Multidisciplinary Design Optimization with Quasiseparable Subsystems

RAPHAEL T. HAFTKA
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611-6250

LAYNE T. WATSON
Departments of Computer Science and Mathematics, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061-0106

Abstract. Numerous hierarchical and nonhierarchical decomposition strategies for the optimization of large scale systems, comprised of interacting subsystems, have been proposed. With a few exceptions, all of these strategies lack a rigorous theoretical justification. This paper focuses on a class of quasiseparable optimization problems narrow enough for a rigorous decomposition theory, yet general enough to encompass many large scale engineering design problems. The subsystems for these problems involve local design variables and global system variables, but no variables from other subsystems. The objective function is a sum of a global system criterion and the subsystems' criteria. The essential idea is to give each subsystem a budget and global system variable values, and then ask the subsystems to independently maximize their constraint margins. Using these constraint margins, a system optimization then adjusts the values of the system variables and subsystem budgets. The subsystem margin problems are totally independent, always feasible, and could even be done asynchronously in a parallel computing context. An important detail is that the subsystem tasks, in practice, would be to construct response surface approximations to the constraint margin functions, and the system level optimization would use these margin surrogate functions. The purpose of the present paper is to present a decomposition strategy in a general context, provide rigorous theory justifying the decomposition, and give some simple illustrative examples.

Keywords: decomposition, global/local optimization, multidisciplinary design, response surface approximation, separable

1. Introduction

Numerous hierarchical and nonhierarchical decomposition strategies for the optimization of large scale systems, comprised of interacting subsystems, have been proposed—Dantzig-Wolfe decomposition, concurrent subspace optimization (CSSO) (Rodríguez et al., 1998a), Sobieski's nonhierarchical decomposition (Shankar et al., 1993), Kroo's collaborative optimization (CO) (Kroo, 1997), multilevel decomposition (Alexandrov, 1997), Renaud and Watson's augmented Lagrangian approach (Rodríguez et al., 1998b), to name but a few. Many decomposition formulations in the engineering literature are heuristic, without rigorous mathematical justification. This paper focuses on a class of quasiseparable optimization problems narrow enough for a rigorous decomposition theory, yet general enough to encompass many large scale engineering design problems, including multidisciplinary design optimization (MDO) problems. The subsystems for these problems involve local design variables and global system variables, but no variables from other subsystems. The objective function is a sum of a global system criterion and the subsystems' criteria.

The essential idea is to give each subsystem a budget and global system variable values, and then ask the subsystems to independently maximize their constraint margins. (A constraint margin is the amount of slack a constraint has at a particular design point.) Using these constraint margins, a system optimization then adjusts the values of the system variables and subsystem budgets. The subsystem margin problems are totally independent, always feasible, and could even be done asynchronously in a parallel computing context. An important implementation detail is that the subsystem tasks, in practice, would be to construct response surface approximations to the constraint margin functions, and the system level optimization would use these margin function surrogates. The details of how these margin surrogate functions would be constructed, communicated to the system level, and rigorous theory justifying these processes, are topics for future research that are not addressed here.

An example of the type of MDO problem targeted by the decomposition theory presented here is aircraft design. Aircraft design may be viewed as a multilevel optimization problem. At the highest level the entire aircraft is designed; at a lower level major systems and components, such as the wing, fuselage, and propulsion system, are designed; at an even lower level subcomponents, such as individual wing panels, are designed. Unfortunately, because of computational limitations, it has not been possible to perform this multilevel optimization rigorously. Instead, various ad hoc design procedures are used to decouple the three levels.

Typical work has focused on the use of response surface approximations, fit to the values of multiple lower level optima, as a way to integrate the various design levels (Giunta et al., 1997b). For example, Balabanov et al. (1999) used response surface approximations fit to thousands of wing structural optimizations in the overall design of a high-speed civil transport. Response surface approximations to integrate overall wing structural optimization with the design of individual panels were also used by Ragon et al. (1997) and Liu et al. (2000). This involved performing hundreds of panel optimizations, and fitting the optima response surface with a polynomial for use by the wing optimizer. It was verified by Liu and Haftka (2003) that the two-level procedure led to the same design as a single level optimization.

Some early work performed lower level optimization using the same objective function (structural weight) at both wing and panel level optimizations (Ragon et al., 1997). This led to problems, in that for some configurations there were no feasible solutions to the lower level optimization. This problem was overcome by switching to a maximum margin formulation at the lower level. That is, the panel level optimization maximized the constraint margin rather than minimized the weight, so that even when no feasible design was available, it found the design with the smallest constraint violation.

A decomposition procedure is particularly important because of the need for global and/or discrete optimization in panel design. Structural panels are characterized by a large number of choices, such as type of stiffeners (blade, hat, zee, etc.), materials, and stacking sequences of laminates. These problems can be formulated as discrete optimization problems, or alternatively as continuous problems with a large number of local optima. It is practically impossible to design a complex structure, such as a wing made of hundreds of panels, as a single global or discrete optimization problem. The decomposition approach in Ragon et al. (1997) made it possible to apply global optimization for the design of each panel, and maintain coupling between the panels through the maximum margin response surface approximation.

Even though the application of the decomposition procedure of Ragon et al. (1997) was successful, no attempt was made previously to explore its theoretical properties. The objective of the present paper is to present the method in a general context, offer some theorems on the relation between the original and decomposed problem, and present some simple illustrative examples.

2. Intuitive basis

There are two concepts involved here. The first is arranging subtasks, with assigned goals and constraints, that can be pursued independently and concurrently. The second concept is, from practical necessity, using surrogate functions to represent the performance of all the subtasks. Before presenting a mathematical formulation, it may help to provide two real-life scenarios that illustrate the two basic concepts of the decomposition approach. The decomposition approach is similar to the fairly common standard approach of a manager assigning budget and performance requirements to each department or team that reports to her. The decomposition procedure may be viewed as a negotiation session between the manager and the team or department leaders, where she presents each with sets of performance requirements and budgets. For each set the team leader provides a measure of how well he can satisfy the requirements and the budget. This allows the manager to intelligently divide resources between teams.

The second scenario involves designing a stereo system from catalogs that include possible components such as amplifiers and tuners. Supposedly, each item in the catalog represents an optimization by its designers to maximize performance and minimize cost for a different set of overall requirements (such as frequency ranges or power levels). A response surface approximation to all the items in the catalog may be used to approximate the system design as a continuous problem, thus helping the system designer when the number of possible choices is very large.

The decomposition method described in the next section is based on these two concepts. First, each subsystem designer is provided with a budget and a set of performance constraints, and his task is to maximize the margin of safety for satisfying both. Second, each subsystem is optimized in advance for many possible budget and performance requirements, and the results are catalogued in the form of response surface approximations.

This paper is the result of collaboration between an engineer and a mathematician. Therefore, mathematical theorems and proofs are accompanied by notes intended to provide engineering intuition into some aspects of the basis of these theorems and proofs. In particular, a key concept in this intuitive understanding is that two distant locally optimal solutions of the original problem may become close in the decomposed problem. In the context of the management problem discussed above, this represents the viewpoint of the manager. A department may use its budget to achieve exactly or approximately the same goals or outputs in several very different ways. However, the manager, who sees only the results, considers these solutions to be identical or close.

The management analogy also illustrates one important advantage of the decomposition over a single level solution, due to the fact that both the system level problem and the subsystem level problems are of lower dimensions than the single level problem. A manager would be interested in a global rather than a local solution to the problem. By delegating part of the global optimization to each department, both the manager and the departments seek this global optimum in a lower dimensional space than if the manager micromanaged the entire search for the optimum.

3. Theory for quasiseparable subsystems

Denote real n-dimensional Euclidean space by E^n, let g_j denote the jth component of a vector g, and let $B(x, \delta) \subset E^n$ denote the open ball of radius δ centered at x. At some point $\bar{x}$, the *constraint margin* for a (scalar) constraint $g(x) \le 0$ is the amount $-\mu$ by which the constraint is slack: $g(\bar{x}) - \mu = 0$. The margin $-\mu$ for a vector constraint $G(x) \le 0$ is the smallest component margin: $\max_j G_j(\bar{x}) - \mu = 0$. Writing the (positive at feasible points) constraint margin as $-\mu$ rather than μ leads to more elegant notation later.

Consider first a trivial special case where the objective function $f(s)$ depends on the global system design variables s but not on any of the subsystem local design variables $\ell = \big(\ell^{(1)}, \ldots, \ell^{(N)}\big)$. Such a problem has the form

$$\min_{s,\ell} f(s) \qquad \text{subject to} \qquad \begin{array}{c} g^{(0)}(s) \le 0, \\ g^{(i)}\big(s, \ell^{(i)}\big) \le 0,\ i = 1, \ldots, N. \end{array} \tag{ISV}$$

(The name ISV derives from the objective function, which is *independent* of *subsystem variables.*) Here the system variables $s \in E^{n_0}$, the local subsystem variables $\ell^{(i)} \in E^{n_i}$ for $i = 1, \ldots, N$, and to simplify the discussion assume that $f(s)$, $g^{(0)}(s)$, $g^{(1)}\big(s, \ell^{(1)}\big), \ldots, g^{(N)}\big(s, \ell^{(N)}\big)$ are C^2 real vector-valued functions of dimensions 1, m_0, $m_1, \ldots, m_N$, respectively. In this trivial case, there is no optimization done at the subsystem level, since each subsystem consists merely of constraints $g^{(i)}\big(s, \ell^{(i)}\big) \le 0$. Thinking in terms of global system design optimization, the system optimization should have as much leeway as possible, so the subsystem tasks should not be merely to find feasible $\ell^{(i)}$, but to maximize the constraint margins $-\mu_i$. This idea leads to the decomposed (upper level) problem

$$\min_{s} f(s) \qquad \text{subject to} \qquad \begin{array}{c} g^{(0)}(s) \le 0, \\ \mu_i(s) \le 0,\ i = 1, \ldots, N, \end{array} \tag{DISV}$$

where $\mu_i(s)$ is the (global) solution to the ith lower level problem, for $i = 1, \ldots, N$, given by

$$\min_{\ell^{(i)}} \mu_i \qquad \text{subject to} \qquad \max_{1 \le j \le m_i} g_j^{(i)}\big(s, \ell^{(i)}\big) - \mu_i \le 0. \tag{DISV}$$

Denoting the feasible set for (ISV) by X, the feasible set for (DISV) is

$$X_D = \{ s \mid (s, \ell) \in X \text{ for some } \ell \},$$

the projection of X onto the system design space. Note that distant feasible points $(s, \ell) \in X$ may project to the same feasible point $s \in X_D$, and that a disconnected open set in X may project to a connected open set in X_D. The fact that projections from X to X_D do not increase distances between points is important because it guarantees that the decomposition (DISV) does not introduce spurious local solutions. This is formalized in the next lemma.

LEMMA 1. If s^* is a local solution for (DISV), then there exists ℓ^* such that (s^*, ℓ^*) is a local solution for (ISV).

Proof. s^* being a local solution means $\exists\ \epsilon > 0$ such that $s \in X_D \cap B(s^*, \epsilon) \implies f(s^*) \le f(s)$. Also $s^* \in X_D \implies \exists\ \ell^*$ such that $(s^*, \ell^*) \in X$. Since clearly $\|s - s^*\|_2 \le \|(s, \ell) - (s^*, \ell^*)\|_2$, it follows that $(s,\ell) \in X \cap B\big((s^*,\ell^*),\epsilon\big) \implies s \in X_D \cap B(s^*,\epsilon) \implies f(s^*) \le f(s) \implies (s^*,\ell^*)$ is a local solution of (ISV). Q. E. D.

Since close feasible points for the upper level of (DISV) can correspond to feasible points for the original (ISV) that are far apart, Lemma 1 only establishes a correspondence in one direction, which is the best that can be done without further assumptions about the nature of the set of locally optimal solutions of (ISV). Correspondence in the other direction requires that multiple solutions of (ISV) be connected in some sense. This is guaranteed if X is convex and $f(s)$ is strictly quasiconvex at every local solution in X_D. (Quasiconvexity roughly means that $f(s)$ is unimodal in any direction over the feasible set). The theorem below establishes the correspondence in both directions, and then an example shows that the hypothesis is necessary.

Theorem 1. Let X be convex and $f(s)$ be strictly quasiconvex at $s^* \in X_D$. Then s^* is a local solution for (DISV) $\iff$ there exists ℓ^* such that (s^*,ℓ^*) is a local solution for (ISV).

Proof. ($\Longrightarrow$). This direction follows from Lemma 1.

($\Longleftarrow$). (s^*,ℓ^*) being a local solution of (ISV) means $(s^*,\ell^*) \in X$ and $\exists\ \epsilon > 0$ such that $(s,\ell) \in X \cap B\big((s^*,\ell^*),\epsilon\big) \implies f(s^*) \le f(s)$. Suppose s^* is not a local solution for (DISV), so $\exists\ (\hat{s},\hat{\ell}) \in X$ with $\|\hat{s}-s^*\|_2 < \epsilon$ and $f(\hat{s}) < f(s^*)$. By the convexity of X, there exists $0 < \bar{\lambda} < 1$ such that $(\bar{s},\bar{\ell}) = (1-\bar{\lambda})(s^*,\ell^*) + \bar{\lambda}(\hat{s},\hat{\ell}) \in X \cap B\big((s^*,\ell^*),\epsilon\big)$. Since f is strictly quasiconvex at s^*, $f(\hat{s}) < f(s^*) \implies f(\bar{s}) = f\big((1-\bar{\lambda})s^* + \bar{\lambda}\hat{s}\big) < f(s^*)$. This contradicts (s^*,ℓ^*) being a local solution. Therefore $f(s^*) \le f(s)$ for $s \in X_D \cap B(s^*,\epsilon)$, and s^* is a local solution for (DISV). Q. E. D.

Indeed, even for the trivial case (ISV), local minima of the original problem (ISV) may not correspond to local minima of the decomposed problem (DISV) without assumptions on X and f. The following example (for which X is nonconvex and f is strictly convex) illustrates this.

Example 1. Take $f(s) = (s-2)^2$ with constraints $0 \le s \le 2$, $0 \le \ell \le 2$, $(1-s-\ell)(3-s-\ell) \ge 0$. There are local minima at $(s,\ell) = (1,0)$ and $(2,1)$, but the decomposed problem (DISV) has a unique local minimum at $s = 2$. The constraint margin function is given by

$$\mu_1(s) = \begin{cases} \dfrac{-5+2s+\sqrt{13-4s}}{2}, & 0 \le s \le 1; \\[2ex] \dfrac{-1-2s+\sqrt{5+4s}}{2}, & 1 \le s \le 2. \end{cases}$$

To understand the derivation of $\mu_1(s)$, observe first that X is the union of the two disconnected triangles with vertices at $(0,0)$, $(1,0)$, $(0,1)$ and $(1,2)$, $(2,1)$, $(2,2)$. Assume that $0 \le s < 1$, the other range for s is handled similarly. Then necessarily $0 \le \ell \le 1$ and the optimal margin $-\mu$ will exactly balance the two constraints: $-\ell = (s+\ell-1)(3-s-\ell) \le \mu$. Solving this equation for ℓ gives the top expression above for $\mu_1(s) = -\ell$.

Global minima of (ISV) and (DISV) do correspond without convexity assumptions (easily seen from the proofs of Lemma 1 and Theorem 1).

Corollary 1. s^* is a global solution for (DISV) $\iff \exists\ \ell^*$ such that (s^*,ℓ^*) is a global solution for (ISV).

Consider next the more typical situation where the subsystem local design variables are involved in the system objective function. Denote the system variables by $s \in E^{n_0}$, the local subsystem variables by $\ell^{(i)} \in E^{n_i}$ for $i = 1, \ldots, N$, and the total number of variables by

$$n = n_0 + n_1 + \cdots + n_N.$$

Let $\ell = \left(\ell^{(1)}, \ldots, \ell^{(N)}\right)$ denote all the local variables, let $f_0(s)$, $f_1\left(s, \ell^{(1)}\right)$, $\ldots$, $f_N\left(s, \ell^{(N)}\right)$ be C^2 real-valued functions, and let $g^{(0)}(s)$, $g^{(1)}\left(s, \ell^{(1)}\right)$, $\ldots$, $g^{(N)}\left(s, \ell^{(N)}\right)$ be C^2 real vector-valued functions of dimensions m_0, m_1, $\ldots$, m_N, respectively.

The quasiseparable subsystem problem is

$$\min_{s,\ell} f_0(s) + \sum_{i=1}^{N} f_i\left(s, \ell^{(i)}\right) \qquad \text{subject to} \qquad \begin{array}{c} g^{(0)}(s) \leq 0, \\ g^{(i)}\left(s, \ell^{(i)}\right) \leq 0, \ i = 1, \ldots, N. \end{array} \tag{SSP}$$

Precisely, a nonlinear programming problem has *quasiseparable subsystems* (or is *quasiseparable*) if it has the form (SSP). The objective function in (SSP) is partially separable, but quasiseparable implies more, namely a subsystem structure within the constraints, and a special relationship between the subsystems and the objective function. The essential ideas are to introduce a budget $b_i \in E$ for each subsystem criterion $f_i\left(s, \ell^{(i)}\right)$, and to attack the subsystem constraints $g^{(i)}\left(s, \ell^{(i)}\right) \leq 0$ by maximizing the constraint margin $-\mu_i$ for each subsystem. The original problem (SSP) is decomposed into an upper level problem and N lower level problems, which are always feasible and are totally independent. The decomposed problem is (upper level)

$$\min_{s,b} f_0(s) + \sum_{i=1}^{N} b_i \qquad \text{subject to} \qquad \begin{array}{c} g^{(0)}(s) \leq 0, \\ \mu_i(s, b) \leq 0, \ i = 1, \ldots, N, \end{array} \tag{DSSP}$$

where $\mu_i(s, b)$ is the (global) solution to the ith lower level problem, for $i = 1, \ldots, N$, given by

$$\min_{\ell^{(i)}} \mu_i \qquad \text{subject to} \qquad \begin{array}{c} \max\limits_{1 \leq j \leq m_i} g_j^{(i)}\left(s, \ell^{(i)}\right) - \mu_i \leq 0, \\ f_i\left(s, \ell^{(i)}\right) - b_i - \mu_i \leq 0. \end{array} \tag{DSSP}$$

Let X and X_D denote the feasible sets for (SSP) and (DSSP), respectively. Let $\theta(s, \ell)$ be the objective function of (SSP), and $\theta_D(s, b)$ the upper level objective function of (DSSP). The following lemma is straightforward to verify.

Lemma 2. If (s, ℓ) is feasible for (SSP), then

$$(s, b) = (s, b_1, \ldots, b_N) = \left(s, f_1\left(s, \ell^{(1)}\right), \ldots, f_N\left(s, \ell^{(N)}\right)\right)$$

is feasible for (DSSP).

It is also easy to check that at an optimal solution of (DSSP) there is no margin on the budget constraints, so that $b_i = f_i\left(s, \ell^{(i)}\right)$. In the simpler (ISV) case, points that were close to locally optimal solutions for (ISV) corresponded to even closer points for the upper level of (DISV) because the ℓ components were dropped. This guaranteed that the decomposition did not introduce spurious locally optimal solutions. Now, with the b variables, this is not as trivial. Here the continuity of the f_i ensures that points that are close to locally optimal solutions of (SSP) correspond to nearby points in the upper level of (DSSP). The following lemma implies that every locally optimal solution of (DSSP) corresponds to a locally optimal solution of (SSP).

LEMMA 3. If $(s^*, b^*) = \big(s^*,\ f_1(s^*, \ell^{(1)*}),\ \ldots,\ f_N(s^*, \ell^{(N)*})\big) \in X_D$ is a local solution for (DSSP), then $(s^*, \ell^*) \in X$ is a local solution for (SSP).

Proof. The claim is that there exist $\delta > 0$, $\epsilon > 0$ such that

$$\theta_D(s^*, b^*) \le \theta_D(s, b) \text{ for } (s, b) \in X_D \cap B\big((s^*, b^*), \delta\big) \Longrightarrow$$
$$\theta(s^*, \ell^*) \le \theta(s, \ell) \text{ for } (s, \ell) \in X \cap B\big((s^*, \ell^*), \epsilon\big).$$

This follows directly from the continuity of the functions f_i and the observation that, by Lemma 2, the functions $f_1, \ldots, f_N$ map feasible points (s, ℓ) in a neighborhood of (s^*, ℓ^*) to feasible points (s, b) in a neighborhood of (s^*, b^*). Q. E. D.

The reverse direction still requires convexity assumptions similar to those for (ISV).

THEOREM 2. Let X be convex and $\theta(s, \ell)$ be strictly quasiconvex at $(s^*, \ell^*) \in X$. Then $(s^*, \ell^*) \in X$ is a local solution for (SSP) $\iff$ $(s^*, b^*) = \big(s^*,\ f_1(s^*, \ell^{(1)*}),\ \ldots,\ f_N(s^*, \ell^{(N)*})\big) \in X_D$ is a local solution for (DSSP).

Proof. ($\Longleftarrow$). This direction follows from Lemma 3.

($\Longrightarrow$). It is required to show there exist $\delta > 0$, $\epsilon > 0$ such that

$$\theta(s^*, \ell^*) \le \theta(s, \ell) \text{ for } (s, \ell) \in X \cap B\big((s^*, \ell^*), \delta\big) \Longrightarrow$$
$$\theta_D(s^*, b^*) \le \theta_D(s, b) \text{ for } (s, b) \in X_D \cap B\big((s^*, b^*), \epsilon\big).$$

Suppose $\exists\ (\hat{s}, \hat{b}) \in X_D \cap B\big((s^*, b^*), \epsilon\big)$ with $\theta_D(\hat{s}, \hat{b}) < \theta_D(s^*, b^*)$. Then there is also a feasible point $(\hat{s}, \hat{\ell}) \in X$ with $\theta(\hat{s}, \hat{\ell}) \le \theta_D(\hat{s}, \hat{b}) < \theta_D(s^*, b^*) = \theta(s^*, \ell^*)$. By the convexity of X, there exists $0 < \bar{\lambda} < 1$ such that $(\bar{s}, \bar{\ell}) = (1 - \bar{\lambda})(s^*, \ell^*) + \bar{\lambda}(\hat{s}, \hat{\ell}) \in X \cap B\big((s^*, \ell^*), \delta\big)$. Since θ is strictly quasiconvex at (s^*, ℓ^*), $\theta(\hat{s}, \hat{\ell}) < \theta(s^*, \ell^*) \Longrightarrow \theta(\bar{s}, \bar{\ell}) = \theta\big((1 - \bar{\lambda})(s^*, \ell^*) + \bar{\lambda}(\hat{s}, \hat{\ell})\big) < \theta(s^*, \ell^*)$. This contradicts (s^*, ℓ^*) being a local solution. Therefore $\theta_D(s^*, b^*) \le \theta_D(s, b)$ for $(s, b) \in X_D \cap B\big((s^*, b^*), \epsilon\big)$, and (s^*, b^*) is a local solution for (DSSP). Q. E. D.

An alternative to the convexity assumptions is given in the next technical corollary, which is primarily of theoretical mathematical interest.

COROLLARY 2. Suppose that (s, ℓ) is feasible for (SSP) $\iff$

$$(s, b) = (s, b_1, \ldots, b_N) = \Big(s, f_1\big(s, \ell^{(1)}\big), \ldots, f_N\big(s, \ell^{(N)}\big)\Big)$$

is feasible for (DSSP), and that $D_{\ell^{(i)}} f_i(s^*, \ell^{(i)*}) \ne 0$ for $i = 1, \ldots, N$. Then $(s^*, \ell^*) \in X$ is a local solution for (SSP) $\iff$ $(s^*, b^*) = \big(s^*,\ f_1(s^*, \ell^{(1)*}),\ \ldots,\ f_N(s^*, \ell^{(N)*})\big) \in X_D$ is a local solution for (DSSP).

Proof. By Lemma 3, it suffices to show that there exist $\delta > 0$, $\epsilon > 0$ such that

$$\theta(s^*, \ell^*) \leq \theta(s, \ell) \text{ for } (s, \ell) \in X \cap B\big((s^*, \ell^*), \delta\big) \Longrightarrow$$
$$\theta_D(s^*, b^*) \leq \theta_D(s, b) \text{ for } (s, b) \in X_D \cap B\big((s^*, b^*), \epsilon\big).$$

By hypothesis the correspondence $\big(s, \ell^{(i)}\big) \leftrightarrow (s, b_i) = \big(s, f_i\big(s, \ell^{(i)}\big)\big)$ preserves feasibility, and so by continuity the functions $f_1, \ldots, f_N$ map feasible points (s, ℓ) in a neighborhood of (s^*, ℓ^*) to feasible points (s, b) in a neighborhood of (s^*, b^*). It therefore suffices to show that $f_1, \ldots, f_N$ map an (open) neighborhood of (s^*, ℓ^*) *onto* a neighborhood $B\big((s^*, b^*), \epsilon\big)$. That the map $(s, \ell) \to (s, b)$ is onto in a neighborhood $B\big((s^*, b^*), \epsilon\big)$ follows from the assumption $D_{\ell^{(i)}} f_i\big(s^*, \ell^{(i)*}\big) \neq 0$ and the Inverse Function Theorem. Q. E. D.

Similar to Example 1, a counterexample shows that the hypotheses of Theorem 2 are required.

EXAMPLE 2. Consider the strictly quasiconvex objective function $\theta(s, \ell) = s + .1\ell$ with nonconvex feasible set $X = \big\{(s, \ell) \mid s \geq 16\ell^2(1-\ell)^2,\ 0 \leq \ell \leq .75\big\}$. The problem (SSP) has two local minima, at $(s, \ell) = (0, 0)$ and $(9/16, 3/4)$, but (DSSP), with $\theta_D(s, b) = s + b$ and $X_D = \big\{(s, b) \mid \mu_1(s, b) \leq 0\big\} = \{(s, b) \mid (s, b) \geq 0\}$, has a unique local minimum at $(s, b) = (0, 0)$. The constraint margin function $\mu_1(s, b)$ is complicated, but simply observe that for any $(s, b) > 0$, $\ell = 0$ is feasible for the lower level problem of (DSSP), and $\mu_1(s, b) < 0$. Therefore every $(s, b) > 0$ is an interior feasible point for (DSSP), and the only local minimum occurs at (0,0).

Just as Corollary 1 followed from the proofs of Lemma 1 and Theorem 1, the proofs of Lemma 3 and Theorem 2 immediately yield the following analog of Corollary 1.

COROLLARY 3. $(s^*, b^*) = \Big(s^*,\ f_1\big(s^*, \ell^{(1)*}\big),\ \ldots,\ f_N\big(s^*, \ell^{(N)*}\big)\Big) \in X_D$ is a global solution for (DSSP) $\Longleftrightarrow$ $(s^*, \ell^*) \in X$ is a global solution for (SSP).

4. Constraint margin function approximations

While all of the preceding theory is valid regardless of the relative dimensions of s, b, and the $\ell^{(i)}$, the context here is that of multidisciplinary design optimization, where often $n_0 + N \ll n_1 + \cdots + n_N$, with n_0 perhaps as small as 10 or 20. An example of how such a small n_0 might be achieved is in Liu et al. (2003). For the approximations to $\mu_i(s, b)$ (note that only the one component b_i of b is used for μ_i) described here to be practical, it is essential that the size n_0 of s be small.

The constraint margin functions $\mu_i(s, b)$ are generally nonsmooth (e.g., $\mu_1(s)$ in Example 1 is a C^0 piecewise analytic function over $[0, 2]$ with a breakpoint at $s = 1$), and will typically be known only pointwise rather than analytically. Constructing (response surface) approximations $\tilde{\mu}_i$ to μ_i thus constitutes a plausible strategy. Replacing μ_i by $\tilde{\mu}_i$ in the upper level problem (DSSP) gives the approximate problem

$$\min_{s,b} f_0(s) + \sum_{i=1}^{N} b_i \qquad \text{subject to} \qquad \begin{aligned} &g^{(0)}(s) \leq 0, \\ &\tilde{\mu}_i(s, b) \leq 0,\ i = 1, \ldots, N. \end{aligned} \tag{ADSSP}$$

Let X_D and $\tilde{X}_D$ denote the feasible sets for (DSSP) and (ADSSP), respectively. Note that $\tilde{\mu}_i$ need not really approximate μ_i well everywhere—when $\mu_i \ll 0$ or $\mu_i \gg 0$, $\tilde{\mu}_i$ must only behave similarly.

Only near the feasible set "boundaries" must $\tilde{\mu}_i$ accurately approximate μ_i. Precisely, let $\tau > 0$ and $\xi > 0$ be fixed tolerances. Then for points $(s, b) \in X_D \cup \tilde{X}_D$, each $\tilde{\mu}_i$, for $i = 1, \ldots, N$, must satisfy

$$\begin{cases} \mu_i(s,b)\tilde{\mu}_i(s,b) > 0, & \text{if } |\mu_i(s,b)| > \tau; \\ |\tilde{\mu}_i(s,b) - \mu_i(s,b)| < \xi, & \text{if } |\mu_i(s,b)| \le \tau. \end{cases} \tag{1}$$

The following theorem is then easily verified.

THEOREM 3. For $i = 1, \ldots, N$, let the approximations $\tilde{\mu}_i(s, b)$ satisfy (1), and let

$$(s^*, b^*) = \Big(s^*, f_1\big(s^*, \ell^{(1)*}\big), \ldots, f_N\big(s^*, \ell^{(N)*}\big)\Big)$$

be a local solution for (ADSSP). Then (s^*, ℓ^*) violates the constraints of (SSP) by at most ξ: $g^{(0)}(s^*) \le 0$, $\max_j g_j^{(i)}(s^*, \ell^*) \le \xi$ for $i = 1, \ldots, N$.

NOTE 1. If a lower level problem of (DSSP) has a feasible point with $\mu_i(s, b) \le 0$, then that fact must be communicated to the upper level problem of (DSSP). That is, it is not sufficient to find a local minimum $\mu_i(\bar{s}, \bar{b}) > 0$ when another local minimum $\mu_i(\hat{s}, \hat{b}) \le 0$ exists. This may require a global search to determine whether the global minimum $\mu_i(s^*, b^*) \le 0$. All of the preceding theory assumes that $\mu_i(s, b)$ is the global minimum for the ith lower level problem, for which a global search is far preferable to a global search for the higher dimensional original problem (SSP). Computationally, to find a step $(\Delta s, \Delta b)$ improving $\theta_D(s, b)$, it is acceptable to report a local minimum $\mu_i(s, b) < 0$ to the upper level problem. Typically the constraint margins $-\mu_i$ will be driven to zero, thus requiring global optimizations at the lower level.

NOTE 2. The subsystem constraints written earlier as $g^{(i)}\big(s, \ell^{(i)}\big) \le 0$ typically are of the form

$$\begin{pmatrix} \hat{g}^{(i)}\big(\ell^{(i)}\big) \\ \check{g}^{(i)}\big(s, \ell^{(i)}\big) \end{pmatrix} \le 0,$$

where $\hat{g}^{(i)}$ includes simple bounds on the local variables $\ell^{(i)}$. There is no need to put a margin on these purely local constraints, so the constraints of the (DSSP) lower level problem should look like

$$\begin{aligned} \hat{g}^{(i)}\big(\ell^{(i)}\big) &\le 0, \\ \max_j \check{g}_j^{(i)}\big(s, \ell^{(i)}\big) - \mu_i &\le 0, \\ f_i\big(s, \ell^{(i)}\big) - b_i - \mu_i &\le 0. \end{aligned}$$

This considerably simplifies the margin function $-\mu_i(s, b)$, and assuming that all the purely local constraints $\hat{g}^{(i)}\big(\ell^{(i)}\big) \le 0$ can be satisfied (so $\mu_i(s, b)$ is always defined), has no effect on the theory in this section. In practice the local subsystem constraints $g^{(i)}$ will always be separated into $\hat{g}^{(i)}$ and $\check{g}^{(i)}$, and a margin only applied to the constraints involving the system variables s.

5. Example

The following example is artificial, but illustrates how effective the decomposition can be in eliminating local minima from the original formulation. Consider the objective function

$$\theta(s,\ell) = \sum_{i=1}^{10} s_i + .1\ell_i$$

subject to the constraints

$$s_i \geq 16\ell_i^2(1-\ell_i)^2, \quad 0 \leq \ell_i \leq .75, \quad i = 1,\ldots,10.$$

There are 2^{10} local minima at points $(s,\ell) \in E^{20}$, where each pair (s_i,ℓ_i) is either $(0,0)$ or $(9/16,3/4)$. By an argument similar to that used in Example 2, the decomposed problem (DSSP) with $\theta_D(s,b) = \sum_{i=1}^{10} s_i + b_i$ has a single local minimum at $(s,b) = (0,0)$.

A realistic, large-scale engineering example using response surface approximations for the lower level problem optima is in Liu et al. (2003). Constructing a high dimensional approximation to a subproblem optima response surface is (for typical large-scale engineering problems) highly nontrivial, and the technical difficulty of such response surface approximation construction dominates both the computation time and the applicability of the quasiseparable decomposition strategy proposed here.

References

N. M. Alexandrov, "Multilevel methods for MDO", in *Multidisciplinary Design Optimization*, N. M. Alexandrov and M. Y. Hussaini (eds.), SIAM, Philadelphia, PA, 1997, pp. 79–89.

V. O. Balabanov, A. A. Giunta, O. Golovidov, B. Grossman, W. H. Mason, L. T. Watson, and R. T. Haftka, "Reasonable design space approach to response surface approximation", *J. Aircraft* vol. 36, pp. 308–315, 1999.

A.A. Giunta, V. Balabanov, D. Haim, B. Grossman, W.H. Mason, L.T. Watson, and R.T. Haftka, "Multidisciplinary optimisation of a supersonic transport using design of experiments theory and response surface modelling", *Aeronautical J.* vol. 101, no. 1008, pp. 347–356, 1997a.

A. A. Giunta, V. Balabanov, M. Kaufman, S. Burgee, B. Grossman, R. T. Haftka, W. H. Mason, and L. T. Watson, "Variable-complexity response surface design of an HSCT configuration", in *Multidisciplinary Design Optimization*, N. M. Alexandrov and M. Y. Hussaini (eds.), SIAM, Philadelphia, PA, 1997b, pp. 348–367.

S. Hosder, L. T. Watson, B. Grossman, W. H. Mason, H. Kim, R. T. Haftka, and S. Cox, "Polynomial response surface approximations for the multidisciplinary design optimization of a high speed civil transport", *Optim. Engrg.* vol. 2, pp. 431–452, 2001.

I. Kroo, "MDO for large-scale design", in *Multidisciplinary Design Optimization*, N. M. Alexandrov and M. Y. Hussaini (eds.), SIAM, Philadelphia, PA, 1997, pp. 22–44.

B. Liu and R.T. Haftka, "Single-level composite wing optimization based on flexural lamination parameters", *Structural Multidisciplinary Optim.* vol. 26, nos. 1–2, pp. 111–120, 2004.

B. Liu, R.T. Haftka, and M.A. Akgün, "Two-level composite wing structural optimization using response surfaces", *Structural Optim.* vol. 20, no. 2, pp. 87–96, 2000.

B. Liu, R. T. Haftka, and L. T. Watson, "Global-local structural optimization using response surfaces of local optimization margins", *Structural Multidisciplinary Optim.* vol. 27, no. 5, pp. 352–359, 2004.

S.A. Ragon, Z. Gürdal, R.T. Haftka, and T.J. Tzong, "Global/local structural wing design using response surface techniques", in *38th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conf.*, AIAA Paper 97-1051, Kissimmee, FL, April 7–10, 1997, Part 2, pp. 1204–1214..

J. F. Rodríguez, J. E. Renaud, and L. T. Watson, "Trust region augmented Lagrangian methods for sequential response surface approximation and optimization", *ASME J. Mech. Design* vol. 120, pp. 58–66, 1998a.

J. F. Rodríguez, J. E. Renaud, and L. T. Watson, "Convergence of trust region augmented Lagrangian methods using variable fidelity approximation data", *Structural Optim.* vol. 15, pp. 141–156, 1998b.

J. Shankar, C. J. Ribbens, R. T. Haftka, and L. T. Watson, "Computational study of a nonhierarchical decomposition algorithm", *Comput. Optim. Appl.* vol. 2, pp. 273–293, 1993.