

A Local Convergence Analysis of Bilevel Decomposition Algorithms

Victor DeMiguel

Walter Murray

Decision Sciences

Management Science and Engineering

London Business School

Stanford University

`avmiguel@london.edu`

`walter@stanford.edu`

18th April 2004

Multidisciplinary design optimization (MDO) problems are engineering design problems that require the consideration of the interaction between several design disciplines. Due to the organizational aspects of MDO problems, decomposition algorithms are often the only feasible solution approach. Decomposition algorithms reformulate the MDO problem as a set of independent subproblems, one per discipline, and a coordinating master problem. A popular approach to MDO problems is *bilevel decomposition algorithms*. 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 analyze their properties. In particular, we show that the proposed problem formulations are mathematically equivalent to the original problem and that the proposed algorithms 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 (MDO).

1 Introduction

Quoting Sobieszczanski-Sobieski and Haftka [19], Multidisciplinary Design Optimization (MDO) problems may be defined as problems that

... involve the design of systems where the interaction between several disciplines must be considered, and where the designer is free to significantly affect the system performance in more than one discipline.

The interdisciplinary nature of MDO problems makes them challenging from both a computational and an organizational perspective. From a computational perspective MDO problems integrate several disciplinary analysis, which usually require solving large-scale nonlinear problems. From an organizational perspective, MDO problems usually require the participation of several persons within one or more organizations. Moreover, often different computer codes are used to perform each of the disciplinary analysis. These are sophisticated codes that have been under development for many years and that are subject to constant modification. As a result, it is often impractical to integrate all of the disciplinary codes into a single hardware platform.

In this context, *decomposition* algorithms may be the only feasible alternative to finding the optimal design. Decomposition algorithms exploit the structure of MDO problems by reformulating them as a set of independent *subproblems*, one per discipline. Then, a so-called *master problem* is used to coordinate the subproblem solutions and find an overall problem minimizer. For a detailed discussion of the computational and organizational aspects of MDO problems see the monograph [1], the reviews [7, 19], and the research articles [31, 28, 27, 26, 20].

A popular decomposition approach to MDO problems is *bilevel decomposition algorithms*. Once decomposed into a master problem and a set of subproblems, the MDO is a particular type of *bilevel program* [21, 29, 13]. 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 [4] 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 [9]. Nondegeneracy is a common assumption when proving 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 [3]. Moreover, some convergence difficulties have been observed in the application of the algorithm to some simplified real-world problems [5, 30, 22].

In this paper, we propose two bilevel decomposition algorithms closely related to CO that overcome some of the difficulties associated with CO. 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. We analyze the local convergence properties of the proposed algorithms. The relevance of our analysis is twofold. Firstly, we show that the proposed problem formulations are *mathematically equivalent* to the original MDO problem. That is, under mild assumptions, KKT points of the MDO problem are KKT points of the decomposed problem and vice versa. Secondly, we show that the proposed master problem and subproblems are

nondegenerate. Consequently, we can show that standard nonlinear programming algorithms *converge locally at a fast rate* when applied to solve the proposed mater problem and subproblems.

The remainder of 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

To simplify our analysis, herein we consider a special type of MDO problem termed *quasi-separable* MDO problem [20]. The term quasi-separable refers to the fact that the objective function terms and constraints of each of the different systems are only coupled through a few of the variables, which we term *global variables*. This problem is general enough to cover many engineering design problems, yet simple enough to allow for rigorous analysis. The quasi-separable MDO problem may be stated as follows:

$$\begin{aligned}
 \min_{x, y_1, \dots, y_N} \quad & F_1(x, y_1) + F_2(x, y_2) + \dots + F_N(x, y_N) \\
 \text{s.t.} \quad & c_1(x, y_1) \geq 0 \\
 & c_2(x, y_2) \geq 0 \\
 & \vdots \\
 & c_N(x, y_N) \geq 0,
 \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$ are the *global* variables, $y_i \in \mathbb{R}^{n_i}$ are the i th system *local* variables, $c_i : \mathbb{R}^n \times \mathbb{R}^{n_i} \rightarrow \mathbb{R}^{m_i}$ are the i th system constraints, and $F_i : \mathbb{R}^n \times \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ is the objective function term corresponding to the i th 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. Herein, we refer to problem (1) as the MDO problem.

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 a CO-like bilevel decomposition algorithm may be applied to 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.

There are several types of bilevel decomposition algorithms. For expositional purposes, we consider a bilevel decomposition algorithm that is a close relative of collaborative optimization. The algorithm may be described as follows. Solve the following *master problem*:

$$\min_x F_1^*(x) + F_2^*(x) + \dots + F_N^*(x),$$

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

$$\begin{aligned} F_1^*(x) = \min_{y_1} F_1(x, y_1) & \quad \dots \quad F_N^*(x) = \min_{y_N} F_N(x, y_N) \\ \text{s.t. } c_1(x, y_1) \geq 0 & \quad \quad \quad \text{s.t. } c_N(x, y_N) \geq 0. \end{aligned}$$

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 disciplines 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.

Roughly speaking, 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 [15] 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 [32] 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, \dots, 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 [12] 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 $\nabla_{x,y_1,\dots,y_N} c_i(x, y_i)$ 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 $j \in \{1, 2, \dots, N\}$, 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 [4] 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 forced to converge to the same value (given by the *target* variable vector z) by using quadratic penalty functions. Several CO formulations have been considered in the literature [4, 5, 2]. To simplify the exposition we consider the following CO formulation. Solve the following master problem

$$\begin{aligned} \min_z \quad & \sum_{i=1}^N F_i(z, y_i) \\ \text{s.t.} \quad & p_i^*(z) = 0, \quad i = 1:N, \end{aligned}$$

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

$$\begin{aligned} p_1^*(z) = \min_{x_1, y_1} \|x_1 - z\|_2^2 & \quad \dots \quad p_N^*(z) = \min_{x_N, y_N} \|x_N - z\|_2^2 \\ \text{s.t.} \quad c_1(x_1, y_1) \geq 0. & \quad \text{s.t.} \quad c_N(x_N, y_N) \geq 0, \end{aligned}$$

where y_i and c_i and F_i are as defined in (1), the global variables $x_1, x_2, \dots, x_N \in \mathbb{R}^n$, and the target variables $z \in \mathbb{R}^n$. Note 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. 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 [15, 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 [23]. In fact, Alexandrov and Lewis [3] 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).

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 analyze the properties of IPD. The relevance of our analysis is twofold. Firstly, we show that under mild conditions the MDO problem and the IPD formulation are *mathematically equivalent*. That is, KKT points of the MDO problem are KKT points of the IPD formulation and vice versa. Secondly, we show that under standard nondegeneracy assumptions on the MDO minimizer, the IPD 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 force 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 $\gamma_k \rightarrow \infty$:

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

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

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

F_i , c_i , and y_i are as in problem (1), $x_i \in \mathbb{R}^n$ are the i th 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 $\gamma_k \rightarrow \infty$ 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 $\{\gamma_k\}$. To illustrate the behavior of the algorithm, we need to specify how to solve the master problem and subproblems. In practice a variety of methods could be used to solve these problems. However, here we use a BFGS quasi-Newton method [18, 16] to solve the master problem. 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 solve the subproblems we use the sequential quadratic programming (SQP) algorithm NPSOL [17].

The master problem objective and its gradient can be computed from the subproblem from the subprob-

lem minimizers. In particular, the master problem objective $F^*(z)$ can be evaluated at the subproblem solution $(x_i(z), y_i(z))$ as

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

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

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

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

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

$$+ \nabla_z L_i(x_i(z), y_i(z), \lambda_i(z), z), \quad (8)$$

where $x_i'(z)$, $y_i'(z)$, $\lambda_i'(z)$ denote the Jacobian matrices of $x_i(z)$, $y_i(z)$, $\lambda_i(z)$ 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

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

The resulting decomposition algorithm is outlined in Figure 1.

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

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  $\gamma$ .
  2. Solve master problem with current  $\gamma$ :
    repeat
      (a) Function evaluation: for  $i = 1 : N$  call NPSOL with  $\gamma$  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

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 Mathematical Equivalence and Nondegeneracy

In this section we analyze the theoretical properties of IPD. In particular, we show that, given a nondegenerate MDO minimizer (that is, given an MDO minimizer satisfying the linear independence constraint qualification, the strict complementarity slackness conditions, and the second order sufficient conditions; see Appendix A), there exists an equivalent trajectory of minimizers to the IPD master problem and subproblems that are also nondegenerate. Conversely, we show that, under mild conditions, an accumulation point of IPD nondegenerate local minimizers, is a KKT point for the MDO problem. The result is summarized in Theorems 4.10 and 4.11.

The importance of this result is twofold. Firstly, our analysis implies that under mild conditions the MDO problem and the proposed IPD formulation are *mathematically equivalent*. That is, nondegenerate KKT points of the MDO problem are nondegenerate KKT points of the IPD formulation and *vice versa*. Secondly, the nondegeneracy of the equivalent IPD minimizers implies that Newton-like algorithms will

converge locally at a fast rate when applied to solve the IPD formulation.

To prove the 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 result can be shown by realizing that mathematical equivalence and nondegeneracy are 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 establish that mathematical equivalence and nondegeneracy of a local minimizer is preserved by each of the three transformations listed above. Finally, in Section 4.3.5, we give two theorems summarizing the result.

In Section 4.4, we use the nondegeneracy results to prove that the algorithms proposed to solve the master problem and the subproblems converge superlinearly for each value of the penalty parameter.

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 [8, 11] for an exposition with N systems. The MDO problem with two systems is:

$$\begin{aligned} \min_{x, y_1, y_2} \quad & F_1(x, y_1) + F_2(x, y_2) \\ \text{s.t.} \quad & c_1(x, y_1) \geq 0, \\ & c_2(x, y_2) \geq 0, \end{aligned} \tag{10}$$

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

We assume there exists a minimizer (x^*, y_1^*, y_2^*) and an associated 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 the analysis in [32, 12] 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:

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

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

Herein we denote the Jacobian matrix of the system constraints $c_i(x, y_i)$ as the matrix $\nabla c_i(x, y_i) \in \mathbb{R}^{m_i \times (n+n_i)}$ whose m_i rows contain the gradients of the individual constraints $c_{ij}(x, y_i)$ for $j = 1 : m_i$.

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 $(\psi_1, \psi_2) \neq 0$ such that

$$\begin{pmatrix} (\nabla_x \hat{c}_1(x, y_1))^T & (\nabla_x \hat{c}_2(x, y_2))^T \\ (\nabla_{y_1} \hat{c}_1(x, y_1))^T & (\nabla_{y_2} \hat{c}_2(x, y_2))^T \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = 0, \tag{13}$$

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

$$\begin{pmatrix} (\nabla_x \hat{c}_1(x, y_1))^T & & & I & & \\ & (\nabla_x \hat{c}_2(x, y_2))^T & & & I & \\ (\nabla_{y_1} \hat{c}_1(x, y_1))^T & & & & & \\ & (\nabla_{y_2} \hat{c}_2(x, y_2))^T & & & & \\ & & & -I & & -I \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = 0, \tag{14}$$

with $\psi_3 = -(\nabla_x c_1(x, y_1))^T \psi_1$ and $\psi_4 = -(\nabla_x c_2(x, y_2))^T \psi_2$. Thus, condition A.1 does not hold for problem (11). Conversely, assume there exists $(\psi_1, \psi_2, \psi_3, \psi_4) \neq 0$ such that equation (14) holds. Then, by the first and second rows in (14) we know that $\psi_3 = -(\nabla_x c_1(x, y_1))^T \psi_1$ and $\psi_4 = -(\nabla_x c_2(x, y_2))^T \psi_2$. Then by the third, fourth, and fifth rows in (14) it follows that equation (13) holds for (ψ_1, ψ_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 (47)–(48) 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 (49) at (x, y_1, y_2) such that

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

Then we obviously have that

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

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 (49) and (50) 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

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

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

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

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 . \blacksquare

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:

$$\begin{aligned} \min_{x_1, x_2, y_1, y_2, z} \quad & \sum_{i=1}^2 [F_i(x_i, y_i) + \gamma \|x_i - z\|_2^2] \\ \text{s.t.} \quad & c_i(x_i, y_i) \geq 0 \quad i = 1 : 2. \end{aligned} \quad (19)$$

The quadratic penalty functions has been extensively analyzed in the nonlinear optimization literature. The following two theorems follow from standard results (see [15, 24]) and establish the mathematical equivalence of problems (11) and (19). The first theorem follows from Theorems 14 and 17 in [15] and shows that, given a nondegenerate minimizer to problem (11), there exists a trajectory of nondegenerate minimizers to (19) that converge to it.

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 $\lim_{\gamma \rightarrow \infty} (x_1(\gamma), x_2(\gamma), y_1(\gamma), y_2(\gamma), z(\gamma)) = (x_1, x_2, y_1, y_2, z)$.*

The second theorem follows from Theorem 17.2 in [15] and shows that, under mild assumptions, any

accumulation point of nondegenerate local minimizers to problem (19) is a KKT point of problem (11). To state the theorem we need to introduce some notation. We denote by $\hat{c}_i(x_i^*, y_i^*)$ the constraints that are active at (x_i^*, y_i^*) . That is, $\hat{c}_i(x_i^*, y_i^*)$ is the subset of the constraints $c_i(x_i, y_i)$ for which $c_i(x_i^*, y_i^*) = 0$.

Theorem 4.3 *Let $\{(x_1, x_2, y_1, y_2, z)\}_k$ be a sequence of KKT points for problem (19) at which condition A.1 holds and corresponding to a sequence of penalty parameters $\{\gamma\}_k$ with $\lim_{k \rightarrow \infty} \gamma_k = \infty$. Then, any limit point $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ of the sequence $\{(x_1, x_2, y_1, y_2, z)\}_k$ at which the gradients of the equality constraints $x_1^* = z^*$, $x_2^* = z^*$ and the gradients of the active inequality constraints $\hat{c}_1(x_1^*, y_1^*)$, $\hat{c}_2(x_2^*, y_2^*)$ are linearly independent is a KKT point of problem (11) at which condition A.1 holds.*

4.3.4 Decomposition

We now consider the third transformation operated on the MDO problem. 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 formulation. Namely, solve the following master problem

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

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

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

In this section, we show that problems (19) and (20)-(21) are mathematically equivalent. First, we introduce some useful definitions.

Definition 4.4 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 i th subproblem (21) with $z = z^*$.

Definition 4.5 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}_ϵ 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 4.6 *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 4.5 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,

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

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 4.5 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 4.5. ■

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

Theorem 4.7 *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 i th subproblem (21) with $z = z^*$.*

Proof: Note that the constraints in the i th 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 i th 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

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

where $\lambda = (\lambda_1, \lambda_2)$. The first four rows in (23) show that the KKT conditions hold for the i th subprob-

lem (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 i th 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 i th 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

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

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 4.8 *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 4.7, we know that the vector (x_i^*, y_i^*) is a minimizer satisfying the nondegeneracy conditions A.1-A.3 for the i th subproblem (21) with $z = z^*$. Therefore, by the implicit function theorem and Theorem 6 in [15], 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 i th 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 i th 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 $z \in \mathcal{N}_{\epsilon_3}(z^*)$,

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

Moreover, by Lemma 4.6 $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ 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

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

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

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

■

Finally, we show that a local minimizer of problem (20)-(21) at which the LICQ, SCS, and SOSC hold for the subproblems (21) is a KKT point for problem (19).

Theorem 4.9 *Let (x_1, x_2, y_1, y_2, z) be a local minimizer of (20)-(21) satisfying conditions A.1-A.3 for the subproblems (21), then (x_1, x_2, y_1, y_2, z) is a KKT point for problem (19) at which condition A.1 holds.*

Proof: Note that, if conditions A.1-A.3 hold at (x_1, x_2, y_1, y_2, z) for the subproblems, then, by the same arguments used in the proof of Theorem 4.8, we have that the master problem objective function $\sum_{i=1}^2 F_i^*(z)$ is twice continuously differentiable. Moreover, if (x_1, x_2, y_1, y_2, z) is a local minimizer to (20)-(21), the gradient of the master problem objective must be zero. Because $\nabla \sum_{i=1}^2 F_i^*(z) = 2\gamma(x_1 - z) + 2\gamma(x_2 - z)$, this in turn implies that

$$2\gamma(x_1 - z) + 2\gamma(x_2 - z) = 0. \quad (26)$$

Moreover, because A.1 holds at the subproblems, there exist unique Lagrange multipliers for which the KKT conditions hold for the subproblems (21) at (x_1, x_2, y_1, y_2, z) . The subproblem KKT conditions together with (26) imply that the KKT conditions hold at (x_1, x_2, y_1, y_2, z) for problem (19) with the same Lagrange multipliers that satisfy the subproblem KKT conditions.

Moreover, it is easy to see that if A.1 holds for the subproblems, it must also hold for problem (19), because the constraints of problem (19) are simply the aggregation of the constraints of both subproblems,

and the constraints of each subproblem depend on mutually exclusive sets of variables. Thus, the A.1 and the KKT conditions hold at (x_1, x_2, y_1, y_2, z) for problem (19). ■

4.3.5 Summary

The following two theorems summarize the mathematical equivalence and nondegeneracy results for IPD. The first theorem shows that for any nondegenerate minimizer to problem (10) there exists an equivalent trajectory of nondegenerate minimizers to problem (20)-(21).

Theorem 4.10 *Let (x, y_1, y_2) be a minimizer to problem (10) satisfying the nondegeneracy conditions A.1-A.3, then for γ sufficiently large, there exists a locally unique once continuously-differentiable trajectory of local minimizers to problem (20)-(21) $(x_1(\gamma), x_2(\gamma), y_1(\gamma), y_2(\gamma), z(\gamma))$ satisfying the nondegeneracy conditions A.1-A.3 for the subproblems (21) and condition A.3 for problem (20) and such that*

$$\lim_{\gamma \rightarrow \infty} (x_1(\gamma), x_2(\gamma), y_1(\gamma), y_2(\gamma), z(\gamma)) = (x, x, y_1, y_2, x).$$

Proof: The result follows from Theorems 4.1, 4.2, 4.6, 4.7, and 4.8. ■

The second theorem shows that any limit point of a sequence of nondegenerate minimizers to problem (20)-(21) converges to a minimizer to problem (10) provided that at the limit point the gradients of the active constraints are linearly independent.

Theorem 4.11 *Let $\{(x_1, x_2, y_1, y_2, z)\}_k$ be a sequence of local minimizers of problem (20)-(21) corresponding to a sequence of penalty parameters $\{\gamma\}_k$ with $\lim_{k \rightarrow \infty} \gamma_k = \infty$ and such that the nondegeneracy conditions A.1-A.3 are satisfied for the subproblems (21) at $\{(x_1, x_2, y_1, y_2, z)\}_k$. Then, any limit point $(x_1^*, x_2^*, y_1^*, y_2^*, z^*)$ of the sequence $\{(x_1, x_2, y_1, y_2, z)\}_k$ at which the gradients of the equality constraints $x_1^* = z^*$, $x_2^* = z^*$ and the gradients of the active inequality constraints $\hat{c}_1(x_1^*, y_1^*)$, $\hat{c}_2(x_2^*, y_1^*)$ are linearly independent is a KKT point of problem (10) at which condition A.1 holds.*

Proof: The result follows from Theorems 4.1, 4.3, and 4.9. ■

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 a sequential quadratic programming (SQP) algorithm converges locally at a superlinear rate when applied to the subproblem (21).

Theorem 4.12 *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 $\lim_{\gamma \rightarrow \infty} z^*(\gamma) = x^*$. 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 i th 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 the SQP algorithm converge locally and superlinearly to $(x_i(z_k), y_i(z_k))$ when applied to the i th subproblem (21) with $z = z_k$.*

Proof: Part 1 follows from Theorems 4.1, 4.2, and 4.8. To prove Part 2, note that from Theorem 4.8 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 [16].

From Theorems 4.1, 4.2, and 4.7 we know that there exists a minimizer to the i th 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 [14]. Finally, the local and superlinear convergence of the SQP algorithm (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 [23]. ■

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 [24, 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 interested reader is referred to [8] for a detailed exposition.

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 analyze the properties of EPD. The importance of our analysis is twofold. Firstly, we show that under mild conditions the MDO problem and the EPD formulation are *mathematically equivalent*. Secondly, we show that under standard nondegeneracy assumptions on the MDO minimizer, the IPD master problem and subproblems are also nondegenerate. Using these nondegeneracy results, we show in Section 5.4 that the optimization algorithms used to solve the master problem and the subproblems will *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 force 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:

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

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

$$\begin{aligned} F_i^*(z) = \min_{x_i, y_i, s_i, t_i} & F_i(x_i, y_i) + \gamma e^T (s_i + t_i) \\ \text{s.t.} & c_i(x_i, y_i) \geq 0 \\ & x_i + s_i - t_i = z \\ & s_i, t_i \geq 0. \end{aligned} \tag{27}$$

It is obvious that, at a minimizer to subproblem (27), $\|x_i - z\|_1 = e^T (s_i + t_i)$. Moreover, when using the exact penalty function, we have $x_i = z$ for $i = 1 : N$ provided that γ is sufficiently large. 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 (27) are not linearly independent in general. This, by standard sensitivity theory [15], 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 circumvented by introducing barrier terms to remove the nonnegativity constraints from the subproblems. The result is the EPD algorithm: solve the following master problem for a decreasing sequence of barrier parameters $\{\mu_k\}$ such that $\mu_k \rightarrow 0$:

$$\min_z \sum_{i=1}^N F_i^*(z), \quad (28)$$

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

$$\begin{aligned} \min_{x_i, y_i, s_i, t_i} \quad & F_i(x_i, y_i) + \gamma e^T (s_i + t_i) - \mu \sum_{j=1}^n (\log s_{ij} + \log t_{ij}) \\ \text{s.t.} \quad & c_i(x_i, y_i) \geq 0, \\ & x_i + s_i - t_i = z. \end{aligned} \quad (29)$$

5.2 Algorithm Statement

To solve the EPD master problem (28), we use the same BFGS quasi-Newton method we proposed for the IPD master problem. However, we use a primal-dual interior-point method [33] to solve the subproblems (29). These methods are especially designed to deal with problems that include barrier functions. We coded the method in the MATLAB file PDSOL.

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

$$F^*(z_k) = \sum_{i=1}^N F_i(x_{ik}, y_{ik}) + \gamma e^T (s_{ik} + t_{ik}) - \mu \sum_{j=1}^n (\log(s_{ik})_j + \log(t_{ik})_j). \quad (30)$$

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 [15] can be used to compute the master problem objective gradient as:

$$\nabla F^*(z_k) = \sum_{i=1}^N \nabla F_i^*(z_k) = - \sum_{i=1}^N \lambda_{z_i}, \quad (31)$$

where λ_{z_i} 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.

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

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

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 (30) and (31).

(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

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 Mathematical Equivalence and Nondegeneracy

In this section, we analyze the properties of EPD. Firstly, we show that under mild conditions the MDO problem and the proposed EPD formulation are mathematically equivalent. Secondly, we show that under standard nondegeneracy assumptions on the MDO minimizer, the EPD master problem and subproblems are also nondegenerate.

In particular, we show that, given a nondegenerate MDO minimizer, there exists an equivalent trajectory of minimizers to the EPD master problem and subproblems that are also nondegenerate. Conversely, we show that, under mild conditions, an accumulation point of EPD nondegenerate local minimizers, is a KKT point for the MDO problem. The result is summarized in Theorems 5.9 and 5.10.

To prove the result, it is essential to note that the EPD master problem (28) 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. We shall show that mathematical equivalence

and nondegeneracy are preserved by each of these transformations. 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 [8] 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:

$$\begin{aligned} \min_{x_i, y_i, s_i, t_i, z} \quad & \sum_{i=1}^2 F_i(x_i, y_i) + \gamma e^T (s_i + t_i) \\ \text{s.t.} \quad & c_i(x_i, y_i) \geq 0, \quad i = 1, 2, \\ & x_i + s_i - t_i = z, \quad i = 1, 2, \\ & s_i, t_i \geq 0, \quad i = 1, 2, \end{aligned} \tag{32}$$

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 (32) that is nondegenerate and vice versa.

Theorem 5.1 *Provided*

$$\gamma > \left\| \begin{pmatrix} \nabla_x F_1(x, y_1) - (\nabla_x c_1(x, y_1))^T \lambda_1 \\ \nabla_x F_2(x, y_2) - (\nabla_x c_2(x, y_2))^T \lambda_2 \end{pmatrix} \right\|_{\infty}, \tag{33}$$

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 (32) satisfying the nondegeneracy conditions A.1-A.3 with Lagrange multipliers

where $\psi = (\psi_1, \psi_2, \dots, \psi_8)$ with $\psi_3 = -\psi_5 = \psi_7 = -(\nabla_x \hat{c}_1(x, y_1))^T \psi_1$, and $\psi_4 = -\psi_6 = \psi_8 = -(\nabla_x \hat{c}_2(x, y_2))^T \psi_2$. Thus, condition A.1 does not hold for problem (32). Conversely, it is easy to show by similar arguments that if there exists $\psi_1, \dots, \psi_8 \neq 0$ such that equation (36) holds, then (35) holds for $(\psi_1, \psi_2) \neq 0$.

Now we turn to the KKT conditions. First, it is obvious that the feasibility conditions (47)–(48) 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 (32). Moreover, assume there exists $(\lambda_1, \lambda_2) \geq 0$ satisfying the complementarity condition (49) at (x, y_1, y_2) such that

$$\begin{pmatrix} \sum_{i=1}^2 \nabla_x F_i(x, y_i) \\ \nabla_{y_1} F_1(x, y_1) \\ \nabla_{y_2} F_2(x, y_2) \end{pmatrix} = J_1^T \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}. \quad (37)$$

Then we obviously have that

$$\begin{pmatrix} \nabla_x F_1(x, y_1) \\ \nabla_x F_2(x, y_2) \\ \nabla_{y_1} F_1(x, y_1) \\ \nabla_{y_2} F_2(x, y_2) \\ \gamma e \\ \gamma e \\ \gamma e \\ \gamma e \\ 0 \end{pmatrix} = J_2^T \lambda, \quad (38)$$

with λ as given by (34). Conversely, if there exists $(\lambda_1, \lambda_2, \dots, \lambda_8)$ satisfying equation (38), then $(\lambda_1, \lambda_2, \dots, \lambda_8)$ must be as given in (34) and (37) is satisfied at (x, y_1, y_2) for problem (10) with (λ_1, λ_2) . Furthermore, it is clear that, provided condition (33) holds, the complementarity, nonnegativity, and strict complementarity conditions [(49), (50), 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 (32) with the Lagrange multiplier given by (34).

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 $s_1, s_2, t_1, t_2 = 0$ and $x_1, x_2, z = x$ for problem (32). In particular, given $\tau_1 = (\tau_x, \tau_{y_1}, \tau_{y_2})$ such that

$$\hat{J}_1 \tau_1 = 0, \quad (39)$$

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

$$\hat{J}_2 \tau_2 = 0. \quad (40)$$

Conversely, if τ_2 satisfies (40), 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 (39). 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 (32) 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 (32).

■

The following theorem shows that for γ large enough, and assuming the gradients of the equality constraints $x_1 = z$, $x_2 = z$, and the active inequalities $\hat{c}_1(x_1, y_1) \geq 0$ and $\hat{c}_2(x_2, y_2) \geq 0$ are linearly independent, a local minimizer of problem (32) must be feasible with respect to problem (10).

Theorem 5.2 *Assume there exists M such that $X = (x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ is a local minimizer to problem (32) for all $\gamma > M$ and such that the nondegeneracy conditions A.1-A.3 are satisfied at X and the gradients of the equality constraints $x_1 = z$, $x_2 = z$, and the active inequalities $\hat{c}_1(x_1, y_1) \geq 0$ and $\hat{c}_2(x_2, y_2) \geq 0$ are linearly independent at X . Then, we must have that $s_1 = s_2 = t_1 = t_2 = 0$, $x_1 = x_2 = z$, and the point (x, y_1, y_2) is a local minimizer to problem (10) satisfying the nondegeneracy conditions A.1-A.3.*

Proof: Assume there exists M such that for all $\gamma > M$, X is a nondegenerate minimizer of (32) at which the gradients of the equality constraints $x_1 = z$, $x_2 = z$, and the active inequalities $\hat{c}_1(x_1, y_1) \geq 0$ and $\hat{c}_2(x_2, y_2) \geq 0$ are linearly independent. Then, for any $\gamma > M$ and $\Delta\gamma > 0$, we must have that X is a local minimizer of (32) for both γ and $\hat{\gamma} = \gamma + \Delta\gamma$. Therefore, the KKT conditions must hold for both γ and $\hat{\gamma}$ at X . Subtracting these two sets of KKT conditions we get the following equation

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \Delta\gamma e \\ \Delta\gamma e \\ \Delta\gamma e \\ \Delta\gamma e \\ 0 \end{pmatrix} = \hat{J}_2^T \Delta\lambda, \quad (41)$$

where \hat{J}_2 is the Jacobian matrix of the equality and active inequality constraints at X . The proof continues by contradiction. Assume that some of the nonnegativity bounds $s_i \geq 0$, $t_i \geq 0$ were not active at X (that is, assume the gradients of some of the nonnegativity bounds were not in \hat{J}_2^T). Then, because the left-hand side of (41) is a linear combination of the gradients of the nonnegativity bounds and \hat{J}_2 does not contain the gradients of all these nonnegativity bounds, we have that (41) would imply

that the gradients of the equality constraints $x_i + s_i - t_i = z$, the nonnegativity bounds $s_i, t_i \geq 0$ for $i = 1, 2$, and the active inequalities $\hat{c}_i(x_i, y_i) \geq 0$ for $i = 1, 2$ are linearly *dependent*. But note that this contradicts the assumption that the gradients of the equality constraints $x_1 = z, x_2 = z$, and the active inequalities $\hat{c}_1(x_1, y_1) \geq 0$ and $\hat{c}_2(x_2, y_2) \geq 0$ are linearly independent at X , because it is easy to see that the gradients of the equality constraints $x_1 = z, x_2 = z$ are linearly independent at X if and only if the gradients of the equality constraints $x_1 + s_1 - t_1 = z, x_2 + s_2 - t_2 = z$, and the gradients of all (active and inactive) nonnegativity bounds $s_i \geq 0, t_i \geq 0$ are linearly independent. Thus, we must have that $s_1 = s_2 = t_1 = t_2 = 0$ and thus $x_1 = x_2 = z$. Then, the result then follows from Theorem 5.1. \blacksquare

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:

$$\begin{aligned} \min_{x_i, y_i, s_i, t_i, z} \quad & \sum_{i=1}^2 [F_i(x_i, y_i) + \gamma e^T (s_i + t_i) - \mu \sum_{j=1}^n (\log s_{ij} + \log t_{ij})] \\ \text{s.t.} \quad & c_i(x_i, y_i) \geq 0, \quad i = 1, 2, \\ & x_i + s_i - t_i = z, \quad i = 1, 2, \end{aligned} \tag{42}$$

where μ is the barrier parameter.

Barrier functions have been extensively analyzed in the nonlinear optimization literature. The following two theorems establish the mathematical equivalence of problems (32) and (42). The first theorem follows from Theorems 14 and 17 in [15] and shows that, given a nondegenerate minimizer of problem (32), there exists a trajectory of nondegenerate minimizers to problem (42) that converge to it.

Theorem 5.3 *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 (32), then there exists $\epsilon > 0$ such that for $\mu \in (0, \epsilon)$ there exists a unique once continuously differentiable trajectory of minimizers to problem (42) $X(\mu)$ satisfying the the nondegeneracy conditions A.1-A.3 such that $\lim_{\mu \rightarrow 0} X(\mu) = X$.*

The second theorem follows from Theorem A.4 in the Appendix and shows that any accumulation point of nondegenerate local minimizers to problem (42) is a KKT point of problem (32).

Theorem 5.4 *Let $\{(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)\}_k$ be a sequence of KKT points for problem (42) at which condition A.1 holds and corresponding to a sequence of positive barrier parameters $\{\mu\}_k$ with $\lim_{k \rightarrow \infty} \mu_k = 0$. Then, any limit point $(x_1^*, x_2^*, y_1^*, y_2^*, s_1^*, s_2^*, t_1^*, t_2^*, z^*)$ of the sequence $\{(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)\}_k$ at which the gradients of the active constraints of problem (32) are linearly independent is a KKT point of problem (32) at which condition A.1 holds.*

5.3.4 Decomposition

We now consider the last transformation operated on the MDO problem. Note that problem (42) 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

$$\min_z \sum_{i=1}^2 F_i^*(z), \quad (43)$$

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

$$\begin{aligned} \min_{x_i, y_i, s_i, t_i} \quad & F_i(x_i, y_i) + \gamma e^T (s_i + t_i) - \mu \sum_{j=1}^n (\log s_{ij} + \log t_{ij}) \\ \text{s.t.} \quad & c_i(x_i, y_i) \geq 0, \\ & x_i + s_i - t_i = z. \end{aligned} \quad (44)$$

In this section, we show that problems (42) and (43)-(44) are mathematically equivalent. The proof to the following lemma parallels that of Lemma 4.6.

Lemma 5.5 *If $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ is a minimizer to problem (42) satisfying the second order sufficient conditions A.3, then it is also a strict local minimizer to problem (43).*

In the following two theorems we show that the proposed subproblem and master problem are non-degenerate.

Theorem 5.6 *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 (42), 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 (44) with z .*

Proof: Note that, because of the elastic variables, the gradients of the equality constraints in the i th subproblem (44) are always linearly independent from the gradients of the active inequality constraints. Moreover, if the linear independence constraint qualification holds for problem (42) 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 i th subproblem.

Now we turn to proving that the KKT conditions hold for the i th subproblem. Let $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ be the unique Lagrange multiplier vector for problem (42) at $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$. By the KKT conditions we have

$$\begin{pmatrix} \nabla_{x_1} F_1(x_1, y_1) \\ \nabla_{x_2} F_2(x_2, y_2) \\ \nabla_{y_1} F_1(x_1, y_1) \\ \nabla_{y_2} F_2(x_2, y_2) \\ \gamma e + \mu S_1^{-1} e \\ \gamma e + \mu S_2^{-1} e \\ \gamma e + \mu T_1^{-1} e \\ \gamma e + \mu T_2^{-1} e \\ 0 \end{pmatrix} = \begin{pmatrix} (\nabla_{x_1} c_1(x_1^*, y_1^*))^T & 0 & I & 0 \\ 0 & (\nabla_{x_2} c_2(x_2^*, y_2^*))^T & 0 & I \\ (\nabla_{y_1} c_1(x_1^*, y_1^*))^T & 0 & 0 & 0 \\ 0 & (\nabla_{y_2} c_2(x_2^*, y_2^*))^T & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ 0 & 0 & -I & 0 \\ 0 & 0 & 0 & -I \\ 0 & 0 & -I & -I \end{pmatrix} \lambda, \quad (45)$$

where S_i and T_i are the diagonal matrices whose diagonal are the vectors s_i , t_i , and $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Notice that by the first, third, fifth, and seventh rows in (45) imply 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 (42) with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, then they obviously also hold for the first subproblem (44) 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 i th subproblem (44). 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 (44). 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 (42). 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 (44), and $\nabla^2 \mathcal{L}_2$ is the Hessian of the Lagrangian for problem (42). The same argument can be applied for the second subproblem. \blacksquare

The proof to the following theorem parallels that of Theorem 4.8

Theorem 5.7 *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 (42) satisfying the nondegeneracy conditions A.1-A.3, then:*

1. *the objective function of the master problem (43) 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.*

Finally, the proof of the following Theorem parallels that of Theorem 4.11.

Theorem 5.8 *Let $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ be a local minimizer of (43)-(44) satisfying conditions A.1-A.3 for the subproblems (44), then $(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)$ is a KKT point for problem (42) at which condition A.1 holds.*

5.3.5 Summary

The following two theorems summarize the mathematical equivalence and nondegeneracy results for EPD. Note that the assumptions are equivalent to those made in the analysis of IPD.

Theorem 5.9 *Let (x, y_1, y_2) be a minimizer to problem (10) satisfying the nondegeneracy conditions A.1-A.3, then there exists M such that for $\gamma > M$ and for μ sufficiently small, there exists a locally unique once continuously-differentiable trajectory of local minimizers to problem (43)-(44) satisfying the nondegeneracy conditions A.1-A.3 for the subproblems (44) and condition A.3 for problem (43) and such that*

$$\lim_{\mu \rightarrow 0} (x_1(\mu), x_2(\mu), y_1(\mu), y_2(\mu), s_1(\mu), s_2(\mu), t_1(\mu), t_2(\mu), z(\mu)) = (x, x, y_1, y_2, 0, 0, 0, 0, x).$$

Proof: The result follows from Theorems 5.1, 5.3, 5.5, 5.6, and 5.7, with the existence of a finite M is guaranteed by Theorem 5.1. ■

Theorem 5.10 *Let $\{X\}_k = \{(x_1, x_2, y_1, y_2, s_1, s_2, t_1, t_2, z)\}_k$ be a sequence of local minimizers of problem (43)-(44) corresponding to a sequence of barrier parameters $\{\mu\}_k$ with $\lim_{k \rightarrow \infty} \mu_k = 0$ and such that the nondegeneracy conditions A.1-A.3 are satisfied for the subproblems (44) at $\{X\}_k$. Moreover, let X^* be a limit point of the sequence $\{X\}_k$ at which the gradients of the equality constraints $x_1^* = z^*$, $x_2^* = z^*$, and the active inequality constraints $\hat{c}_1(x_1^*, y_1^*)$ and $\hat{c}_2(x_2^*, y_2^*)$ are linearly independent. Then, there exists M such that if $\gamma > M$, X^* is a KKT point of problem (10) at which condition A.1 holds.*

Proof: The result follows from Theorems 5.1, 5.2, 5.4, and 5.8, with the existence of a finite M guaranteed by Theorem 5.1. ■

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 (44).

Theorem 5.11 *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 $\lim_{\mu \rightarrow 0} z^*(\mu) = x^*$. 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 i th subproblem (44) 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 (44) with $z = z_k$.

Proof: Part 1 follows from Theorems 5.1, 5.3, and 5.7. To prove Part 2, note that from Theorem 5.7 we know that $F^*(z)$ may be defined locally as a twice continuously-differentiable function and its Hessian matrix is positive definite at $z^*(\mu)$. The local and superlinear convergence of the iterates generated by the master problem algorithm follows from standard theory for quasi-Newton unconstrained optimization algorithms [16].

From Theorems 5.1, 5.3, and 5.6 we know that there exists a minimizer to the i th subproblem (44) $(x_i(z^*(\mu)), y_i(z^*(\mu)), s_i(z^*(\mu)), t_i(z^*(\mu)))$ 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 [14]. Finally, the local and superlinear convergence of PDSOL (Part 4) when applied to the subproblem (44) follows from the nondegeneracy of the i th 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 [34]. ■

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.

We test the algorithms on the nonlinear test problem set of DeMiguel and Murray [10], which is a modification of the bilevel programming test problems by Calamai and Vicente [6]. 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) [25]. 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 [10]. 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 where 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 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.

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

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.

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

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, IPD and EPD successfully solved the two sample test problems proposed in [3] on which CO failed.

7 Conclusion

We propose two bilevel decomposition algorithms and analyze their properties. Firstly, we show that they are mathematically equivalent to MDO under mild assumptions. Secondly, we show that under standard nondegeneracy assumptions on the MDO minimizer, the proposed master problem and subproblems are nondegenerate as well. 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 [25]. 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 thank the Associate Editors and three referees for their detailed comments and sugges-

tions. 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

$$\begin{aligned} \min_x \quad & F(x) \\ \text{s.t.} \quad & c(x) \geq 0, \\ & d(x) = 0, \end{aligned} \tag{46}$$

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

$$c(x^*) \geq 0, \tag{47}$$

$$d(x^*) = 0, \tag{48}$$

$$c(x^*)^T \lambda^* = 0, \tag{49}$$

$$\lambda^* \geq 0, \tag{50}$$

$$\nabla \mathcal{L}(x^*, \lambda^*, \nu^*) = 0, \tag{51}$$

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

The triple (x, λ, ν) is a *KKT point* if it satisfies conditions (47)–(51).

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 (47)–(51) and for every τ satisfying*

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

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

Theorem A.4 *Let $\{x\}_k$ be a sequence of KKT points of the barrier problem*

$$\begin{aligned}\min_x \quad & F(x) - \mu \log(c(x)) \\ \text{s.t.} \quad & d(x) = 0,\end{aligned}\tag{52}$$

corresponding to a sequence of positive barrier parameters $\{\mu\}_k$ such that $\lim_{k \rightarrow \infty} \mu_k = 0$. Then, any limit point of $\{x\}_k$ at which the gradients of the equality constraints $d(x)$ and the active inequality constraints $\hat{c}(x)$ are linearly independent is a KKT point of problem (46).

Proof: Because $\{x\}_k$ is a sequence of KKT points of problem (52), we must have that for all k there exist ν_k such that

$$\nabla F(x_k) - (\nabla c(x_k))^T \mu_k C(x_k)^{-1} e + (\nabla d(x_k))^T \nu_k = 0,\tag{53}$$

where $C(x_k)$ is the diagonal matrix whose main diagonal is $c(x_k)$ and e is the vector of ones. Let x be a limit point of $\{x\}_k$. By taking limits as $k \rightarrow \infty$ we obtain

$$\nabla F(x) - (\nabla c(x))^T \lim_{k \rightarrow \infty} \mu_k C(x_k)^{-1} e + (\nabla d(x))^T \lim_{k \rightarrow \infty} \nu_k = 0.\tag{54}$$

Note that for the inactive inequality constraints ($c_i(x)$ such that $c_i(x^*) > 0$), we have that $\lim_{k \rightarrow \infty} \mu_k / c_i(x_k) = 0$ because $\lim_{k \rightarrow \infty} \mu_k = 0$ and $c_i(x^*) > 0$. This together with (54) imply that

$$\nabla F(x) - (\nabla \hat{c}(x))^T \lim_{k \rightarrow \infty} \mu_k \hat{C}(x_k)^{-1} e + (\nabla d(x))^T \lim_{k \rightarrow \infty} \nu_k = 0,\tag{55}$$

where $\hat{C}(x)$ is the diagonal matrix whose main diagonal is the vector of active inequality constraints $\hat{c}(x)$. Moreover, the linear independence of the gradients of $\hat{c}(x)$ and $d(x)$ at the limit point x implies that there exist unique and finite λ and ν such that $\lambda = \lim_{k \rightarrow \infty} \mu_k \hat{C}(x_k)^{-1} e$ and $\nu = \lim_{k \rightarrow \infty} \nu_k$. Consequently, condition (51) holds at x for problem (46) with the multipliers of the inactive inequality constraints equal to zero, the multipliers of $\hat{c}(x)$ equal to λ , and the multipliers of $d(x)$ equal to ν .

Also, note that if $\{x\}_k$ is a sequence of KKT points of (52), then $\lim_{k \rightarrow \infty} d(x_k) = d(x) = 0$. Moreover, because $\mu_k > 0$ and, by the properties of the barrier function, $c(x_k) > 0$, we must have that $\lambda = \lim_{k \rightarrow \infty} \mu_k \hat{C}(x_k)^{-1} e \geq 0$. Finally, by the definition of λ and the fact that $\lim_{k \rightarrow \infty} \mu_k = 0$, we know that the complementarity condition (49) must hold. ■

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, *Comparative properties of collaborative optimization and other approaches to MDO*, Proc., First ASMO UK/ISSMO conference on Engineering Design Optimization, MCB Press, Jul 1999.
- [3] ———, *Analytical and computational aspects of collaborative optimization*, Tech. Report TM-2000-210104, NASA, 2000.
- [4] R.D. Braun, *Collaborative optimization: An architecture for large-scale distributed design.*, Ph.D. thesis, Stanford University, 1996.
- [5] 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.
- [6] P.H. Calamai and L.N. Vicente, *Generating quadratic bilevel programming test problems*, ACM Transactions on Mathematical Software **20** (1994), no. 1, 103–119.
- [7] 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.
- [8] A.V. DeMiguel, *Two decomposition algorithms for nonconvex optimization problems with global variables*, Ph.D. thesis, Stanford University, 2001.
- [9] 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.
- [10] ———, *Generating optimization problems with global variables*, Tech. Report SOL 01-3, Dept. of Management Science and Engineering, Stanford University, 2001.

- [11] ———, *Two decomposition algorithms for nonconvex optimization problems with global variables*, Tech. Report SOL 01-1, Dept. of Management Science and Engineering, Stanford University, 2001.
- [12] 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.
- [13] S. Dempe, *Foundations of bilevel programming*, Kluwer Academic Publishers, Boston, 2002.
- [14] A.V. Fiacco, *Introduction to sensitivity and stability analysis in nonlinear programming*, Academic Press, New York, 1983.
- [15] A.V. Fiacco and G.P. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*, John Wiley & Sons, New York, 1968.
- [16] R. Fletcher, *Practical methods of optimization*, Wiley, New York, 1987.
- [17] 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.
- [18] P.E. Gill, W. Murray, and M.H. Wright, *Practical optimization*, Academic Press, New York, 1981.
- [19] R.T. Haftka and J. Sobieszczanski-Sobieski, *Multidisciplinary aerospace design optimization: Survey of recent developments*, *Structural Optimization* **14** (1997), 1–23.
- [20] R.T. Haftka and L.T. Watson, *Multidisciplinary design optimization with quasiseparable subsystems*, Tech. Report TR-02-09, Computer Science, Virginia Polytechnic Institute and State University, 2002.
- [21] Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical programs with equilibrium constraints*, Cambridge University Press, Cambridge, 1996.
- [22] V.M. Manning, *Large-scale design of supersonic aircraft via collaborative optimization.*, Ph.D. thesis, Stanford University, 2000.
- [23] 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.
- [24] J. Nocedal and S.J. Wright, *Numerical optimization*, Springer-Verlag, New York, 1999.
- [25] R.T. Rockafellar and R.J-B. Wets, *Scenarios and policy aggregation in optimization under uncertainty*, *Mathematics of Operations Research* **16** (1991), 119–147.
- [26] J.F. Rodriguez, J.E. Renaud, and L.T. Watson, *Convergence of trust region augmented lagrangian methods using variable fidelity approximation data*, *Structural Optim.* **15** (1998), 141–156.

- [27] ———, *Trust region augmented lagrangian methods for sequential response surface approximation and optimization*, ASME M. Mech. Design **120** (1998), 58–66.
- [28] J. Shankar, C.J. Ribbens, R.T. Haftka, and L.T. Watson, *Computational study of a nonhierarchical decomposition algorithm*, Comput. Optim. Appl. **2** (1993), 273–293.
- [29] K. Shimizu, Y. Ishizuka, and J.F. Bard, *Nondifferentiable and two-level mathematical programming*, Kluwer Academic Publishers, Boston, 1997.
- [30] I.P. Sobieski, *Multidisciplinary design using collaborative optimization.*, Ph.D. thesis, Stanford University, 1998.
- [31] J. Sobieszczanski-Sobieski, *Optimization by decomposition: a step from hierarchic to nonhierarchic systems*, Second NASA/Air Force Symp. on Recent Advances in Multidisciplinary Analysis and Optimization, 1988.
- [32] 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.
- [33] S.J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.
- [34] 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.