the gradients of all active inequality constraints ($\nabla c_i(x)$, all i such that $c_i(x) = 0$) and the gradients of all equality constraints ($\nabla d_i(x)$, $i = 1 : p$) are linearly independent.

Theorem [KKT conditions] Provided F, c and d are differentiable at x^* and the linear independence constraint qualification holds at x^*, if x^* is a local minimizer then there exist vectors λ^* and ν^* such that

$$c(x^*) \geq 0, \tag{45}$$

$$d(x^*) = 0, \tag{46}$$

$$c(x^*)^T \lambda^* = 0, \tag{47}$$

$$\lambda^* \geq 0, \tag{48}$$

$$\nabla \mathcal{L}(x^*, \lambda^*, \nu^*) = 0, \tag{49}$$

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

The triple (x,λ,ν) is a *KKT point* if it satisfies conditions (45)–(49).

Definition A.2 The *strict complementarity slackness conditions* hold at a first-order KKT point (x^*,λ^*,ν^*) if exactly one of λ_i^* and $c_i(x^*)$ is zero for each $i = 1:m$.

Theorem A.3 *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 (45)–(49) and for every τ satisfying*

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

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

References

[1] N.M. Alexandrov and M.Y. Hussaini (eds.), *Multidisciplinary optimization: State of the art*, Philadelphia, SIAM, 1997.

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

[3] K.A. Ariyawansa and A.J. Felt, *On a new collection of stochastic linear programming test problems*, Tech. Report TR-01-4, Department of Mathematics Washington State University, 2001.

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

[5] J.R. Birge and F. Louveaux, *Introduction to stochastic programming*, Springer-Verlag, New York, 1997.

[6] R.D. Braun, *Collaborative optimization: An architecture for large-scale distributed design.*, Ph.D. thesis, Stanford University, 1996.

[7] R.D. Braun and I.M. Kroo, *Development and application of the collaborative optimization architecture in a multidisciplinary design environment*, Multidisciplinary Design Optimization: State of the Art (N.M. Alexandrov and M.Y. Hussaini, eds.), 1997.

[8] P.H. Calamai and L.N. Vicente, *Generating quadratic bilevel programming test problems*, ACM Transactions on Mathematical Software **20** (1994), no. 1, 103–119.

[9] B.J. Chun and S.M. Robinson, *Scenario analysis via bundle decomposition*, Annals of Operations Research **56** (1995), 39–63.

[10] 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** (1994), no. 4, 754–776.

[11] A.V. DeMiguel, *Two decomposition algorithms for nonconvex optimization problems with global variables*, Ph.D. thesis, Stanford University, 2001.

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

[13] ———, *Generating optimization problems with global variables*, Tech. Report SOL 01-3, Dept. of Management Science and Engineering, Stanford University, 2001.

[14] ———, *Two decomposition algorithms for nonconvex optimization problems with global variables*, Tech. Report SOL 01-1, Dept. of Management Science and Engineering, Stanford University, 2001.

[15] A.V. DeMiguel and F.J. Nogales, *A superlinearly convergent interior-point decomposition algorithm using low-cost subproblem solutions*, Working paper, London Business School, 2002.

[16] S. Dempe, *Foundations of bilevel programming*, Kluwer Academic Publishers, Boston, 2002.

[17] A.V. Fiacco, *Introduction to sensitivity and stability analysis in nonlinear programming*, Academic Press, New York, 1983.

[18] A.V. Fiacco and G.P. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*, John Wiley & Sons, New York, 1968.

[19] R. Fletcher, *Practical methods of optimization*, Wiley, New York, 1987.

[20] A.M. Geoffrion, *Generalized Benders decomposition*, JOTA **10** (1972), no. 4, 237–260.

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

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

[23] T. Helgason and S.W. Wallace, *Approximate scenario solutions in the progressive hedging algorithm*, Annals of Operations Research **31** (1991), 425–444.

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

[25] Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical programs with equilibrium constraints*, Cambridge University Press, Cambridge, 1996.

[26] V.M. Manning, *Large-scale design of supersonic aircraft via collaborative optimization.*, Ph.D. thesis, Stanford University, 2000.

[27] W. Murray and F.J. Prieto, *A sequential quadratic programming algorithm using an incomplete solution of the subproblem*, SIAM J. Optimization **5** (1995), no. 3, 590–640.

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

[29] R.T. Rockafellar and R.J-B. Wets, *Scenarios and policy aggregation in optimization under uncertainty*, Mathematics of Operations Research **16** (1991), 119–147.

[30] A. Ruszczynski, *A regularized decomposition for minimizing a sum of polyhedral functions*, Mathematical Programming **35** (1986), 309–333.

[31] ______, *On convergence of an augmented lagrangian decomposition method for sparse convex optimization*, Mathematics of Operations Research **20** (1995), no. 3, 634–656.

[32] K. Shimizu, Y. Ishizuka, and J.F. Bard, *Nondifferentiable and two-level mathematical programming*, Kluwer Academic Publishers, Boston, 1997.

[33] I.P. Sobieski, *Multidisciplinary design using collaborative optimization.*, Ph.D. thesis, Stanford University, 1998.

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

[35] R. Van Slyke and R.J-B. Wets, *L-shaped linear programs with application to optimal control and stochastic programming*, SIAM Journal on Applied Mathematics **17** (1969), 638–663.

[36] S.J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.

[37] H. Yamashita and H. Yabe, *Superlinear and quadratic convergence of some primal–dual interior point methods for constrained optimization*, Mathematical Programming **75** (1996), :377–397.