

On Decomposition Methods for a Class of Partially Separable Nonlinear Programs

Victor DeMiguel

Management Science and Operations, London Business School, Regent’s Park, London NW1 4SA, United Kingdom,
avmiguel@london.edu, <http://faculty.london.edu/avmiguel>

Francisco J. Nogales

Department of Statistics, Universidad Carlos III de Madrid, Spain,
fcojavier.nogales@uc3m.es, <http://www.est.uc3m.es/nogales>

We study two different decomposition algorithms for the general (nonconvex) partially separable nonlinear program (PSP): bilevel decomposition algorithms (BDAs) and Schur interior-point methods (SIPMs). BDAs solve the problem by breaking it into a master problem and a set of independent subproblems, forming a type of bilevel program. SIPMs, on the other hand, apply an interior-point technique to solve the problem in its original (integrated) form, but then use a Schur complement approach to solve the Newton system in a decentralized manner. Our first contribution is to establish a theoretical relationship between these two types of decomposition algorithms. This is a first step toward closing the gap between the incipient local convergence theory of BDAs and the mature local convergence theory of interior-point methods. Our second contribution is to show how SIPMs can be modified to solve problems for which the Schur complement matrix is not invertible in general. The importance of this contribution is that it substantially enlarges the class of problems that can be addressed with SIPMs.

Key words: Bilevel programming; Schur complement; decomposition algorithms

MSC2000 subject classification: Primary: 49M27; secondary: 90C30, 90C51

OR/MS subject classification: Primary: programming: nonlinear algorithms; secondary: programming: nonlinear applications; industries: transportation/shipping

History: Received December 23, 2005; revised October 5, 2006.

1. Introduction. We consider the following *partially separable nonlinear program* (PSP):

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

where $x \in \mathcal{R}^n$ are the *global* variables, $y_k \in \mathcal{R}^{n_k}$ for $k = 1 : K$ are the *local* variables, and $c_k(x, y_k) : \mathcal{R}^{n+n_k} \rightarrow \mathcal{R}^{m_k}$ and $F_k(x, y_k) : \mathcal{R}^{n+n_k} \rightarrow \mathcal{R}$ for $k = 1 : K$ are smooth functions. Note that the PSP is separable on the local variables; that is, if one sets the value of the global variables to a fixed value, the problem breaks into K independent subproblems. This suggests decomposition algorithms may be used to solve the PSP advantageously.

We focus on two decomposition algorithms for the general (nonconvex) version of the PSP: *bilevel decomposition algorithms* (BDAs) and SIPMs.¹ BDAs (Tammer [32], Braun and Kroo [6], DeMiguel and Murray [14]) solve the PSP by breaking it into a *master problem* and a set of K independent *subproblems*. These subproblems, together with the master problem, form a particular type of *bilevel program*; for a review of bilevel programming see the recent paper by Colson et al. [10] or the monographs (Shimizu et al. [31], Dempe [15]). Just as bilevel programming methods, BDAs 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 are used to compute the master problem derivatives and their associated Newton direction.

Our motivation to consider BDAs is that they are very popular among engineers solving the so-called *multi-disciplinary design optimization* problem (MDO) (Alexandrov and Hussaini [1], Cramer et al. [12], Haftka and

¹ Two popular types of decomposition algorithms for the convex version of the PSP are cutting-plane methods (Benders [3], Van Slyke and Wets [33], Geoffrion [20], Ruszczyński [29], Medhi [25], Robinson [27]) and augmented Lagrangian methods (Cohen and Miara [9], Rockafellar and Wets [28], Ruszczyński [30]). We do not consider cutting-plane methods because they rely heavily on convexity assumptions. Likewise, the theory of augmented Lagrangian decomposition methods (Cohen and Miara [9], Rockafellar and Wets [28], Ruszczyński [30]) depends on convexity assumptions. Moreover, augmented Lagrangian methods may converge slowly in practice (see Helgason and Wallace [23], Chun and Robinson [8], DeMiguel and Murray [14]).

Sobieszcanski-Sobieski [22]). The MDO arises in engineering design projects that require the consideration of several *disciplinary analysis*. For instance, when designing an airplane, one must consider both a structural and an aerodynamical analysis. Often, a different group of engineers is in charge of each of these disciplinary analysis. Moreover, the different groups often rely on sophisticated software codes (known as legacy codes) that have been under development for many years and whose method of use is subject to constant modification. Integrating all of these codes into a single platform is judged to be impractical. In this context, BDAs are the method of choice by engineers because of the high degree of flexibility these methods allow in the solution process. In particular, when using BDAs, the different groups working on the disciplinary analysis are free to choose the particular optimization algorithms they use to solve the disciplinary subproblems. This allows engineers to reuse their software codes with only minor modifications. But there is an important downside to BDAs; namely, there is analytical and numerical evidence that certain commonly used BDAs may fail to converge even when the starting point is very close to the minimizer (Alexandrov and Lewis [2], DeMiguel and Murray [13]). Finally, it is important to note that not all MDO problems can be formulated as PSPs. For instance, some MDO problems may involve integer or categorical variables or for certain problems, it may not be possible to synthesize the objectives of all disciplines into a single function. In this paper, we focus on the PSP formulation because it remains analytically tractable while covering a significant subclass of MDO problems.

The second type of decomposition methods for the nonconvex PSP that we consider is Schur interior-point methods (SIPMs). These methods apply an interior-point technique to solve the PSP in its original (integrated) form, but then use a Schur complement approach to solve the Newton system (and thus compute the search direction) in a decentralized manner. Thus SIPMs allow for a parallel solution of the linear algebra involved in interior-point methods. The main advantages of SIPMs over BDAs is that they have lower cost per iteration and can be shown to converge superlinearly or quadratically to a solution of the PSP. One disadvantage of SIPMs is that they can only be applied to problems that satisfy the so-called *strong linear independence constraint qualification* (SLICQ), whereas certain BDAs can deal with problems that do not satisfy SLICQ (Braun [5], DeMiguel and Murray [14]). Roughly speaking, the SLICQ implies that the constraints of the PSP can be eliminated and that, in the vicinity of the minimizer, the problem is an unconstrained one (DeMiguel and Murray [14]). This assumption is not likely to hold for certain applications (such as MDO) where finding a feasible solution is not straightforward. Another disadvantage is that the degree of flexibility provided by these approaches is not as high as that of BDAs. In particular, when using an SIPM, one must use the same optimization algorithm for all disciplinary analysis (in fact, to solve the whole problem) and there is flexibility only at the linear algebra level.

Summarizing, BDAs allow for a high degree of flexibility, which makes them suitable for certain applied problems such as MDO, while SIPMs are efficient and backed up by robust convergence theory for the case where SLICQ holds.

Our contribution is twofold. Our first contribution is to establish a theoretical relationship between BDAs and SIPMs. To do so, we consider a particular *inexact* BDA, which only takes one iteration to solve the subproblems. We show that the search direction of this inexact BDA can be approximated as the solution to a linear system that is closely related to the linear Newton system corresponding to SIPMs. Using this relationship, we analyze the convergence properties of the inexact BDA. Because this inexact BDA is a close relative of the BDAs used by engineers in practice, our analysis gives some insight into why some of the practical BDAs encounter convergence difficulties. In this sense, we think our analysis is a first step toward closing the gap between the incipient local convergence theory of BDAs (Alexandrov and Lewis [2], DeMiguel and Murray [14]) and the mature local convergence theory of interior-point methods (Martinez et al. [24], El-Bakry et al. [16], Yamashita and Yabe [36], Gould et al. [21]).

Our second contribution is to show how SIPMs can be applied to solve PSPs that satisfy only the conventional linear independence constraint qualification LICQ but not the strong LICQ. We accomplish this task in two steps. First, we show that any PSP satisfying the conventional LICQ can be *regularized* by introducing an exact penalty function. Second, we show how degenerate optimization techniques can be used to ensure an SIPM will achieve fast local convergence when applied to solve the regularized problem. The importance of this contribution is that it substantially enlarges the class of problems that can be addressed with SIPMs.

The paper is organized as follows. In §2, we introduce some notation and assumptions. Section 3 describes the two benchmark approaches to the nonconvex PSP: BDAs and SIPMs. In §§4 and 5, we introduce an inexact BDA and analyze its relationship to SIPMs. In §6, we show how a PSP satisfying LICQ can be *regularized* by introducing an exact penalty function. In §7, we show how the SIPM can be modified to ensure it achieves fast local convergence when applied to solve the regularized problem. Section 8 concludes.

2. Notation and assumptions. To facilitate the exposition and without loss of generality, herein we consider the following simplified problem composed of only one subsystem ($K = 1$):

$$\begin{aligned} & \underset{x, y, r}{\text{minimize}} && F(x, y), \\ & \text{subject to} && c(x, y) - r = 0, \\ & && r \geq 0, \end{aligned} \tag{2}$$

where $x \in \mathcal{R}^{n_x}$ are the *global* variables, $y \in \mathcal{R}^{n_y}$ are the *local* variables, $r \in \mathcal{R}^m$ are the *slack* variables, and $c(x, y): \mathcal{R}^{n_x+n_y} \rightarrow \mathcal{R}^m$ and $F(x, y): \mathcal{R}^{n_x+n_y} \rightarrow \mathcal{R}$ are smooth functions. Note that, in addition to considering only one subsystem, we have introduced slack variables, so that only equality constraints and nonnegativity bounds are present.

We will use the following notation: $\text{diag}(x)$ is the diagonal matrix whose diagonal contains the vector x , e_n is the n -dimensional vector of ones, and $\nabla_x c(x, y) \in \mathcal{R}^{m \times n_x}$ and $\nabla_y c(x, y) \in \mathcal{R}^{m \times n_y}$ are the matrices whose j th rows contain the gradients of the j th component of $c(x, y)$, with respect to x and y , respectively. Finally, we denote by $O(\alpha_k)$ a sequence $\{a_k\}$ satisfying $\|a_k\| \leq \gamma \alpha_k$ for some constant $\gamma > 0$, and we denote by $o(\alpha_k)$ a sequence $\{a_k\}$ satisfying $\|a_k\| \leq \gamma_k \alpha_k$ for some sequence $\{\gamma_k\}$ such that $\gamma_k > 0$ and $\lim_{k \rightarrow \infty} \gamma_k = 0$.

We assume there exists a minimizer (x^*, y^*, r^*) to problem (2) and a Lagrange multiplier vector (λ^*, σ^*) such that the vector

$$w^* = (x^*, y^*, r^*, \lambda^*, \sigma^*)$$

is a Karush-Kuhn-Tucker (KKT) point; that is, it satisfies the KKT conditions

$$\nabla_{xy} \mathcal{L}(w^*) = 0 \tag{3}$$

$$-\sigma^* + \lambda^* = 0 \tag{4}$$

$$c(x^*, y^*) - r^* = 0 \tag{5}$$

$$R^* \sigma^* = 0 \tag{6}$$

$$r^*, \sigma^* \geq 0, \tag{7}$$

where $\lambda^* \in \mathcal{R}^m$ are the multipliers of the equality constraints; $\sigma^* \in \mathcal{R}^m$ are the multipliers of the nonnegativity bounds on r^* , $R^* = \text{diag}(r^*)$; and the Lagrangian function is $\mathcal{L}(w^*) = F(x^*, y^*) - (\lambda^*)^T (c(x^*, y^*) - r^*) - (\sigma^*)^T r^*$.

We make the following assumptions on the problem functions and on the KKT point $w^* = (x^*, y^*, r^*, \lambda^*, \sigma^*)$.

ASSUMPTION A.1. *The second derivatives of the functions in problem (2) are Lipschitz continuous in an open convex set containing w^* .*

ASSUMPTION A.2. *The LICQ is satisfied at w^* ; that is, the matrix*

$$J = \begin{pmatrix} \nabla_x c(x^*, y^*) & \nabla_y c(x^*, y^*) & -I \\ 0 & 0 & I_{\mathcal{I}} \end{pmatrix} \tag{8}$$

has full row rank, where \mathcal{I} is the active set $\{i: r_i^* = 0\}$ and $I_{\mathcal{I}}$ is the matrix formed by the rows of the identity corresponding to indices in \mathcal{I} .

ASSUMPTION A.3. *The strict complementary slackness condition is satisfied at w^* ; that is, $\sigma_i^* > 0$ for $i \in \mathcal{I}$.*

ASSUMPTION A.4. *The second-order sufficient conditions for optimality are satisfied at w^* ; that is, for all $d \neq 0$ satisfying $Jd = 0$, we have*

$$d^T \nabla^2 \mathcal{L}(w^*) d > 0, \tag{9}$$

where $\nabla^2 \mathcal{L}(w^*)$ is the Hessian of the Lagrangian function with respect to the primal variables x, y, r at point w^* .

Finally, the following condition is often assumed to ensure the Schur complement matrix is well defined near the solution.

CONDITION C.1. *The SLICQ holds at w^* ; that is, the matrix*

$$\begin{pmatrix} \nabla_y c(x^*, y^*) & -I \\ 0 & I_{\mathcal{I}} \end{pmatrix} \tag{10}$$

has full row rank.

Clearly, the SLICQ (Condition C.1) implies the LICQ (Assumption A.2). Moreover, by the implicit function theorem, it is easy to see that the SLICQ implies that the problem constraints can be used in the vicinity of the minimizer to eliminate the local and slack variables y and r and transform the problem into an unconstrained problem on the global variables only. To illustrate this point, we consider the following example:

$$\begin{aligned} \min_{x,y} \quad & \frac{1}{2}(x-a)^2 + \frac{1}{2}(y-b)^2 \\ \text{s.t.} \quad & x+y \leq 2, \\ & x-y \leq 0. \end{aligned} \tag{11}$$

Note that the LICQ holds for this problem at all feasible points. In addition, the parameters (a, b) are useful to control whether the SLICQ holds at the minimizer. In particular, for $(a, b) = (1, 0)$, the minimizer is $(x^*, y^*) = (0.5, 0.5)$. At this point, only one of the constraints is active and the SLICQ holds. Moreover, note that for $(a, b) = (1, 0)$, one can perturb the value of the global variable x around the minimizer and there are values of the local variable y for which the problem is feasible. That is, one may use the active constraint to eliminate the local variable from the problem in the vicinity of the minimizer, and thus obtain an equivalent unconstrained problem on the global variables only. For $(a, b) = (2, 1)$, on the other hand, the minimizer is $(x^*, y^*) = (1, 1)$. At this point, there are two active constraints and the SLICQ does not hold. Also, if one increases the value of the global variable x above its optimal value of 1, there is no value of the local variables that makes the problem feasible. That is, for this case, we cannot use the constraints to eliminate the local variables from the problem. The example is illustrated in Figure 1.

3. Two benchmark approaches. We discuss two benchmark approaches to the nonconvex PSP: BDAs and SIPMs. Because most of our analysis will be based on interior-point techniques, it is convenient to introduce the *barrier PSP*

$$\begin{aligned} \text{minimize}_{x,y,r} \quad & F(x, y) - \mu \sum_{i=1}^m \log(r_i) \\ \text{subject to} \quad & c(x, y) - r = 0, \end{aligned} \tag{12}$$

where μ is the barrier parameter.

3.1. BDAs. The structure of the barrier PSP suggests it can be decomposed into a master problem, which only depends on the global variables, and a subproblem, which depends on the local and slack variables. In particular, problem (12) can be reformulated as the following *master problem*:

$$\min_x G^*(x), \tag{13}$$

where $G^*(x) \equiv F(x, y^*(x)) - \mu \sum_{i=1}^m \log(r_i^*(x))$ is the subproblem optimal value function

$$\begin{aligned} G^*(x) = \text{minimum}_{y,r} \quad & F(x, y) - \mu \sum_{i=1}^m \log(r_i) \\ \text{subject to} \quad & c(x, y) - r = 0, \end{aligned} \tag{14}$$

and we have omitted the dependence of $G^*(x)$ on μ to simplify notation.

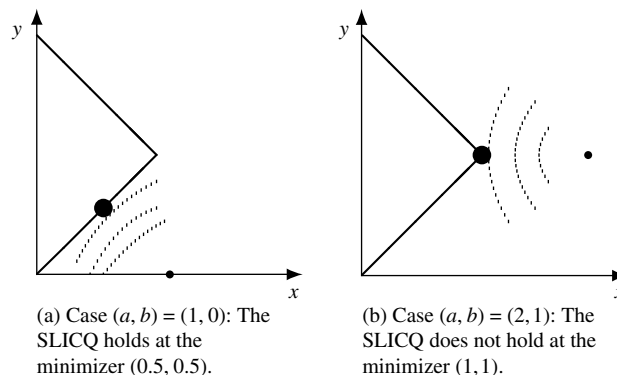


FIGURE 1. Example illustrating SLICQ.

- Initialization:** Choose a starting point x_0 . Set $k \leftarrow 0$ and choose the parameter $\mu_0 > 0$.
- Repeat**
- (i) **Solve subproblem:** Solve the subproblem (14) with μ_k and evaluate $G^*(x_k)$, $\nabla G^*(x_k)$ and $\nabla^2 G^*(x_k)$.
 - (ii) **Master problem search direction:** Compute the search direction for the master problem (13) by solving the system $\nabla^2 G^*(x_k)\Delta x = -\nabla G^*(x_k)$.
 - (iii) **Line search:** Compute a diagonal matrix, Λ_k , of step sizes.
 - (iv) **Update iterate:** Set $x_{k+1} = x_k + \alpha_k \Delta x$.
 - (v) **Parameter update:** Set $\mu_{k+1} > 0$ and $k \leftarrow k + 1$.
- Until convergence**

FIGURE 2. BDA.

BDA's divide the job of finding a minimizer to the PSP into two different tasks: (i) finding an optimal value of the global variables x^* and (ii) finding an optimal value of the local variables $y^*(x)$ for a given value of the global variables x . The first task is accomplished by solving the master problem (13) and the second by solving the subproblem (14).

BDA's apply an iterative unconstrained nonlinear optimization method to solve the master problem. At each iteration, a new estimate of the global variables x_k is generated and the subproblem is solved *exactly* using x_k as a parameter. Then, sensitivity analysis formulae (Fiacco and McCormick [17]) are used to compute the master problem objective and its derivatives from the *exact* subproblem minimizer. 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. The general BDA algorithm is stated in Figure 2.

BDA's are very popular among engineers solving the MDO. The main advantage of BDA's in this context is that, in the general case where there are more than one subsystem or discipline ($K > 1$), the above formulation allows the different disciplines to be dealt with almost independently. Moreover, within the framework given in Figure 2, the engineer is free to choose the particular optimization algorithm used to solve each of the disciplinary subproblems and the sensitivity formulae used to compute the first and second derivative of $G^*(x)$. Again, this usually allows engineers to reuse their disciplinary software codes with minimum change. Another advantage of BDA's is that there exist BDA's that can deal with problems whose minimizer does *not* satisfy the SLICQ (Braun [5], DeMiguel and Murray [14]).

The downside of BDA's is that there is analytical and numerical evidence that certain commonly used BDA's may fail to converge even when the starting point is very close to the minimizer (Alexandrov and Lewis [2], DeMiguel and Murray [13]). Moreover, although there are some local convergence proofs for certain BDA's that solve the subproblems exactly (DeMiguel and Murray [14]), it is safe to say that the local convergence theory of BDA's is not nearly as satisfactory as that of standard interior-point methods.

3.2. SIPM. In this section, we describe how a primal-dual interior-point method (Byrd et al. [7], El-Bakry et al. [16], Forsgren and Gill [18], Gay et al. [19], Vanderbei and Shanno [34]) can be combined with a Schur complement approach (Cottle [11]) to solve problem (1) in a decentralized manner. We term the resulting approach the SIPM.

3.2.1. The interior-point method. Primal-dual interior-point methods apply Newton's method to solve a perturbed version of the KKT conditions for the PSP (3)–(7). At each iteration, a search direction is computed by solving a linearization of these perturbed KKT conditions. Then, a step size is chosen such that all nonnegative variables remain strictly positive.

Given a suitable constraint qualification condition, a minimizer to the barrier PSP (12) must satisfy the following KKT conditions:

$$g(\mu) \equiv \begin{pmatrix} \nabla_x F(x, y) - \nabla_x c(x, y)^T \lambda \\ \nabla_y F(x, y) - \nabla_y c(x, y)^T \lambda \\ -\sigma + \lambda \\ -c(x, y) + r \\ -R\sigma + \mu e_m \end{pmatrix} = 0, \quad (15)$$

where $R = \text{diag}(r)$; $\lambda \in \mathcal{R}^m$ are the multipliers of the equality constraints; $\sigma \in \mathcal{R}^m$ are the multipliers of the nonnegativity bounds on r ; $e_m \in \mathcal{R}^m$ is the vector whose components are all ones; and the variables r, λ, σ

are strictly positive. Note that a minimizer to the PSP (2) must satisfy $g(0) = 0$. This is why people refer to Conditions (15) as the perturbed KKT conditions.

Notice that the problem variables can be split into two different components: the global component x and the local component $\hat{y} = (y, r, \lambda, \sigma)$. Likewise, $g(\mu)$ can also be split into two different components. Note that, to simplify notation, we omit the dependence of g on the variables and multipliers.

$$g(\mu) = \begin{pmatrix} g_1 \\ g_2(\mu) \end{pmatrix} = 0, \tag{16}$$

where

$$g_1 = \nabla_x F(x, y) - \nabla_x c(x, y)^T \lambda \tag{17}$$

and

$$g_2(\mu) = \begin{pmatrix} \nabla_y F(x, y) - \nabla_y c(x, y)^T \lambda \\ -\sigma + \lambda \\ -c(x, y) + r \\ -R\sigma + \mu e_m \end{pmatrix}. \tag{18}$$

Let $w_k = (x_k, \hat{y}_k)$ be the current estimate of the global and local components. Then, the *Newton search direction*, $\Delta w_k^N = (\Delta x_k^N, \Delta \hat{y}_k^N)$, is the solution to the following system of linear equations:

$$\begin{pmatrix} W_k & -\hat{A}_k^T \\ -\hat{A}_k & M_k \end{pmatrix} \begin{pmatrix} \Delta x_k^N \\ \Delta \hat{y}_k^N \end{pmatrix} = - \begin{pmatrix} g_{1,k} \\ g_{2,k}(\mu_k) \end{pmatrix}, \tag{19}$$

where $g_{1,k}$ and $g_{2,k}$ denote the functions g_1 and g_2 evaluated at w_k , $W_k = \nabla_x g_{1,k}$, $\hat{A}_k = -\nabla_x g_{2,k}(\mu_k) = -(\nabla_{\hat{y}} g_{1,k})^T$ and

$$M_k = \nabla_{\hat{y}} g_{2,k}(\mu_k). \tag{20}$$

For convenience, we rewrite the Newton system (19) as

$$K_k^N \Delta w_k^N = -g_k(\mu_k). \tag{21}$$

In addition to computing the Newton step, interior-point methods choose a step size such that all nonnegative variables remain strictly positive. In our case, r , λ , and σ must remain positive. To ensure this, we assume that the step sizes are chosen such as those in Yamashita and Yabe [36]. Therefore, at iteration k ,

$$\alpha_{r,k} = \min \left\{ 1, \gamma_k \min \left\{ -\frac{r_{k,i}}{\Delta r_{k,i}^N} \right\} \text{ s.t. } \Delta r_{k,i}^N < 0 \right\}, \tag{22}$$

$$\alpha_{\lambda,k} = \min \left\{ 1, \gamma_k \min \left\{ -\frac{\lambda_{k,i}}{\Delta \lambda_{k,i}^N} \right\} \text{ s.t. } \Delta \lambda_{k,i}^N < 0 \right\}, \tag{23}$$

$$\alpha_{\sigma,k} = \min \left\{ 1, \gamma_k \min \left\{ -\frac{\sigma_{k,i}}{\Delta \sigma_{k,i}^N} \right\} \text{ s.t. } \Delta \sigma_{k,i}^N < 0 \right\}, \tag{24}$$

where $\gamma_k \in (0, 1)$. As the global and local variables are not required to be nonnegative, we can set

$$\alpha_{x,k} = \alpha_{y,k} = 1. \tag{25}$$

If we define the matrix Λ_k as

$$\Lambda_k = \begin{pmatrix} \alpha_{x,k} I & 0 & 0 & 0 & 0 \\ 0 & \alpha_{y,k} I & 0 & 0 & 0 \\ 0 & 0 & \alpha_{r,k} I & 0 & 0 \\ 0 & 0 & 0 & \alpha_{\lambda,k} I & 0 \\ 0 & 0 & 0 & 0 & \alpha_{\sigma,k} I \end{pmatrix}, \tag{26}$$

the k th iteration of a primal-dual algorithm has the following form:

$$w_{k+1} = w_k + \Lambda_k \Delta w_k^N. \tag{27}$$

We also define the matrix $\Lambda_{\hat{y},k}$, which contains the step lengths on the \hat{y} variables. That is, the matrix $\Lambda_{\hat{y},k} = \text{diag}(\alpha_{y,k} I, \alpha_{r,k} I, \alpha_{\lambda,k} I, \alpha_{\sigma,k} I)$.

3.2.2. The Schur complement approach. Finally, assuming the matrix M_k is invertible, one can use the Schur complement approach to solve the Newton system (19) in a distributed manner. In particular, the Schur complement of matrix K_k^N is the matrix

$$S_k = W_k - \hat{A}_k^T M_k^{-1} \hat{A}_k. \quad (28)$$

If in addition S_k is invertible, the global component of the Newton search direction Δx_k^N can be computed as

$$S_k \Delta x_k^N = -(g_{1,k} + \hat{A}_k^T M_k^{-1} g_{2,k}(\mu_k)). \quad (29)$$

Then, the local component $\Delta \hat{y}_k^N$ is

$$M_k \Delta \hat{y}_k^N = -(g_{2,k} - \hat{A}_k \Delta x_k^N). \quad (30)$$

The following proposition gives conditions under which the matrices M_k and S_k are invertible in the vicinity of the minimizer to problem (2).

PROPOSITION 3.1. *Under Assumptions A.1–A.4 and Condition C.1, $\|M_k^{-1}\|$ and $\|S_k^{-1}\|$ are bounded for w_k in a neighborhood of the minimizer w^* .*

PROOF. Let $(K^N)^*$, M^* , and S^* be the matrices K_k^N , M_k , and S_k evaluated at w^* . Assumptions A.1–A.4 imply by Fiacco and McCormick [17, Theorem 14] and El-Bakry et al. [16, Proposition 4.1] that $(K^N)^*$ must be invertible.

In addition, it is easy to see that if $w^* = (x^*, y^*, r^*, \lambda^*, \sigma^*)$ is a KKT point satisfying Assumptions A.1–A.4 and Condition C.1, then $(y^*, r^*, \lambda^*, \sigma^*)$ is a KKT point, satisfying the LICQ (by C.1) and the strict complementarity slackness (SCS) condition (by A.3) for subproblem (14) with $x = x^*$ and $\mu = 0$:

$$\begin{aligned} & \underset{y, r}{\text{minimize}} && F(x^*, y) \\ & \text{subject to} && c(x^*, y) - r = 0, \\ & && r \geq 0. \end{aligned} \quad (31)$$

Moreover, the second-order sufficient conditions (SOSCs) for subproblem (31) are also satisfied. To see this, note that for any nonzero $\bar{d} \in \mathcal{R}^{n_y+m}$ such that $\bar{J}\bar{d} = 0$, where

$$\bar{J} = \begin{pmatrix} \nabla_y c(x^*, y^*) & -I \\ 0 & I_{\mathcal{I}} \end{pmatrix} \quad (32)$$

is the Jacobian matrix for subproblem (31), we have that

$$J \begin{pmatrix} 0 \\ \bar{d} \end{pmatrix} = 0,$$

where J is the Jacobian matrix for problem (2). That is, the null space of the Jacobian of the active constraints for the subproblem (31) is contained in the null space of the Jacobian of the active constraints for problem (2). Therefore, from A.4, we then have that

$$0 < (0 \quad \bar{d}^T) \nabla^2 \mathcal{L}(w^*) \begin{pmatrix} 0 \\ \bar{d} \end{pmatrix} = \bar{d}^T \nabla^2 \bar{\mathcal{L}}(\hat{y}^*) \bar{d}, \quad (33)$$

where $\nabla^2 \bar{\mathcal{L}}(\hat{y}^*)$ is the Hessian of the Lagrangian function of subproblem (31), with respect to (y, r) at point $\hat{y}^* = (y^*, r^*, \lambda^*, \sigma^*)$. Hence the SOSCs for subproblem (31) are satisfied at \hat{y}^* .

The LICQ, the SCS and the SOSC for subproblem (31) imply the matrix M^* must be invertible by Fiacco and McCormick [17, Theorem 14] and El-Bakry et al. [16, Proposition 4.1].

Moreover, the invertibility of $(K^N)^*$ and M^* implies that S^* must be invertible or otherwise there would be multiple solutions to the Newton system (19), in contradiction with the invertibility of $(K^N)^*$. Finally, $\|M_k^{-1}\|$ and $\|S_k^{-1}\|$ are bounded in the vicinity of the minimizer by the invertibility of M_k^* and S_k^* and Assumption A.1 \square

Note that to ensure the Schur complement matrix is well defined in the vicinity of a minimizer, we need to assume the SLICQ (Condition C.1). As we mentioned in §2, the SLICQ implies that the PSP constraints can be used in the vicinity of the minimizer to eliminate the local and slack variables y and r , and thus transform the PSP into an unconstrained problem on only the global variables. This is a fairly strong assumption for application areas such as MDO, where often the difficulty is precisely finding a design that is feasible with respect to the constraints of all disciplines. In §7, we show how the SIPM can be modified to deal with problems that do not satisfy the SLICQ.

The resulting SIPM is stated in Figure 3. Also, note that, for the general problem (1) with K subsystems, M_k is a block diagonal matrix composed of K blocks. Thus the Schur complement allows one to decompose the linear system (30) into K smaller independent linear systems (see Birge and Louveaux [4, §5.6]).

Initialization: Choose the starting point $w_0 = (x_0, y_0, r_0, \lambda_0, \sigma_0)$ such that $r_0 > 0$, $\lambda_0 > 0$, $\sigma_0 > 0$. Set $k \leftarrow 0$ and choose the parameters $\mu_0 > 0$ and $0 < \hat{\gamma} \leq \gamma_0 < 1$.

Repeat

- (i) **Master problem iteration:** Form the matrix S_k and compute Δx_k^N from system (29). Set $x_{k+1} = x_k + \Delta x_k^N$.
- (ii) **Subproblem iteration:**
 - (a) **Search direction:** Compute $\Delta \hat{y}_k^N$ by solving system (30).
 - (b) **Line search:** Compute the diagonal matrix, Λ_k , from the subproblem step sizes as in (22)–(26).
 - (c) **Update iterate:** Set $\hat{y}_{k+1} = \hat{y}_k + \Lambda_{\hat{y},k} \Delta \hat{y}_k^N$.
- (iii) **Parameter update:** Set $\mu_{k+1} > 0$, $0 < \hat{\gamma} \leq \gamma_{k+1} < 1$ and $k \leftarrow k + 1$.

Until convergence

FIGURE 3. SIPM.

3.2.3. Convergence. The local convergence theory of interior-point methods is developed in the papers by Martinez et al. [24], El-Bakry et al. [16], Yamashita and Yabe [36], and Gould et al. [21]. These papers show that under Assumptions A.1–A.4 and certain conditions on parameters μ_k and γ_k , the Newton matrix K_k^N is invertible in the vicinity of the minimizer and the iteration (27) converges superlinearly or quadratically to a solution of (2). If in addition, SLICQ holds, we know by Proposition 3.1 that Equations (29) and (30) have a unique solution in the vicinity of the minimizer, and thus the SIPM converges also superlinearly or quadratically. As in this paper, our analysis focuses on the local convergence properties of the algorithms, no procedures are given to ensure global convergence, though the techniques in Vanderbei and Shanno [34], Gay et al. [19], Forsgren and Gill [18], and Byrd et al. [7] could be adapted.

4. An inexact BDA. In this section, we describe the inexact BDA. We also analyze its relationship to the SIPM described in the previous section.

4.1. The algorithm. The inexact BDA takes just one Newton iteration of a primal-dual interior-point method to solve the subproblems (14). Following the notation introduced in §3, the subproblem perturbed KKT conditions can be written in compact form as $g_2(\mu) = 0$ (see (18)). Then, the Newton system for the subproblem perturbed KKT conditions is simply

$$M_k \Delta \hat{y}_k^D = -g_{2,k}(\mu_k). \quad (34)$$

The Newton search direction for the master problem (13) is the solution to the Newton system

$$\nabla_{xx}^2 G^*(x_k) \Delta x_k^D = -\nabla_x G^*(x_k), \quad (35)$$

where, to simplify notation, we do not explicitly write the dependence of G^* on the barrier parameter μ . Unfortunately, because the inexact BDA does not solve the subproblems exactly, the master problem Hessian and gradient $\nabla_{xx}^2 G^*(x_k)$ and $\nabla_x G^*(x_k)$ cannot be computed from standard sensitivity formulae, as is customary in BDAs. In the remainder of this section, we show how approximations to $\nabla_x G^*(x_k)$ and $\nabla_{xx}^2 G^*(x_k)$ can be obtained from the estimate of the subproblem minimizer. The inexact BDA uses these approximations to generate the search direction for the master problem. In particular, the following two propositions show that the right-hand side in Equation (29) can be seen as an approximation to the master problem gradient $\nabla_x G^*(x_k)$ and that the Schur complement matrix S_k can be interpreted as an approximation to the master problem Hessian $\nabla_{xx}^2 G^*(x_k)$.

PROPOSITION 4.1. *Let $(x^*, y^*, r^*, \lambda^*, \sigma^*)$ be a KKT point satisfying Assumptions A.1–A.4 and Condition C.1 for problem (2). Then, for $(x_k, y_k, r_k, \lambda_k, \sigma_k)$ close to $(x^*, y^*, r^*, \lambda^*, \sigma^*)$, the subproblem optimal value function $G^*(x_k)$ and its gradient $\nabla_x G^*(x_k)$ are well defined and*

$$\|\nabla_x G^*(x_k) - (g_{1,k} + \hat{A}_k^T M_k^{-1} g_{2,k}(\mu_k))\| = o(\|\hat{y}(x_k) - \hat{y}_k\|),$$

where $\hat{y}(x_k) = (y(x_k), r(x_k), \lambda(x_k), \sigma(x_k))$ is the locally unique, once continuously differentiable trajectory of minimizers to subproblem (14) with $\hat{y}(x^*) = (y^*, r^*, \lambda^*, \sigma^*)$ and $\hat{y}_k = (y_k, r_k, \lambda_k, \sigma_k)$.

PROOF. Note that if Condition C.1 holds at $(x^*, y^*, r^*, \lambda^*, \sigma^*)$, then the LICQ holds at $(y^*, r^*, \lambda^*, \sigma^*)$ for subproblem (14) with $x = x^*$ and $\mu = 0$. Moreover, it is easy to see that if $(x^*, y^*, r^*, \lambda^*, \sigma^*)$ is a KKT point satisfying Assumptions A.1–A.4, then $(y^*, r^*, \lambda^*, \sigma^*)$ is a minimizer satisfying the SCS and SOSCs for subproblem (14) with $x = x^*$ and $\mu = 0$. It follows from Fiacco and McCormick [17, Theorem 6] that there exists a locally unique, once continuously differentiable trajectory of subproblem minimizers $\hat{y}(x_k) = (y(x_k), r(x_k), \lambda(x_k), \sigma(x_k))$, satisfying LICQ, SCS, and SOSC for the subproblem with x_k close to x^* and

μ_k sufficiently small. As a result, the subproblem optimal value function $G^*(x_k)$ can be defined as $G^*(x_k) = F(x_k, y(x_k)) - \mu \sum_{i=1}^m \log(r_i(x_k))$ and it is once continuously differentiable. By the properties of $\hat{y}(x_k)$, its gradient is simply

$$\nabla_x G^*(x_k) = \frac{d[F(x_k, y(x_k)) - \mu \sum_{i=1}^m \log(r_i(x_k))]}{dx} = \frac{d\mathcal{L}_y(x_k, \hat{y}(x_k))}{dx}, \quad (36)$$

where d/dx denotes the total derivative and \mathcal{L}_y is the subproblem Lagrangian function

$$\mathcal{L}_y(x, \hat{y}(x)) = F(x, y) - \mu \sum_{i=1}^m \log(r_i) - \lambda^T (c(x, y) - r). \quad (37)$$

Applying the chain rule, we get

$$\frac{d\mathcal{L}_y(x_k, \hat{y}(x_k))}{dx} = \nabla_x \mathcal{L}_y(x_k, \hat{y}(x_k)) \quad (38)$$

$$+ \nabla_y \mathcal{L}_y(x_k, \hat{y}(x_k)) y'(x_k) \quad (39)$$

$$+ \nabla_r \mathcal{L}_y(x_k, \hat{y}(x_k)) r'(x_k) \quad (40)$$

$$+ \nabla_\lambda \mathcal{L}_y(x_k, \hat{y}(x_k)) \lambda'(x_k) \quad (41)$$

$$+ \nabla_\sigma \mathcal{L}_y(x_k, \hat{y}(x_k)) \sigma'(x_k), \quad (42)$$

where $y'(x_k)$, $r'(x_k)$, $\lambda'(x_k)$, and $\sigma'(x_k)$ denote the Jacobian matrices of y , r , λ , and σ evaluated at x_k , respectively. Note that (39) and (40) are zero because of the optimality of $\hat{y}(x_k)$, (41) is zero by the feasibility of $\hat{y}(x_k)$, and (42) is zero because the Lagrangian function does not depend on σ . Thus we can write the master problem objective gradient as

$$\nabla_x G^*(x_k) = \nabla_x \mathcal{L}_y(x_k, \hat{y}(x_k)). \quad (43)$$

If we knew the subproblem minimizer $\hat{y}(x_k)$, we could easily compute the master problem gradient by evaluating the gradient of the Lagrangian function (37) at x_k and $\hat{y}(x_k)$. After taking only one interior-point iteration on the subproblem, we do not know $\hat{y}(x_k)$ exactly but rather the following approximation:

$$\hat{y}_k + \Delta \hat{y}_k^D, \quad (44)$$

where $\Delta \hat{y}_k^D$ is the subproblem search direction computed by solving system (34). However, by Taylor's theorem, we know that the master problem gradient can be approximated as

$$\nabla_x G^*(x_k) = \nabla_x \mathcal{L}_y(x_k, \hat{y}_k) + \nabla_{x\hat{y}} \mathcal{L}_y(x_k, \hat{y}_k) (\hat{y}(x_k) - \hat{y}_k) + O(\|\hat{y}(x_k) - \hat{y}_k\|^2). \quad (45)$$

Moreover, if \hat{y}_k is close enough to $\hat{y}(x_k)$, we know from the local convergence theory of Newton's method that $\|\hat{y}(x_k) - (\hat{y}_k + \Delta \hat{y}_k^D)\| = o(\|\hat{y}(x_k) - \hat{y}_k\|)$, and thus

$$\nabla_x G^*(x_k) = \nabla_x \mathcal{L}_y(x_k, \hat{y}_k) + \nabla_{x\hat{y}} \mathcal{L}_y(x_k, \hat{y}_k) \Delta \hat{y}_k^D + o(\|\hat{y}(x_k) - \hat{y}_k\|). \quad (46)$$

Finally, from Proposition 3.1, we know that the matrix M_k is nonsingular once the iterates are close to the minimizer (Fiacco and McCormick [17, Theorem 14]). Since $\Delta \hat{y}_k^D = -M_k^{-1} g_{2,k}(\mu_k)$ and $\hat{A}_k^T = -\nabla_{\hat{y}} g_{1,k} = -\nabla_{x\hat{y}} \mathcal{L}_y(x_k, \hat{y}_k)$, the result follows from (46). \square

PROPOSITION 4.2. *Let $(x^*, y^*, r^*, \lambda^*, \sigma^*)$ be a KKT point satisfying Assumptions A.2–A.4 and Condition C.1 for problem (2). Moreover, assume all functions in problem (2) are three times continuously differentiable. Then, for $(x_k, y_k, r_k, \lambda_k, \sigma_k)$ close to $(x^*, y^*, r^*, \lambda^*, \sigma^*)$, the Hessian of the subproblem optimal value function $\nabla_{xx}^2 G^*(x_k)$ is well defined and*

$$\|\nabla_{xx}^2 G^*(x_k) - S_k\| = O(\|\hat{y}(x_k) - \hat{y}_k\|),$$

where $\hat{y}(x_k) = (y(x_k), r(x_k), \lambda(x_k), \sigma(x_k))$ is the locally unique, twice continuously differentiable trajectory of minimizers to subproblem (14) with $\hat{y}(x^*) = (y^*, r^*, \lambda^*, \sigma^*)$, $\hat{y}_k = (y_k, r_k, \lambda_k, \sigma_k)$, and S_k is the Schur complement matrix $S_k = W_k - \hat{A}_k^T M_k^{-1} \hat{A}_k$ of the system matrix in (19).

PROOF. By the same arguments as in Proposition 4.1, and the assumption that all problem functions are three times continuously differentiable, we know that the subproblem optimal value function can be defined as

$$G^*(x_k) = F(x_k, y(x_k)) - \mu \sum_{i=1}^m \log(r_i(x_k))$$

and it is twice continuously differentiable.

Moreover, differentiating expression (43), we obtain the following expression for the optimal value function Hessian:

$$\begin{aligned} \nabla_{xx} G^*(x_k) &= \frac{d(\nabla_x \mathcal{L}_y(x_k, \hat{y}(x_k)))}{dx} \\ &= \nabla_{xx} \mathcal{L}_y(x_k, \hat{y}(x_k)) + \nabla_{x\hat{y}} \mathcal{L}_y(x_k, \hat{y}(x_k)) \hat{y}'(x_k), \end{aligned} \quad (47)$$

where $\hat{y}'(x_k)$ is the Jacobian matrix of the subproblem minimizer with respect to x_k .

By A.3, A.4, and C.1, we know that for x_k close enough to x^* and μ_k sufficiently small, $\hat{y}(x_k)$ is a minimizer satisfying the LICQ, SCS, and SOSC for the subproblem, and thus it follows from Fiacco and McCormick [17, Theorem 6] that

$$M_k^* \hat{y}'(x_k) = \hat{A}_k^*, \quad (48)$$

where M_k^* and \hat{A}_k^* are the matrices M_k and \hat{A}_k evaluated at $\hat{y}(x_k)$.

If we knew the subproblem minimizer $\hat{y}(x_k)$ exactly, we could use (47) and (48) to compute the master problem Hessian. Unfortunately, after taking only one Newton iteration on the subproblems, we do not know $\hat{y}(x_k)$ exactly. However, we can approximate $\hat{y}'(x_k)$ as the solution to the following system:

$$M_k \hat{y}'(x_k) \simeq \hat{A}_k. \quad (49)$$

Assumptions A.3, A.4, and Condition C.1 imply by Proposition 3.1 that $\|M_k^{-1}\|$ is uniformly bounded for (x_k, \hat{y}_k) in the vicinity of w^* . Then,

$$\|\hat{y}'(x_k) - M_k^{-1} \hat{A}_k\| \leq \|((M_k^*)^{-1} - M_k^{-1}) \hat{A}_k\| + \|(M_k^*)^{-1} (\hat{A}_k^* - \hat{A}_k)\|. \quad (50)$$

By the assumption that all functions are three times continuously differentiable, and the fact that $\|M_k^{-1}\|$ is uniformly bounded in the vicinity of w^* , the first term in (50) can be written as

$$\|((M_k^*)^{-1} - M_k^{-1}) \hat{A}_k\| \leq \|M_k^{-1}\| \|M_k - M_k^*\| \|(M_k^*)^{-1}\| \|\hat{A}_k\| = O(\|\hat{y}(x_k) - \hat{y}_k\|).$$

Likewise, by the assumption that all functions are three times continuously differentiable and because $\|M_k^{-1}\|$ is uniformly bounded, the second term in (50) is $O(\|\hat{y}(x_k) - \hat{y}_k\|)$. Therefore

$$\|\hat{y}'(x_k) - M_k^{-1} \hat{A}_k\| = O(\|\hat{y}(x_k) - \hat{y}_k\|).$$

Finally, the result follows because $W_k = \nabla_x g_{1,k} = \nabla_{xx} \mathcal{L}_y(x_k, \hat{y}_k)$ and $\hat{A}_k^T = -\nabla_{\hat{y}} g_{1,k} = -\nabla_{x\hat{y}} \mathcal{L}_y(x_k, \hat{y}_k)$. \square

Note that Propositions 4.1 and 4.2 show that the SIPM iteration

$$S_k \Delta x_k^D = -(g_{1,k} + \hat{A}_k^T M_k^{-1} g_{2,k}(\mu_k)) \quad (51)$$

described in §3.2.2, may be seen as an approximation to the master problem Newton Equation (35).

The inexact bilevel decomposition algorithm (IBDA) is stated in Figure 4.

4.2. Relationship to the SIPM. The following proposition establishes the relationship between the search direction of the inexact BDA ($\Delta x_k^D, \Delta y_k^D$) and the search direction of the SIPM ($\Delta x_k^N, \Delta y_k^N$). In particular, we show that the global variable components of both search directions are identical and the local components are equal up to first-order terms.

PROPOSITION 4.3. *Under Assumptions A.1–A.4 and Condition C.1, for w_k in a neighborhood of the KKT point w^* , we have that*

$$\Delta x_k^D = \Delta x_k^N \quad \text{and} \quad \Delta y_k^D = \Delta y_k^N + O(\|\Delta x_k^N\|). \quad (52)$$

Initialization: Choose a starting point $w_0 = (x_0, y_0, r_0, \lambda_0, \sigma_0)$ such that $r_0 > 0$, $\lambda_0 > 0$, $\sigma_0 > 0$. Set $k \leftarrow 0$ and choose the parameters $\mu_0 > 0$ and $0 < \hat{\gamma} \leq \gamma_0 < 1$.

Repeat

(i) **Solve master problem:** Form the matrix S_k and compute Δx_k^D from system (51). Set $x_{k+1} = x_k + \Delta x_k^D$.

(ii) **Solve subproblem:**

(a) **Search direction:** Compute $\Delta \hat{y}_k^D$ by solving system (34).

(b) **Line search:** Compute the diagonal matrix, Λ_k , from the subproblem step sizes as in (22)–(26).

(c) **Update iterate:** Set $\hat{y}_{k+1} = \hat{y}_k + \Lambda_{\hat{y},k} \Delta \hat{y}_k^D$.

(iii) **Parameter update:** Set $\mu_{k+1} > 0$, $0 < \hat{\gamma} \leq \gamma_{k+1} < 1$ and $k \leftarrow k + 1$.

Until convergence

FIGURE 4. IBDA.

PROOF. The first equality follows trivially from (29) and (51).

From Proposition 3.1, we know that M_k is invertible for w_k in a neighborhood of w^* . Thus by multiplying (34) by M_k^{-1} , we can write the local search direction of the inexact BDA as $\Delta \hat{y}_k^D = -M_k^{-1} g_{2,k}(\mu_k)$, and by multiplying (30) by M_k^{-1} , we can write the local search direction of the SIPM as $\Delta \hat{y}_k^N = -M_k^{-1}(g_{2,k}(\mu_k) - \hat{A}_k \Delta x_k^N)$. Hence

$$\Delta \hat{y}_k^D = \Delta \hat{y}_k^N - M_k^{-1} \hat{A}_k \Delta x_k^N.$$

The second equality in (52) follows from Proposition 3.1 and Assumption A.1. \square

Note that the difference between the local components of both search directions is not surprising because the global variables are just a parameter to the subproblem solved by the decomposition algorithm. As a result, the local component of the search direction computed by the inexact BDA lacks *first-order* information about the global component search direction. In §5, we show how a Gauss-Seidel strategy can be used to overcome this limitation inherent to BDAs.

Finally, it is useful to note that from (34) and (51), the inexact BDA search direction is the solution to the following linear system:

$$K_k^D \Delta w_k^D = -g_k(\mu_k), \quad (53)$$

where

$$K_k^D = \begin{pmatrix} S_k & -\hat{A}_k^T \\ 0 & M_k \end{pmatrix}. \quad (54)$$

Note that the fact that the global variables are a parameter to the subproblems is evident in the structure of K_k^D . In particular, notice that the lower left block in matrix K_k^D is zero instead of $-\hat{A}_k$, as in the interior-point method matrix K_k^N . In §5, we give conditions under which the norm of the matrix $\|(K_k^D)^{-1}\|$ is bounded in the neighborhood of the minimizer, and thus the iterates of the proposed decomposition algorithm are well defined.

5. The Gauss-Seidel BDA. The difference in the local component of the search directions computed by the inexact BDA and the SIPM precludes any possibility of superlinear convergence for the decomposition algorithm. In this section, we show how one can first compute the global variable component of the search direction, and then use it to update the subproblem derivative information before computing the local variable component. We show that the resulting *Gauss-Seidel BDA* generates a search direction that is equal (up to second-order terms) to the search direction of the SIPM. Moreover, we prove that the resulting BDA converges locally at a superlinear rate.

5.1. The algorithm. The inexact BDA defined in §4 does not make use of all the information available at each step. Note that, at each iteration of the decomposition algorithm, we first compute the master problem step as the solution to

$$S_k \Delta x_k^G = -(g_{1,k} + \hat{A}_k^T M_k^{-1} g_{2,k}(\mu_k)), \quad (55)$$

and update the global variables as $x_{k+1} = x_k + \Delta x_k^G$. At this point, one could use the new value of the global variables x_{k+1} to perform a nonlinear update of the subproblem derivative information, and thus generate a better subproblem step. In particular, after solving for the master problem search direction, we could compute (see (18))

$$g_{2,k}^+(\mu_k) = \begin{pmatrix} \nabla_y F(x_{k+1}, y_k) - \nabla_y c(x_{k+1}, y_k)^T \lambda_k \\ -\sigma_k + \lambda_k \\ -c(x_{k+1}, y_k) + r_k \\ -R_k \sigma_k + \mu_k e_m \end{pmatrix}. \quad (56)$$

- Initialization:** Choose a starting point $w_0 = (x_0, y_0, r_0, \lambda_0, \sigma_0)$ such that $r_0 > 0, \lambda_0 > 0, \sigma_0 > 0$. Set $k \leftarrow 0$ and choose the parameters $\mu_0 > 0$ and $0 < \hat{\gamma} \leq \gamma_0 < 1$.
- Repeat**
- (i) **Master problem iteration:** Form the matrix S_k and compute Δx_k^G from system (55). Set $x_{k+1} = x_k + \Delta x_k^G$.
 - (ii) **Subproblem iteration:**
 - (a) **Search direction:** Use x_{k+1} to update $g_{2,k}^+(\mu_k)$ and compute $\Delta \hat{y}_k^G$ by solving system (57).
 - (b) **Line search:** Compute the diagonal matrix, Λ_k , from the subproblem step sizes as in (22)–(26).
 - (c) **Update iterate:** Set $\hat{y}_{k+1} = \hat{y}_k + \Lambda_{\hat{y},k} \Delta \hat{y}_k^G$.
 - (iii) **Parameter update:** Set $\mu_{k+1} > 0, 0 < \hat{\gamma} \leq \gamma_{k+1} < 1$ and $k \leftarrow k + 1$.
- Until convergence**

FIGURE 5. Gauss-Seidel BDA.

Then, the subproblem search direction would be given as the solution to

$$M_k \Delta \hat{y}_k^G = -g_{2,k}^+(\mu_k). \tag{57}$$

The resulting algorithm is stated in Figure 5. It must be noted that the only difference between the Gauss-Seidel BDA stated in Figure 5 and the inexact BDA stated in Figure 4 is that in the Gauss-Seidel version, we introduce a nonlinear update into the derivative information of the subproblem $g_{2,k}(\mu_k)$ using the master problem step Δx_k^G . As a consequence, the refinement requires one more subproblem derivative evaluation per iteration. The advantage is that, as we show in the next section, the Gauss-Seidel refinement guarantees that the proposed algorithm converges at a superlinear rate.

5.2. Relationship to the SIPM. The following proposition shows that the search directions of the proposed Gauss-Seidel BDA and the SIPM are equal up to second-order terms.

PROPOSITION 5.1. *Under Assumptions A.1–A.4 and Condition C.1, for w_k in a neighborhood of the KKT point w^* , we have that*

$$\Delta x_k^G = \Delta x_k^N$$

and

$$\Delta \hat{y}_k^G = \Delta \hat{y}_k^N + O(\|\Delta x_k^N\|^2). \tag{58}$$

PROOF. The result for the global components is trivial from (55). For the local components, note that the search direction of the resulting Gauss-Seidel decomposition algorithm satisfies

$$\Delta x_k^G = \Delta x_k^D = \Delta x_k^N \tag{59}$$

and

$$\Delta \hat{y}_k^G = \Delta \hat{y}_k^D - M_k^{-1}(g_{2,k}^+(\mu_k) - g_{2,k}(\mu_k)). \tag{60}$$

Moreover, from (60), we know that

$$\begin{aligned} \Delta \hat{y}_k^G &= \Delta \hat{y}_k^N - M_k^{-1} \hat{A}_k \Delta x_k^N - M_k^{-1}(g_{2,k}^+(\mu_k) - g_{2,k}(\mu_k)) \\ &= \Delta \hat{y}_k^N - M_k^{-1}(g_{2,k}^+(\mu_k) - g_{2,k}(\mu_k) + \hat{A}_k \Delta x_k^N). \end{aligned}$$

The result is obtained by Taylor’s theorem and the fact that $\hat{A}_k^T = -\nabla_x g_{2,k}(\mu_k)$. \square

Proposition 5.1 intuitively implies that the Gauss-Seidel BDA converges locally at a superlinear rate. In §5.3, we formally show this is the case.

5.3. Convergence of the Gauss-Seidel BDA. In this section, we first show that the search direction of the Gauss-Seidel BDA is well defined in the proximity of the minimizer and then, we show that the iterates generated by the Gauss-Seidel approach converge to the minimizer at a superlinear rate.

Note that the search directions of the inexact and Gauss-Seidel BDA are related as follows:

$$\Delta w_k^G = \Delta w_k^D - G_k(g_{2,k}^+(\mu_k) - g_{2,k}(\mu_k)), \tag{61}$$

where

$$G_k = \begin{pmatrix} 0 \\ M_k^{-1} \end{pmatrix},$$

and $\Delta w_k^D = -(K_k^D)^{-1} g_k(\mu_k)$. Consequently, to show that the Gauss-Seidel search direction is bounded in a neighborhood of a minimizer w^* , it suffices to show that the norms of the matrices $(K_k^D)^{-1}$ and M_k^{-1} are bounded for w_k in a neighborhood of the minimizer w^* .

PROPOSITION 5.2. *Under Assumptions A.1–A.4 and Condition C.1, $\|(K_k^D)^{-1}\|$ and $\|M_k^{-1}\|$ are bounded for w_k in a neighborhood of the minimizer w^* .*

PROOF. Note that from (54),

$$(K_k^D)^{-1} = \begin{pmatrix} S_k^{-1} & S_k^{-1} \hat{A}_k^T M_k^{-1} \\ 0 & M_k^{-1} \end{pmatrix}. \quad (62)$$

Consequently, it is sufficient to prove that $\|S_k^{-1}\|$ and $\|M_k^{-1}\|$ are bounded in the vicinity of the minimizer. The result follows from Proposition 3.1. \square

We now give a result that provides sufficient conditions on the barrier and the step-size parameter updates to ensure superlinear convergence of the Gauss-Seidel BDA.

THEOREM 5.1. *Suppose that Assumptions A.1–A.4 and Condition C.1 hold, that the barrier parameter is chosen to satisfy $\mu_k = o(\|g_k(0)\|)$, and that the step-size parameter is chosen such that $1 - \gamma_k = o(1)$. If w_0 is close enough to w^* , then the sequence $\{w_k\}$ described in (61) is well defined and converges to w^* at a superlinear rate.*

PROOF. The sequence in (61) updates the new point as

$$\begin{aligned} w_{k+1} &= w_k + \Lambda_k \Delta w_k^G \\ &= w_k + \Lambda_k [\Delta w_k^D - G_k(g_{2,k}^+(\mu_k) - g_{2,k}(\mu_k))]. \end{aligned} \quad (63)$$

By (56), note that $g_{2,k}^+(\mu_k) - g_{2,k}(\mu_k) = g_{2,k}^+(0) - g_{2,k}(0)$, and because matrices $(K_k^D)^{-1}$ and M_k^{-1} are well defined by Proposition 5.2, we have

$$\begin{aligned} w_{k+1} &= w_k - \Lambda_k (K_k^D)^{-1} g_k(\mu_k) - \Lambda_k G_k(g_{2,k}^+(0) - g_{2,k}(0)) \\ &= w_k - \Lambda_k (K_k^D)^{-1} (g_k(0) + \bar{\mu}_k) - \Lambda_k G_k(g_{2,k}^+(0) - g_{2,k}(0)), \end{aligned} \quad (64)$$

where $\bar{\mu}_k = (0, 0, 0, 0, \mu_k e_m)$. Then,

$$\begin{aligned} w_{k+1} - w^* &= w_k - w^* - \Lambda_k (K_k^D)^{-1} g_k(\mu_k) - \Lambda_k G_k(g_{2,k}^+(0) - g_{2,k}(0)) \\ &= (I - \Lambda_k)(w_k - w^*) + \Lambda_k (K_k^D)^{-1} (K_k^D(w_k - w^*) - g_k(0) - \bar{\mu}_k) - \Lambda_k G_k(g_{2,k}^+(0) - g_{2,k}(0)), \end{aligned} \quad (65)$$

which may be rewritten as

$$\begin{aligned} w_{k+1} - w^* &= (I - \Lambda_k)(w_k - w^*) \\ &\quad - \Lambda_k (K_k^D)^{-1} \bar{\mu}_k \\ &\quad + \Lambda_k (K_k^D)^{-1} (K_k^N(w_k - w^*) - g_k(0)) \\ &\quad + \Lambda_k (K_k^D)^{-1} (K_k^D - K_k^N)(w_k - w^*) \\ &\quad - \Lambda_k G_k(g_{2,k}^+(0) - g_{2,k}(0)). \end{aligned} \quad (66)$$

By Yamashita and Yabe [36, Lemma 4], the first term in (66) satisfies

$$\|(I - \Lambda_k)(w_k - w^*)\| \leq ((1 - \gamma_k) + O(\|g_k(0)\|) + O(\mu_k)) \|w_k - w^*\|. \quad (67)$$

This inequality together with conditions $1 - \gamma_k = o(1)$ and $\mu_k = o(\|g_k(0)\|)$ imply that

$$\|(I - \Lambda_k)(w_k - w^*)\| = o(\|w_k - w^*\|). \quad (68)$$

The second term in (66) satisfies

$$\|\Lambda_k (K_k^D)^{-1} \bar{\mu}_k\| \leq \|\Lambda_k\| \|(K_k^D)^{-1}\| \|\bar{\mu}_k\| \leq \beta \|\bar{\mu}_k\|, \quad (69)$$

which by condition $\mu_k = o(\|g_k(0)\|)$, imply

$$\|\Lambda_k (K_k^D)^{-1} \bar{\mu}_k\| = o(\|w_k - w^*\|). \quad (70)$$

By Taylor's theorem, the third term in (66) satisfies

$$\|\Lambda_k(K_k^D)^{-1}(K_k^N(w_k - w^*) - g_k(0))\| \leq \|\Lambda_k\| \|(K_k^D)^{-1}\| \|(K_k^N(w_k - w^*) - g_k(0))\| = o(\|w_k - w^*\|). \quad (71)$$

Finally, because

$$K_k^D = K_k^N - \begin{bmatrix} W_k - S_k & 0 \\ -\hat{A}_k & 0 \end{bmatrix} \quad \text{and} \quad (K_k^D)^{-1} = \begin{pmatrix} S_k^{-1} & S_k^{-1}\hat{A}_k^T M_k^{-1} \\ 0 & M_k^{-1} \end{pmatrix}, \quad (72)$$

the fourth term in (66) is

$$\begin{aligned} \Lambda_k(K_k^D)^{-1}(K_k^D - K_k^N)(w_k - w^*) &= -\Lambda_k(K_k^D)^{-1} \begin{bmatrix} W_k - S_k & 0 \\ -\hat{A}_k & 0 \end{bmatrix} (w_k - w^*) \\ &= \Lambda_k \begin{bmatrix} 0 & 0 \\ M_k^{-1}\hat{A}_k & 0 \end{bmatrix} (w_k - w^*) \\ &= \Lambda_k \begin{bmatrix} 0 \\ M_k^{-1}\hat{A}_k(x_k - x^*) \end{bmatrix}. \end{aligned} \quad (73)$$

Then, adding the fourth and fifth terms in (66) and using (73), we get

$$\begin{aligned} \Lambda_k(K_k^D)^{-1}(K_k^D - K_k^N)(w_k - w^*) - \Lambda_k G_k(g_{2,k}^+(0) - g_{2,k}(0)) \\ = \Lambda_k \begin{bmatrix} 0 \\ M_k^{-1}[\hat{A}_k(x_k - x^*) - (g_{2,k}^+(0) - g_{2,k}(0))] \end{bmatrix}. \end{aligned} \quad (74)$$

If only the global variable component, x , of Equations (68), (70), (71), and (74) is considered, then the following relationship is attained:

$$\|x_{k+1} - x^*\| = o(\|w_k - w^*\|). \quad (75)$$

Note that this is not a surprising result because we know that the step taken by the Gauss-Seidel decomposition algorithm on the global variables, x , is the same as that of a SIPM.

To finish the proof, it only remains to show that the local variable component, \hat{y} , satisfies a similar relationship. The local component of Equation (74) can be written as

$$\begin{aligned} \Lambda_{k,\hat{y}} M_k^{-1}(\hat{A}_k(x_k - x^*) - (g_{2,k}^+(0) - g_{2,k}(0))) \\ = \Lambda_{k,\hat{y}} M_k^{-1}(\hat{A}_k(x_{k+1} - x^*) - (g_{2,k}^+(0) - g_{2,k}(0)) - \hat{A}_k(x_{k+1} - x_k)), \end{aligned} \quad (76)$$

which by Taylor's theorem and the fact that $\hat{A}_k = -\nabla_x g_{2,k}(\mu_k) = \nabla_x g_{2,k}(0)$ is

$$\begin{aligned} \Lambda_{k,\hat{y}} M_k^{-1}(\hat{A}_k(x_{k+1} - x^*) - (g_{2,k}^+(0) - g_{2,k}(0)) - \hat{A}_k(x_{k+1} - x_k)) \\ = \Lambda_{k,\hat{y}} M_k^{-1}(\hat{A}_k(x_{k+1} - x^*) + O(\|x_{k+1} - x_k\|^2)). \end{aligned} \quad (77)$$

Because

$$x_{k+1} - x_k = \Delta x_k^G = -(S_k)^{-1}(g_{1,k} + \hat{A}_k^T M_k^{-1} g_{2,k}(\mu_k)), \quad (78)$$

we conclude that

$$\|x_{k+1} - x_k\| = O(\|g_k(\mu_k)\|), \quad (79)$$

and thus the second term in the right-hand side of (77) is of order $O(\|w_k - w^*\|^2)$.

Moreover, we know by (75) that the first term in the right-hand side of (77) is of order $o(\|w_k - w^*\|)$. This, together with the local variable component in (68), (70), (71), give

$$\|\hat{y}_{k+1} - \hat{y}^*\| = o(\|w_{k+1} - w^*\|). \quad (80)$$

Relationships (75) and (80) prove the result. \square

5.4. Discussion. Our analysis in §§4 and 5 gives some interesting insight into why certain BDAs may converge slowly in practice. Most practical BDAs follow the general scheme outlined in Figure 2. These algorithms usually set the global variables to a fixed value at each iteration to break the problem into independent subproblems. Our analysis shows that in the case of the inexact BDA, fixing the global variables leads to a set of subproblems that ignore any information regarding the search direction on the global variable space. For this reason, the inexact BDA may fail to achieve fast convergence in the vicinity of the minimizer. We have also shown that a Gauss-Seidel technique may be used to overcome this difficulty in the case of the inexact BDA. Moreover, because in Propositions 4.1 and 4.2 we showed that the inexact BDA is related to the more general BDAs outlined in Figure 2, our analysis of the inexact and Gauss-Seidel BDAs throws some light into why some of the more general BDAs used by engineers in practice may converge slowly occasionally. We hope that our work will help researchers and practitioners alike to develop BDAs with better convergence properties.

6. An exact penalty formulation of the PSP. The SIPM described in §3 can only be applied to problems satisfying the SLICQ. In particular, Proposition 3.1 shows that if SLICQ holds, then the Schur complement matrix is invertible. As we mentioned in §2, the SLICQ implies that the PSP constraints can be used in the vicinity of the minimizer to eliminate the local and slack variables y and r , and thus transform the PSP into an unconstrained problem on only the global variables. This is a fairly strong assumption for certain applications such as MDO, where often the difficulty is precisely finding a design that is feasible with respect to the constraints of all disciplines.

In this section, we show how a PSP satisfying the LICQ but not SLICQ may be *regularized* by introducing an exact penalty function. Specifically, we show how any PSP satisfying LICQ can be reformulated by means of an exact penalty function as a PSP satisfying the so-called strong Mangasarian-Fromovitz constraint qualification (SMFCQ). Then, in §7, we show how degenerate optimization techniques can be used to ensure the SIPM converges locally at a fast rate when applied to solve a PSP satisfying SMFCQ. Essentially, our work allows the application of SIPMs to solve general PSPs satisfying only the conventional LICQ.

We now define the SMFCQ.

CONDITION C.2. The SMFCQ is satisfied at w^* ; that is, the matrix

$$(\nabla_y c(x^*, y^*) \quad -I) \tag{81}$$

has full row rank and there exist Δy and Δr such that $\nabla_y c(x^*, y^*)\Delta y - \Delta r = 0$ and $(\Delta r)_i > 0$ for all $i \in \mathcal{L}$, where \mathcal{L} is the active set $\{i: r_i^* = 0\}$ and $I_{\mathcal{L}}$ is the matrix formed by the rows of the identity corresponding to indices in \mathcal{L} .

Note that by Assumption A.2, the Lagrange multiplier vector (λ^*, σ^*) associated with (x^*, y^*, r^*) is unique. Also the matrix (81) always has full row rank if all problem constraints are inequality constraints.

Note that the SMFCQ (C.2) holds if and only if the conventional Mangasarian-Fromovitz constraint qualification (MFCQ) holds for the subproblem (2) with $x = x^*$; that is, for the following problem:

$$\begin{aligned} & \min_{y,r} F(x^*, y) \\ & \text{subject to } c(x^*, y) - r = 0, \\ & r \geq 0. \end{aligned} \tag{82}$$

Likewise, the SLICQ (C.1) holds if and only if the conventional LICQ holds for the subproblem (2) with $x = x^*$. Consequently, satisfaction of the SLICQ implies satisfaction of the SMFCQ.

In the remainder of this section, we show how any PSP that satisfies the LICQ (A.2) can be reformulated by introducing an exact penalty function as an equivalent problem for which the SMFCQ holds. To see this, note that by using an l_1 exact penalty function and a sufficiently large but finite penalty parameter γ , the PSP (2) can be reformulated as follows:

$$\begin{aligned} & \min_{x,z,y,r} F(z, y) + \gamma \|x - z\|_1 \\ & \text{subject to } c(z, y) - r = 0, \\ & r \geq 0, \end{aligned} \tag{83}$$

where $z \in \mathcal{R}^n$ is a vector of auxiliary local variables that will be equal to the global variables x for a sufficiently large γ .

Moreover, by introducing slack variables, problem (83) can be transformed in the following smooth optimization problem:

$$\begin{aligned} \min_{x, z, y, r, s, t} \quad & F(z, y) + \gamma e_n^T (s + t) \\ \text{subject to} \quad & c(z, y) - r = 0, \\ & z + s - t = x, \\ & r, s, t \geq 0. \end{aligned} \tag{84}$$

A couple of comments are in order. First, note that the proposed regularization only requires the introduction of linear constraints and linear terms in the objective function. Thus the reformulation will preserve any nice structure available in the original problem. Second, all variables introduced (z , s , and t) must be considered as *local* variables. That is, in the general case where there are K subproblems, we must introduce a different set of variables (z_i , s_i , t_i) for each of the subproblems; that is, for $i = 1, \dots, K$. As a result, the variables introduced can be considered as part of the local variable vector \hat{y} , and thus the K subproblems are still only coupled through the global variables x .

The following proposition shows that satisfaction of LICQ for the original problem implies satisfaction of SMFCQ for problem (84).

PROPOSITION 6.1. *If the LICQ (Assumption A.2) holds at (x^*, y^*, r^*) for the PSP (2), then the SMFCQ (Condition C.2) holds at $(x, z, y, r, s, t) = (x^*, x^*, y^*, r^*, 0, 0)$ for (84).*

PROOF. We only need to show that the standard MFCQ holds at $(z, y, r, s, t) = (x^*, y^*, r^*, 0, 0)$ for the subproblem that results from fixing the global variables x to x^* in problem (84); that is, for problem

$$\begin{aligned} \min_{z, y, r, s, t} \quad & F(z, y) + \gamma e_n^T (s + t) \\ \text{subject to} \quad & c(z, y) - r = 0, \\ & z + s - t = x^*, \\ & r, s, t \geq 0. \end{aligned} \tag{85}$$

Note that the gradients with respect to the subproblem variables (z, y, r, s, t) of the *equality* constraints are the rows of the following matrix

$$\begin{pmatrix} \nabla_z c(z, y) & \nabla_y c(z, y) & -I & 0 & 0 \\ I & 0 & 0 & I & -I \end{pmatrix}.$$

Moreover, these gradients are linearly independent at any feasible point for (85) because of the presence of the identity matrices that correspond to the gradients of the equality constraints with respect to the slack and elastic variables r , s , and t . Moreover, because the LICQ holds for (2), we know that the MFCQ also holds for (2), and thus there exists $(\Delta x, \Delta y, \Delta r)$ such that

$$\nabla_x c(x^*, y^*) \Delta x + \nabla_y c(x^*, y^*) \Delta y - \Delta r = 0$$

and $\Delta r_{\mathcal{Z}} > 0$, where \mathcal{Z} is the set of indices i for which the nonnegativity bounds $r_i \geq 0$ are active at r^* . But then the vector

$$(\Delta z, \Delta y, \Delta r, \Delta s, \Delta t) = (\Delta x, \Delta y, \Delta r, \max(0, -\Delta x) + e_n, \max(0, \Delta x) + e_n)$$

satisfies

$$\begin{aligned} \nabla_x c(x^*, y^*) \Delta z + \nabla_y c(x^*, y^*) \Delta y - \Delta r &= 0, \\ \Delta z + \Delta s - \Delta t &= 0, \end{aligned}$$

and $\Delta r_{\mathcal{Z}}, \Delta s, \Delta t > 0$. Thus the MFCQ holds for subproblem (85), and therefore the SMFCQ holds for (84). \square

For instance, consider the simple example given in (11). For $(a, b) = (2, 1)$, the minimizer is $(x^*, y^*) = (1, 1)$ and it only satisfies LICQ, but not SLICQ. The problem can be reformulated equivalently as

$$\begin{aligned} \min_{x, z, y, s, t} \quad & \frac{1}{2}(z - 2)^2 + \frac{1}{2}(y - 1)^2 + 100(s + t) \\ \text{s.t.} \quad & z + y \leq 2, \\ & z - y \leq 0, \\ & -x + z + s - t = 0, \\ & s, t \geq 0. \end{aligned}$$

It is straightforward to show that the minimizer to this problem is $(x^*, z^*, y^*, s^*, t^*) = (1, 1, 1, 0, 0)$ and that it satisfies SMFCQ.

7. A degenerate Schur approach. In this section, we show how degenerate optimization techniques can be used to ensure the SIPM converges locally at a fast rate when applied to solve a PSP satisfying SMFCQ. In the previous section, we showed how a general PSP satisfying simply the LICQ can be reformulated as a PSP satisfying SMFCQ. As a result, our work extends the applicability of SIPMs to cover general PSPs satisfying the conventional LICQ.

Our proposed SIPM borrows from the degenerate interior-point method developed by Vicente and Wright [35]. There are two differences between their method and a standard interior-point method. Firstly, they modify the Newton matrix by perturbing some of the elements in its diagonal. As we show in this section, this perturbation ensures the Schur complement matrix is invertible even when replacing SLICQ by SMFCQ. The second difference is in the step-size rule: it allows some of the nonnegative variables to become negative if sufficient progress is made toward the solution. Raghunathan and Biegler [26] show that the methodology of Vicente and Wright [35] can work with other step-size rules, but to simplify the exposition, we use the same step-size rule as in Vicente and Wright [35].

7.1. The method. The modified Newton matrix is obtained by perturbing the current iterate for the variables r_k and σ_k as follows:

$$\tilde{r}_k = \max(\mu_k^{\min} e_m, r_k), \quad \tilde{\sigma}_k = \max(\mu_k^{\min} e_m, \sigma_k), \quad (86)$$

where μ_k^{\min} is a positive value and \max is the componentwise maximum. With this perturbation, the modified matrix \tilde{M}_k is

$$\tilde{M}_k = \nabla_{\hat{y}} g_2(x_k, y_k, \tilde{r}_k, \lambda_k, \tilde{\sigma}_k), \quad (87)$$

the modified Newton matrix is

$$\tilde{K}_k^N = \begin{pmatrix} W_k & -\hat{A}_k^T \\ -\hat{A}_k & \tilde{M}_k \end{pmatrix}, \quad (88)$$

and the modified Schur complement matrix is

$$\tilde{S}_k = W_k - \hat{A}_k^T (\tilde{M}_k)^{-1} \hat{A}_k. \quad (89)$$

We use the same step size α_k for all components of w_k ; that is, $w_{k+1} = w_k + \alpha_k \Delta w_k$. Moreover, the step-size rule is computed as follows:

If both of the following conditions hold (where $\tau \in (1, 2)$ and $(a)_-$ is the negative part of the vector a):

$$\|(r_k + \Delta r_k, \lambda_k + \Delta \lambda_k, \sigma_k + \Delta \sigma_k)_-\| \leq \|g_k(0)\|^\tau, \quad (90)$$

$$\|g(w_k + \Delta w_k; \mu = 0)\| \leq \|g_k(0)\|^\tau, \quad (91)$$

then $\alpha_k = 1$.

Otherwise,

$$\alpha_k = \min(\alpha_{r,k}, \alpha_{\lambda,k}, \alpha_{\sigma,k}), \quad (92)$$

where $\alpha_{r,k}$, $\alpha_{\lambda,k}$, and $\alpha_{\sigma,k}$ are given by (22)–(24).

The resulting modified SIPM is stated in Figure 6.

Initialization: Choose a starting point $w_0 = (x_0, y_0, r_0, \lambda_0, \sigma_0)$ such that $r_0 > 0$, $\lambda_0 > 0$, $\sigma_0 > 0$. Set $k \leftarrow 0$ and choose the parameters $\mu_0 > 0$, $\mu_0^{\min} > 0$ and $\tau \in (1, 2)$.

Repeat

- (i) **Master problem iteration:** Compute Δx_k by solving system (55) with \tilde{S}_k in (89).
- (ii) **Subproblem iteration:** Compute $\Delta \hat{y}_k$ by solving system (30) with \tilde{M}_k in (87).
- (iii) **Line search:** Compute the step size α_k as in (90)–(92).
- (iv) **Update iterate:** Set $w_{k+1} = w_k + \alpha_k \Delta w_k$, where $\Delta w_k = (\Delta x_k, \Delta \hat{y}_k)$.
- (v) **Parameter update:** Set $\mu_{k+1} > 0$, $\mu_{k+1}^{\min} > 0$ and $k \leftarrow k + 1$.

Until convergence

FIGURE 6. Modified SIPM.

7.2. Nonsingularity of \tilde{M}_k and \tilde{S}_k . In this section, we show that the modified matrices \tilde{M}_k and \tilde{S}_k are invertible in the vicinity of a minimizer. This, in turn, implies that the iterates generated by the modified SIPM are well defined in the vicinity of a minimizer.

First, the following proposition shows that the modified Newton matrix in (88) is invertible in the vicinity of a minimizer.

PROPOSITION 7.1. *If Assumptions A.1–A.4 hold and $\mu_k^{\min} = O(\|g_k(0)\|)$, then $\|(\tilde{K}_k^N)^{-1}\|$ is bounded for w_k in a neighborhood of the minimizer w^* .*

PROOF. By Assumptions A.1–A.4, we know by Fiacco and McCormick [17, Theorem 14] and El-Bakry et al. [16, Proposition 4.1] that $\|(K_k^N)^{-1}\|$ is bounded in the vicinity of the minimizer. Moreover, $\tilde{K}_k^N = K_k^N + D_k$, with $\|D_k\| = O(\mu_k^{\min}) = O(\|g_k(0)\|)$. Consequently, $\|(\tilde{K}_k^N)^{-1}\|$ is bounded in a small enough neighborhood around the minimizer. \square

The following proposition shows that \tilde{M}_k is invertible in the vicinity of the minimizer. The notation $a_k = \Theta(b_k)$ means that there exists $\beta_1, \beta_2 > 0$ such that for all k sufficiently large, we have that $0 \leq \beta_1 \leq a_k/b_k \leq \beta_2$.

PROPOSITION 7.2. *If Assumptions A.1–A.4 and Condition C.2 hold and $\mu_k^{\min} = \Theta(\|g_k(0)\|)$, then \tilde{M}_k is nonsingular for w_k in a neighborhood of the minimizer w^* .*

PROOF. The proof is made by contradiction. Assume there exists a nonzero vector $\Delta w = (\Delta y, \Delta r, \Delta \lambda, \Delta \sigma)$ such that $\tilde{M}_k \Delta w = 0$. Then (we have omitted the subscript k for convenience),

$$\begin{pmatrix} H_{yy} & 0 & -(\nabla_y c(x, y))^T & 0 \\ 0 & 0 & I & -I \\ -\nabla_y c(x, y) & I & 0 & 0 \\ 0 & -\tilde{\Sigma} & 0 & -\tilde{R} \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta r \\ \Delta \lambda \\ \Delta \sigma \end{pmatrix} = 0, \tag{93}$$

where $H_{yy} = \nabla_{yy}^2 F(x, y) - \sum_{i=1}^m \lambda_i \nabla_{yy}^2 c_i(x, y)$.

Let \mathcal{B} be the inactive set $\{i: r_i^* > 0\}$ and \mathcal{I} be the active set $\{i: r_i^* = 0\}$. Using this partition, because $\tilde{R}_{\mathcal{B}}$ is invertible, the system (93) can be symmetrized and written in the following form:

$$\begin{pmatrix} H_{yy} & 0 & -(\nabla_y c(x, y))^T & 0 \\ 0 & E & I & -I_{\mathcal{I}}^T \\ -\nabla_y c(x, y) & I & 0 & 0 \\ 0 & -I_{\mathcal{I}} & 0 & -\tilde{\Sigma}_{\mathcal{I}}^{-1} \tilde{R}_{\mathcal{I}} \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta r \\ \Delta \lambda \\ \Delta \sigma_{\mathcal{I}} \end{pmatrix} = 0 \tag{94}$$

with

$$\Delta \sigma_{\mathcal{B}} = -\tilde{R}_{\mathcal{B}}^{-1} \tilde{\Sigma}_{\mathcal{B}} \Delta r_{\mathcal{B}}, \tag{95}$$

and where

$$E = \begin{pmatrix} \tilde{R}_{\mathcal{B}}^{-1} \tilde{\Sigma}_{\mathcal{B}} & 0 \\ 0 & 0 \end{pmatrix},$$

and $I_{\mathcal{I}}$ is the matrix formed by the rows of the identity corresponding to indices in \mathcal{I} .²

First, it is convenient to transform the system (94) via the singular value decomposition of the Jacobian matrix of the active constraints at (y^*, r^*) . The SVD of this Jacobian matrix can be formulated as

$$J^* = \begin{pmatrix} -\nabla_y c(x^*, y^*) & I \\ 0 & -I_{\mathcal{I}} \end{pmatrix} = (U \quad V) \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{U}^T \\ \hat{V}^T \end{pmatrix}, \tag{96}$$

where $U \in \mathcal{R}^{(m+|\mathcal{I}|) \times p}$, $C \in \mathcal{R}^{p \times p}$, $V \in \mathcal{R}^{(m+|\mathcal{I}|) \times (m+|\mathcal{I}|-p)}$, $\hat{U} \in \mathcal{R}^{(n_y+m) \times p}$, $\hat{V} \in \mathcal{R}^{(n_y+m) \times (n_y+m-p)}$, and p is the rank of the Jacobian matrix J^* . Moreover, we will denote by U_1 and V_1 , the first m rows of U and V , respectively; by U_2 and V_2 , the last $|\mathcal{I}|$ rows of U and V , respectively; by \hat{U}_1 and \hat{V}_1 , the first n_y rows of \hat{U} and \hat{V} , respectively; and finally, by \hat{U}_2 and \hat{V}_2 , the last m rows of \hat{U} and \hat{V} , respectively.

² Note that the matrix (10) used in the definition of SLICQ is the 2×2 block at the bottom left corner of matrix (94).

Using the SVD in (96), the following (orthogonal) change of variables can be applied:

$$\begin{pmatrix} \Delta y \\ \Delta r \\ \Delta \lambda \\ \Delta \sigma_{\mathcal{X}} \end{pmatrix} = \begin{pmatrix} \hat{U}_1 & \hat{V}_1 & 0 & 0 \\ \hat{U}_2 & \hat{V}_2 & 0 & 0 \\ 0 & 0 & U_1 & V_1 \\ 0 & 0 & U_2 & V_2 \end{pmatrix} \begin{pmatrix} c_{\hat{U}} \\ c_{\hat{V}} \\ c_U \\ c_V \end{pmatrix}, \quad (97)$$

where $c_{\hat{U}} \in \mathcal{R}^p$, $c_{\hat{V}} \in \mathcal{R}^{n_y+m-p}$, $c_U \in \mathcal{R}^p$, and $c_V \in \mathcal{R}^{m+|\mathcal{X}|-p}$. With this change of variables, system (93) is equivalent to

$$\begin{pmatrix} \hat{U}^T G \hat{U} & \hat{U}^T G \hat{V} & \hat{U}^T J^T U & \hat{U}^T J^T V \\ \hat{V}^T G \hat{U} & \hat{V}^T G \hat{V} & \hat{V}^T J^T U & \hat{V}^T J^T V \\ U^T J \hat{U} & U^T J \hat{V} & U^T N U & U^T N V \\ V^T J \hat{U} & V^T J \hat{V} & V^T N U & V^T N V \end{pmatrix} \begin{pmatrix} c_{\hat{U}} \\ c_{\hat{V}} \\ c_U \\ c_V \end{pmatrix} = 0, \quad (98)$$

where

$$G = \begin{pmatrix} H_{yy} & 0 \\ 0 & E \end{pmatrix} \quad \text{and} \quad N = \begin{pmatrix} 0 & 0 \\ 0 & -\tilde{\Sigma}_{\mathcal{X}}^{-1} \tilde{R}_{\mathcal{X}} \end{pmatrix}.$$

Under Assumptions A.1, A.3, A4, and Condition C.2, system (98) can be written near the minimizer as follows (see Vicente and Wright [35]):

$$\begin{pmatrix} D_1 + O(\|g_k(0)\|) & O(\|g_k(0)\|) \\ O(\|g_k(0)\|) & -D_2 \end{pmatrix} \begin{pmatrix} \hat{c} \\ c_V \end{pmatrix} = 0, \quad (99)$$

where $\hat{c} = (c_{\hat{U}}, c_{\hat{V}}, c_U)$,

$$D_1 = \begin{pmatrix} \hat{U}^T G^* \hat{U} & \hat{U}^T G^* \hat{V} & C \\ \hat{V}^T G^* \hat{U} & \hat{V}^T G^* \hat{V} & 0 \\ C & 0 & 0 \end{pmatrix}, \quad \text{with } G^* = \begin{pmatrix} H_{yy}^* & 0 \\ 0 & E^* \end{pmatrix} \quad (100)$$

and

$$D_2 = V_2^T (\tilde{\Sigma}_{\mathcal{X}})^{-1} \tilde{R}_{\mathcal{X}} V_2. \quad (101)$$

Note that because of the modification introduced in the Newton matrix, $\|D_2\| = \Theta(\|g_k(0)\|)$, and thus D_2 is invertible in the vicinity of the minimizer. Therefore we can use the second row of the system in (99) to compute c_V as

$$c_V = D_2^{-1} O(\|g_k(0)\|) \hat{c}.$$

Substituting c_V in the first row of the system in (99), we obtain

$$(D_1 + O(\|g_k(0)\|) + O(\|g_k(0)\|) D_2^{-1} O(\|g_k(0)\|)) \hat{c} = 0. \quad (102)$$

But note that $O(\|g_k(0)\|) D_2^{-1} O(\|g_k(0)\|) = O(\|g_k(0)\|)$, and thus we have

$$(D_1 + O(\|g_k(0)\|)) \hat{c} = 0. \quad (103)$$

On the other hand, by the results in Vicente and Wright [35, Lemma 4.1], we know that under Assumptions A.1, A.3, A.4, and Condition C.2, near the solution, the matrix D_1 is a $O(\|g_k(0)\|)$ perturbation of a nonsingular matrix, and hence it is uniformly nonsingular. This contradicts system (103) and concludes that matrix in (94) is nonsingular. Therefore, \tilde{M}_k must be invertible in a neighborhood of the minimizer. \square

Finally, the following proposition shows that \tilde{S}_k is invertible in the vicinity of the minimizer.

PROPOSITION 7.3. *If Assumptions A.1–A.4 and Condition C.2 hold and $\mu_k^{\min} = \Theta(\|g_k(0)\|)$, then \tilde{S}_k is nonsingular for w_k in a neighborhood of the minimizer w^* .*

PROOF. By Propositions 7.1 and 7.2, the modified Newton matrix \tilde{K}_k^N and the matrix \tilde{M}_k are nonsingular, respectively, in a neighborhood of the minimizer. This implies that \tilde{S}_k must be invertible, or otherwise there would be multiple solutions to the modified Newton system

$$\tilde{K}_k^N \Delta w_k = -g_k(\mu_k), \quad (104)$$

in contradiction with the invertibility of \tilde{K}_k^N . \square

7.3. Superlinear convergence. Finally, we show that the proposed modification of the SIPM maintains the good local convergence properties even when the SLICQ is replaced by SMFCQ.

THEOREM 7.1. *Suppose that Assumptions A.1–A.4 and Condition C.2 hold, that the barrier parameter is chosen to satisfy $\mu_k = O(\|g_k(0)\|^2)$ and that $\mu_k^{\min} = \Theta(\|g_k(0)\|)$. If w_0 is close enough to w^* , then the sequence $\{w_k\}$ generated by the modified SIPM method described in Figure 6 is well defined and converges to w^* at a quadratic rate.*

PROOF. Propositions 7.2 and 7.3 imply that the matrices \tilde{M}_k and \tilde{S}_k are nonsingular in a neighborhood of the minimizer, and thus the search direction of the SIPM method is well defined in this neighborhood.

Moreover, Assumptions A.1–A.4 ensure that the conditions in Vicente and Wright [35, Theorem 4.3] hold at the minimizer w^* . Therefore the proposed modification of the SIPM converges locally at a quadratic rate. \square

8. Conclusions. We establish a theoretical relationship between BDAs and SIPMs. This connection is established through the inexact BDA, which we show is a close relative of both BDAs and SIPMs. The relevance of this relationship is that it is a first step toward closing the gap between the incipient local convergence theory of BDAs (Alexandrov and Lewis [2], DeMiguel and Murray [14]) and the mature local convergence theory of interior-point methods (Martinez et al. [24], El-Bakry et al. [16], Yamashita and Yabe [36], Gould et al. [21]). As a result, we think our analysis constitutes one important step toward the development of a robust convergence theory for BDAs.

Our second contribution is to show how SIPMs can be applied to solve PSPs that satisfy only the conventional LICQ but not the Strong LICQ. We accomplish this task in two steps. First, we show that any PSP satisfying the conventional LICQ can be reformulated by introducing an exact penalty function as a problem satisfying SMFCQ. Second, we show how degenerate optimization techniques can be used to ensure an SIPM will achieve fast local convergence when applied to solve a problem satisfying SMFCQ. The importance of this contribution is that it substantially enlarges the class of problems that can be addressed with SIPMs to include general PSPs satisfying LICQ. This makes SIPMs a viable alternative to certain practical BDAs (Braun [5], DeMiguel and Murray [14]), which can deal with problems that satisfy the LICQ and not the SLICQ.

Acknowledgments. The authors are grateful to the Area Editor Jong-Shi Pang, the Associate Editor, and two referees for their careful feedback. The authors would also like to thank Jorge Nocedal, Danny Ralph, and S. Scholtes for their detailed feedback and suggestions. The authors also acknowledge comments from R. W. Cottle, M. P. Friedlander, F. J. Prieto, L. Vicente, S. Wright, and seminar participants at the SIAM Conference on Optimization (Stockholm 2005), Argonne National Laboratory, Judge Institute of Cambridge, and Northwestern University. This research was partially supported by the Research Development Fund at London Business School and by the Ministerio de Educación y Ciencia of Spain through Project MTM2004-02334.

References

- [1] Alexandrov, N. M., M. Y. Hussaini, eds. 1997. *Multidisciplinary Design Optimization: State of the Art*. SIAM, Philadelphia.
- [2] Alexandrov, N. M., R. M. Lewis. 2000. Analytical and computational aspects of collaborative optimization. Technical Report TM-2000-210104, NASA, Hampton, VA.
- [3] Benders, J. F. 1962. Partitioning procedures for solving mixed variables programming problems. *Numerische Mathematik* 4 238–252.
- [4] Birge, J. R., F. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer-Verlag, New York.
- [5] Braun, R. D. 1996. Collaborative optimization: An architecture for large-scale distributed design. Doctoral dissertation, Stanford University, Stanford, CA.
- [6] Braun, R. D., I. M. Kroo. 1997. Development and application of the collaborative optimization architecture in a multidisciplinary design environment. N. M. Alexandrov, M. Y. Hussaini, eds. *Multidisciplinary Design Optimization: State of the Art*. SIAM, Philadelphia, 98–116.
- [7] Byrd, R. H., M. E. Hribar, J. Nocedal. 1999. An interior point algorithm for large-scale nonlinear programming. *SIAM J. Optim.* 9 877–900.
- [8] Chun, B. J., S. M. Robinson. 1995. Scenario analysis via bundle decomposition. *Ann. Oper. Res.* 56 39–63.
- [9] Cohen, G., B. Miara. 1990. Optimization with an auxiliary constraint and decomposition. *SIAM J. Control Optim.* 28(1) 137–157.
- [10] Colson, B., P. Marcotte, G. Savard. 2005. Bilevel programming: A survey. *4OR* 3 87–107.
- [11] Cottle, R. W. 1974. Manifestations of the Schur complement. *Linear Algebra Appl.* 8 189–211.
- [12] Cramer, E. J., J. E. Dennis, P. D. Frank, R. M. Lewis, G. R. Shubin. 1994. Problem formulation for multidisciplinary optimization. *SIAM J. Optim.* 4(4) 754–776.
- [13] DeMiguel, V., W. Murray. 2000. An analysis of collaborative optimization methods. *Eighth AIAA/USAF/NASA/ISSMO Sympos. Multidisciplinary Anal. and Optim., Long Beach, CA*, Paper 00-4720, AIAA, Reston, VA.
- [14] DeMiguel, V., W. Murray. 2006. A local convergence analysis of bilevel decomposition algorithms. *Optim. Engrg.* 7 99–133.
- [15] Dempe, S. 2002. *Foundations of Bilevel Programming*. Kluwer Academic Publishers, Boston.

- [16] El-Bakry, A. S., R. A. Tapia, T. Tsuchiya, Y. Zhang. 1996. On the formulation and theory of Newton interior-point method for nonlinear programming. *J. Optim. Theory Appl.* **89** 507–541.
- [17] Fiacco, A. V., G. P. McCormick. 1968. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, New York.
- [18] Forsgren, A., P. E. Gill. 1998. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM J. Optim.* **8**(4) 1132–1152.
- [19] Gay, D. M., M. L. Overton, M. H. Wright. 1997. A primal-dual interior method for nonconvex nonlinear programming. Technical Report 97-4-08, Computing Sciences Research, Bell Laboratories, Murray Hill, NJ.
- [20] Geoffrion, A. M. 1972. Generalized Benders decomposition. *J. Optim. Theory Appl.* **10**(4) 237–260.
- [21] Gould, N. I. M., D. Orban, A. Sartenaer, P. L. Toint. 2001. Superlinear convergence of primal-dual interior point algorithms for nonlinear programming. *SIAM J. Optim.* **11**(4) 974–1002.
- [22] Hafka, R. T., J. Sobieszczanski-Sobieski. 1997. Multidisciplinary aerospace design optimization: Survey of recent developments. *Structural Optim.* **14** 1–23.
- [23] Helgason, T., S. W. Wallace. 1991. Approximate scenario solutions in the progressive hedging algorithm. *Ann. Oper. Res.* **31** 425–444.
- [24] Martinez, H. H., Z. Parada, R. A. Tapia. 1995. On the characterization of q -superlinear convergence of quasi-Newton interior-point methods for nonlinear programming. *Boletín de la Sociedad Matemática Mexicana* **1** 1–12.
- [25] Medhi, D. 1990. Parallel bundle-based decomposition for large-scale structured mathematical programming problems. *Ann. Oper. Res.* **22** 101–127.
- [26] Raghunathan, A. U., L. T. Biegler. 2003. Interior point methods for mathematical programs with complementarity constraints. Technical report, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh.
- [27] Robinson, S. M. 1986. Bundle-based decomposition: Description and preliminary results. A. Prekopa, J. Szelezsan, B. Strazicky, eds. *System Modelling and Optimization. Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 751–756.
- [28] Rockafellar, R. T., R. J.-B. Wets. 1991. Scenarios and policy aggregation in optimization under uncertainty. *Math. Oper. Res.* **16** 119–147.
- [29] Ruszczyński, A. 1986. A regularized decomposition for minimizing a sum of polyhedral functions. *Math. Programming* **35** 309–333.
- [30] Ruszczyński, A. 1995. On convergence of an augmented Lagrangian decomposition method for sparse convex optimization. *Math. Oper. Res.* **20**(3) 634–656.
- [31] Shimizu, K., Y. Ishizuka, J. F. Bard. 1997. *Nondifferentiable and Two-Level Mathematical Programming*. Kluwer Academic Publishers, Boston.
- [32] Tammer, K. 1987. The application of parametric optimization and imbedding for the foundation and realization of a generalized primal decomposition approach. J. Guddat, H. Jongen, B. Kummer, F. Nozicka, eds. *Parametric Optimization and Related Topics, Mathematical Research*, Vol. 35. Akademie-Verlag, Berlin, 376–386.
- [33] Van Slyke, R., R. J.-B. Wets. 1969. L-shaped linear programs with application to optimal control and stochastic programming. *SIAM J. Appl. Math.* **17** 638–663.
- [34] Vanderbei, R. J., D. F. Shanno. 1997. An interior-point algorithm for nonconvex nonlinear programming. Technical Report SOR-97-21, Statistics and Operations Research, Princeton University, Princeton, NJ.
- [35] Vicente, L. N., S. J. Wright. 2002. Local convergence of a primal-dual method for degenerate nonlinear programming. *Computational Optim. Appl.* **22** 311–328.
- [36] Yamashita, H., H. Yabe. 1996. Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization. *Math. Programming* **75** 377–397.