



25293579

LIBRARY
Michigan State
University

This is to certify that the
thesis entitled
**ON DESIGN OPTIMIZATION BY
NON-HIERARCHICAL DECOMPOSITION**

presented by

JIAN PAN

has been accepted towards fulfillment
of the requirements for

MS degree in **MECHANICAL ENGINEERING**

 **A. DÍAZ**
Major professor

Date **JULY 20, 1989**

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

MSU Is An Affirmative Action/Equal Opportunity Institution

**ON DESIGN OPTIMIZATION BY
NON-HIERARCHICAL DECOMPOSITION**

BY

Jian Pan

A THESIS

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

MASTER OF SCIENCE

Department of Mechanical Engineering

1989

ABSTRACT**ON DESIGN OPTIMIZATION BY
NON-HIERARCHICAL DECOMPOSITION****BY****Jian Pan**

An algorithm for the optimal design of non-hierarchical systems by decomposition is developed based on rigorous analytical results. In the algorithm, a global problem is decomposed into smaller subproblems. The same objective and constraints are present in all subproblems and the solution of the global problem is obtained by solving each subproblem in sequence. The concept of pseudo optimal point is introduced to signify the fact that a point that satisfies the optimality conditions of all subproblems is not optimal for the global problem. A strategy to move away from pseudo optimal points while improving the global objective is described. In the presence of redundancy, the algorithm proceeds as usual after removal of all redundant constraints in subproblem optimization. Implementation issues are discussed and numerical examples are given for illustration.

To my parents

ACKNOWLEDGMENTS

I feel indebted to Professor Alejandro R. Diaz. His patient guidance and illuminating discussion are crucial to the completion of the thesis. His help is beyond academy. He has provided all the convenience for typing and printing this manuscript.

Special thanks to Professor Joseph Whitesell and Professor Mukesh V. Gandhi for serving in my Graduating Committee and reviewing this manuscript.

I would like to express my deep thank to the Li Foundation, Inc. of New York for providing me a two year scholarship. This work is also partially supported by the National Science Foundation and its support is gratefully appreciated.

Last, but not the least, I would like to thank Dr. Keshun Liu for offering me his invaluable friendship.

TABLE OF CONTENTS

LIST OF FIGURES	vii
INTRODUCTION.....	1
LITERATURE REVIEW.....	9
2.1 Cumulative Constraint Approach in Decomposition of Hierarchical Systems.....	11
2.2 Penalty Approach in Decomposition of Hierarchical Systems	14
2.3 Strategies in Decomposition of Non-hierarchical Systems	16
BACKGROUND.....	18
NON-HIERARCHICAL SYSTEM DECOMPOSITION	22
4.1 Problem Statement.....	22
4.2 Derivation of the Algorithm.....	26
4.3 Redundancy.....	35
4.4 A Procedure to Identify the Set of Redundant Constraints	48
4.5 Infeasible Starting Points.....	50

4.6 Convex Approximations of Non-local State Variables	51
4.7 Use of Optimum Sensitivity Derivatives.	53
4.8 Summary of Chapter 4.	54
NUMERICAL EXAMPLES.	56
5.1 Example 1.	56
5.2 Example 2.	59
5.3 Example 3.	62
REMARKS AND CONCLUSIONS.	70
APPENDIX.	73
BIBLIOGRAPHY.	77

LIST OF FIGURES

Figure	Page
1 Hierarchical system	2
2 Network system	2
3 A drive and control system for a transfer mechanism	4
4 Example of solution by decomposition	6
5 Moving away from pseudo optimal point	6
6 A three level hierarchical system	11
7 Coupling between subsystems of a hierarchical structure	11
8 A sequential algorithm for non-hierarchical system decomposition . . .	23
9 Existence of a redundant constraint	37
10 Counterexample 1: proposition 4.3.1 is not true	41
11 Counterexample 2: proposition 4.3.1 is not true	42
12 Iteration history for example 1: exact function	58

13	Iteration history for example 1: convex approximations	58
14	Iteration history for example 2: exact function	62
15	Conical clutch	65
16	Iteration history for example 3: exact function	69
17	Iteration history for example 3: convex approximations	69

CHAPTER 1

INTRODUCTION

Decomposition in optimal design refers to a strategy whereby an optimization problem is solved iteratively by making design changes only in a subset of the design space at each step. The optimization problem solved in this fashion is usually called the *decomposed problem*. There are several reasons why one may choose to solve the decomposed problem, rather than the original problem in the full design space. For example, decomposition is necessary when the solution of the problem as a whole is cost-prohibitive or impractical due to the size of the problem, i.e., because of the large dimension of the design space. It is important to realize, however, that size is not the only reason why a decomposition strategy should be considered. Optimal design decomposition also makes it possible to study systems formed by subsystems whose analyses correspond to different engineering disciplines. In the decomposed problem, each subsystem can be analyzed locally, using local expertise and the theoretical and computational methods most suitable to the specific subproblem. For instance, to obtain the optimal design of a complex engineering system, it is advisable to employ a decomposition strategy and consider its modules (e.g., engine, suspension and

structure) separately, instead of solving the problem in the whole design space. It is this feature, rather than the problem size, that motivates the present work.

Two basic forms of decomposition in optimal design have been considered in the literature. In *non-hierarchical* or *network* decomposition, a global system is decomposed into smaller subsystems such that there exists information transmission between any two subsystems. In *hierarchical* or *multilevel* decomposition, a global problem is decomposed into a hierarchy of smaller subsystems and coupling is permitted only between a high level subsystem and its subordinate subsystems at the lower hierarchical level. Graphic representations of the two decomposition models are given Fig. 1 and Fig. 2.

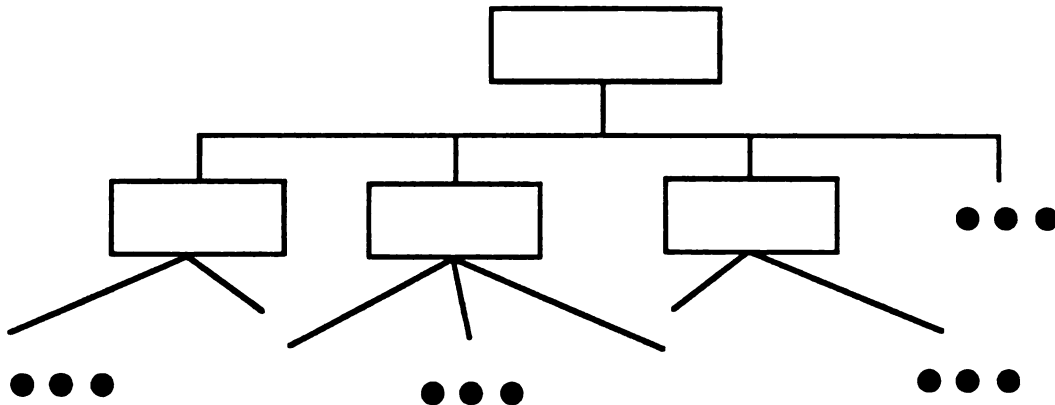


Figure 1. Hierarchical system

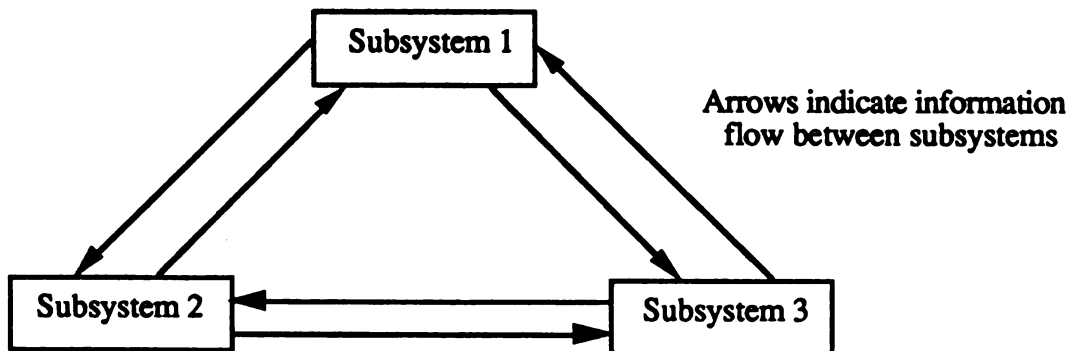


Figure 2. Network system

In Fig. 1 and Fig. 2, a box represents a subsystem which defines an optimization subproblem and its corresponding design subspace. A connection between two boxes describes information transfer to and from the corresponding subproblems. From the figures, it is clear that a hierarchically decomposed problem can be modelled by a tree while the more complex non-hierarchical problem corresponds to a network. In either decomposition strategy, optimizations are performed in subsystems and the solution of the global problem is obtained by combining subsystem optimal results according to some specific decomposition algorithm. This thesis deals with one such algorithm for the optimization of non-hierarchical systems.

The design of many engineering systems can be modelled as a non-hierarchical optimization problem. As an example, consider the optimal design of a drive and control system for a transfer mechanism shown schematically in Fig. 3. An automatic assembly machine consists of two major components: a transfer mechanism to move the work to a prescribed station, and a system to power and control the transfer mechanism. In [W1] and [W2], a crank and connecting rod are used as a transfer mechanism, and an electric motor driving a speed reducer via a clutch is selected to power the crank. The design of the drive and control system as a whole and of each element can be carried out separately. The problems are coupled, however, since clutch and speed reducer must be such that the *overall* system behavior requirements are satisfied, e.g., the minimum motor speed resulting when the clutch is engaged must be greater than the minimum allowable motor speed, and the time required to accelerate the speed reducer input shaft from the time of clutch engagement to normal operating speed of motor must be within bounds. Of all the clutches and speed reducers that satisfy the drive and control system (DC system) design requirements, we seek the detailed designs such that the DC system possesses the minimum weight. It is apparently that optimal design of the DC system can be naturally decomposed into three optimization subproblems:

Problem 1: overall system design

Problem 2: clutch detailed design

Problem 3: speed reducer detailed design

A mathematical model of the example will be formulated later and the problem will be solved by the non-hierarchical optimization algorithm derived in this thesis.

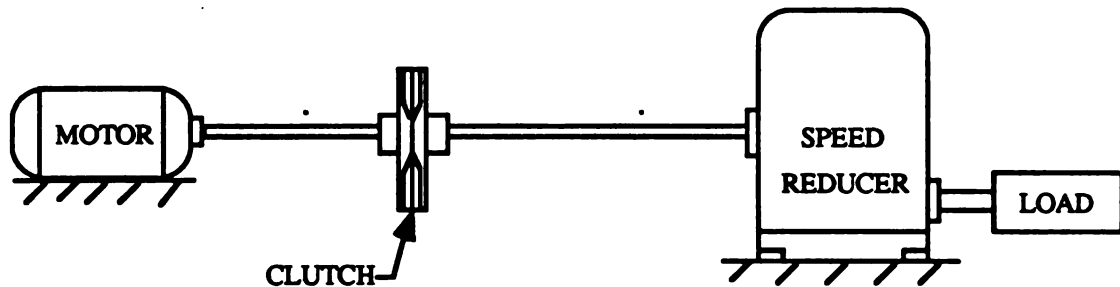


Figure 3. A drive and control system for a transfer mechanism

To illustrate some of the difficulties associated with the decomposition in optimal design, consider the following very simple example:

Problem P:

Find $\mathbf{x} \in \mathbb{R}^2$ that

Minimizes $f(x_1, x_2) = -x_1 - 2x_2$

Subject to $g_1(x_1, x_2) = x_2 - x_1 \leq 0$

$g_2(x_1, x_2) = x_2 + x_1 - 4 \leq 0$

This problem will be decomposed into two subproblems, labeled by P^1 and P^2 , defined as follows:

Problem P¹:

Find $x_1 \in \mathbb{R}^1$ that

Minimizes $f(x_1, x_2^0) = -x_1 - 2x_2^0$

Subject to $g_1(x_1, x_2^0) = x_2^0 - x_1 \leq 0$

$g_2(x_1, x_2^0) = x_2^0 + x_1 - 4 \leq 0$

Problem P²:

Find $x_2 \in \mathbb{R}^1$ that

Minimizes $f(x_1^0, x_2) = -x_1^0 - 2x_2$

Subject to $g_1(x_1^0, x_2) = x_2 - x_1^0 \leq 0$

$g_2(x_1^0, x_2) = x_2 + x_1^0 - 4 \leq 0$

The solution of problem P will be obtained by solving P¹ and P² instead of P. Several strategies to solve P¹ and P² in order to get the solution of problem P are suggested below.

Starting from $(x_1^0, x_2^0) = (1, 0.5)$, suppose that P¹ and P² are solved sequentially using as reference points x_1^0 and x_2^0 , the solution of the previously solved subproblem. If P¹ were solved first, the procedure gets 'stuck' at (3.5, 0.5) (Fig. 4). If P² were solved first, after a sequence P² and P¹, the process would stop at (3, 1). Points (3.5, 0.5) and (3, 1) satisfy the optimality conditions for both P¹ and P² but neither is an optimal solution of P. We call these points *pseudo optimal*. There is an infinite number of such points, and excluding luck, a strategy based on a straight, sequential

solution of the subproblems will get stuck at a pseudo optimal point, and fail to produce the optimal point. Notice also that if P^1 and P^2 were solved simultaneously and their answers combined, the result would be the infeasible point $(3.5, 1)$ (see Fig. 4).

As illustrated above, if all subproblems are solved in sequence, the process will get stuck at a pseudo optimal point. It is necessary to have a strategy to move away from

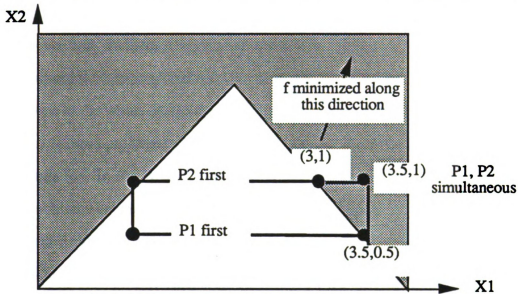


Figure 4. Example of solution by decomposition

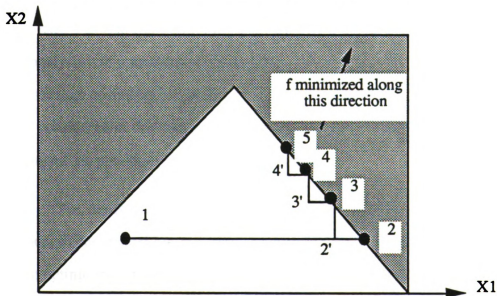


Figure 5. Moving away from pseudo optimal points

pseudo optimal points. A possible strategy may be based on an intermediate step that moves away from one of the active constraints. The search can be restarted from the new point and continued until a new pseudo optimal point is found. When applied to the previous example, the iteration history would look like Fig. 5.

An algorithm for the optimal design of a non-hierarchical system by decomposition is presented in the thesis. The algorithm solves all subproblems in sequence and includes a strategy to move away from pseudo optimal points. Clearly, the procedure cannot compete in speed and efficiency with methods that allow arbitrary motions in the whole design space. However, the intention here is not to increase speed but to address problems that *must* be solved in their decomposed form, with limited information. In the typical problem addressed here, the global problem consists of subproblems corresponding to different engineering disciplines. In each subproblem, the same objective from the global problem is minimized by a group of experts in the particular discipline. Also, only a subset of the global design variables are allowed to change in each subproblem, and only the local state variables can be computed exactly. All subproblems must satisfy the same set of system behavior constraints and coupling may exist between any two subsystems, i.e., each subproblem optimization may require the analysis and optimization results from other subsystems. This is the standard, non-hierarchical optimization problem discussed by Sobieszczanski-Sobieski in [S3] and also considered in this thesis. Strategies for the solution of such problems given later are based on previous results presented in [D1].

The thesis is organized as follows: Chapter 2 gives a literature review on the subject of decomposition in optimal structural design. Chapter 3 presents main results from optimization theory that are essential to the understanding of the development of the thesis. An algorithm to solve non-hierarchical problems by decomposition is

discussed in Chapter 4. Numerical examples are provided in Chapter 5 to illustrate the results derived in the thesis. Remarks and conclusions about the decomposition strategy are included in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

To solve an optimal design problem by decomposition, the problem is first decomposed into smaller subproblems and optimizations are carried out on the subproblems. There are several questions that should be answered in all decomposition strategies to insure subproblem optimal results converge to the global solution:

1. How to maintain a global *feasible design* , at which all constraints of the global problem are satisfied, if the reference point in subproblem optimization is globally feasible.

2. What information can correctly represent the coupling between subproblems and how to use this information to guide subproblem optimization to the global solution.

The *cumulative constraint* concept [S4] and the *penalty method* [H1] are two approaches commonly used in the literature to answer the first question for hierarchical system decomposition. The cumulative constraint is a scalar which represents the satisfaction or violation of a *set* of constraints. A negative value of the cumulative constraint means that all constraints in the set are inactive. Using the cumulative constraint concept, each subsystem at lower hierarchical levels seeks to minimize its

own cumulative constraint, formed by local constraints, while the top level subsystem minimizes the global objective keeping all subsystem cumulative constraints within bounds.

In the penalty method approach, a penalty function for each subproblem is constructed by adding a penalty term formed via constraint violations to the global objective. The penalty function is then minimized in its corresponding subsystem optimization.

To answer the second question, one must first realize that each subsystem optimization is performed by fixing the non-local design variables. Intuitively, subsystem optimizations are closely related to the sensitivity and stability analysis due to parametric variations in nonlinear programming. Thus, optimum sensitivity derivatives and Lagrange multiplier vectors associated with subproblem optimal solutions can be used to represent the coupling among subsystems.

Various decomposition methods and implementations in structural design have appeared in the literature. Much of the work on multilevel optimization is based on exploiting specific structure of the problem. Many engineering systems can be decomposed in the hierarchical fashion shown in Fig. 1. A three level hierarchical system is shown in Fig. 6. At the top level, a simplified optimization model represents the overall behavior of the system. At the lower levels, subproblems optimize detailed representations of system components. An important property of systems of this kind is that no coupling between subsystems at the same hierarchy level exists. A 'parent' problem only requires information from its 'children' subproblems and from its own 'parent'. Optimization methods suited for these problems are outlined in the following sections.

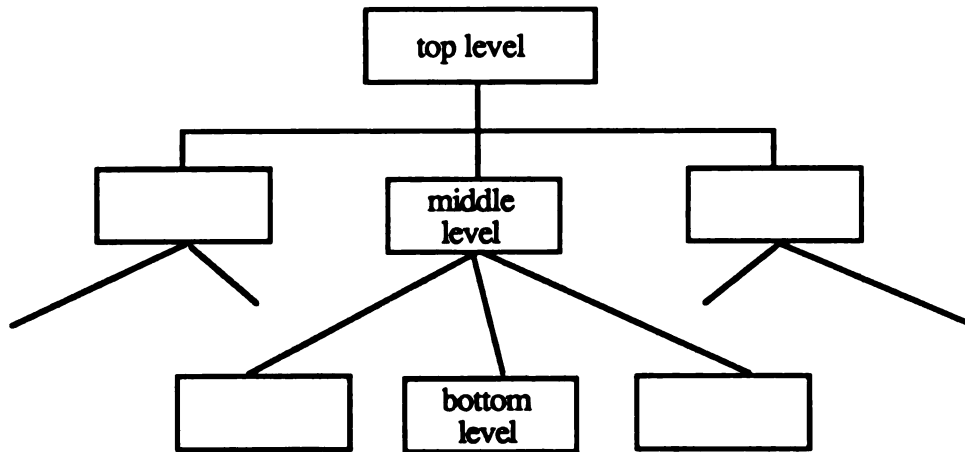


Figure 6. A three level hierarchical system

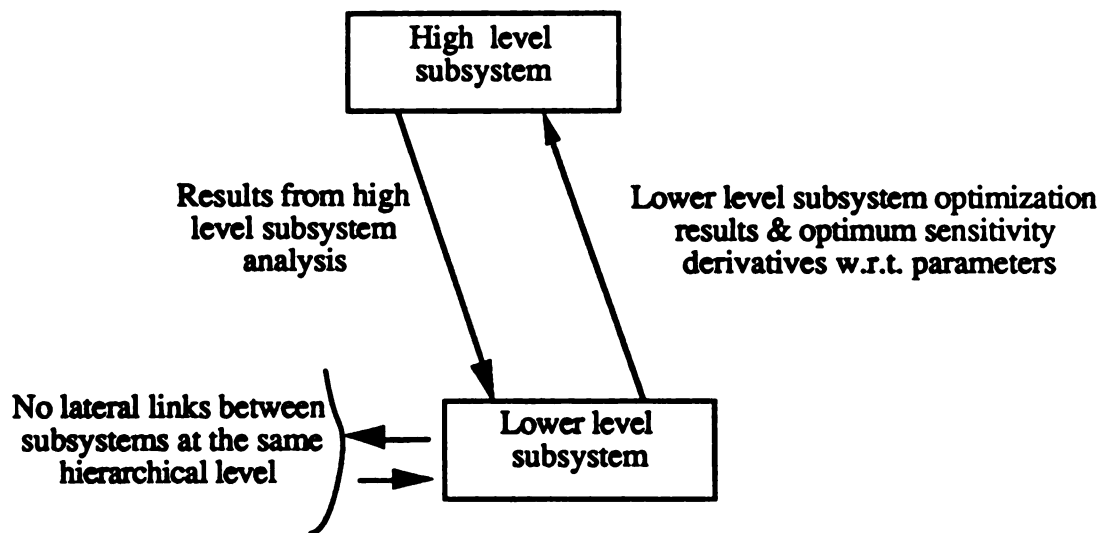


Figure 7. Coupling between subsystems of a hierarchical structure

2.1 Cumulative Constraint Approach in Decomposition of Hierarchical Systems

A linear decomposition method for optimization of hierarchical systems is proposed by J. Sobieszczanski-Sobieski in [S4]. The coupling between subsystems at different hierarchical levels is enforced by the transfer of optimum sensitivity derivative

information. As shown in Fig. 7, the analysis proceeds from top to bottom, so that output from analysis of a higher level subsystem becomes input for analyses of the subordinated subsystems. Since there is no information transmission across subsystems at the same hierarchy level, the subsystems can be analyzed concurrently. After one top-to-bottom cycle of analysis is performed, one optimization cycle proceeds from the bottom up as described below for the three level hierarchical example shown in Fig. 6.

In each subsystem at the bottom level, design variables are local physical quantities and all input received from the middle level subsystem is treated as fixed parameters. The cumulative constraint of the subsystem constraints is the objective function. The cumulative constraint is a single number that measures the degree of satisfaction or violation of an entire set of constraints. The Kresselmeier-Steinhauser (K-S) function is adopted in [S4] as the cumulative constraint. It has the form:

$$\Omega = \frac{1}{\rho} \text{LN} \left(\sum_i (e^{\rho g_i}) \right)$$

where

g_i : subsystem constraint

ρ : user defined coefficient

A negative value of the K-S function means that all subsystem constraints g_i are inactive.

Subsystems in the middle level have their own design variables. Again, input from the top subsystem is considered as fixed parameters. Objective function of each subproblem is the cumulative constraint function for a set of *local* constraints. In addition, a linear approximation of the minimum values of the cumulative constraint

functions transmitted from the bottom level subsystems is added to the set of local constraints.

The top level problem seeks to minimize the global system objective using a set of top level physical quantities as design variables. The top level problem has, as constraints, bounds on first order Taylor's expansions of the minimum values of the cumulative constraint functions of the middle level problems.

The approach outlined above has been successful in several applications reported in [S8] and [W3], although convergence of the procedure is not justified theoretically. To increase the efficiency of the method, J. M. Barthelemy et al [B1] suggest two optimization techniques widely used in single-level optimization: constraint approximation and temporary constraint deletion. The first technique deals with replacing the analytical objective and constraint functions by a sequence of inexpensive convex approximations. The second technique involves reoptimizing only those subproblems that have violated critical or nearly critical constraints.

In the linear decomposition method described above, the minimum value of the cumulative constraint of a subsystem is approximated using a first order Taylor's expansion via optimum sensitivity derivatives. A. R. Parkinson et al [P2] present an alternative strategy that uses low order polynomial approximations of the cumulative constraint functions obtained from a least square fit of actual values at carefully selected points in the design space. The fundamental difference between the decomposition strategy discussed in [P2] and in [S4] is the way in which the minimum values of the cumulative constraints are approximated. The method has been successfully applied to several example problems to demonstrate its usefulness.

2.2 Penalty Approach in Decomposition of Hierarchical Systems

The work of R. T. Haftka [H1] is an example of a decomposition algorithm based on a penalty function approach. The following two-level system discussed in [H1] illustrates the approach. Let y denote the vector of global design variables and x_i , $i=1, \dots, s$ be the vectors of local design variables, where s is the number of subproblems. The global problem is:

Find y and x_i that

Minimizes:

$$\text{Objective function: } f = f_0(y) + \sum_i^s f_i(y, x_i)$$

Subject to:

$$\text{Global constraints: } g_j(y) \geq 0 \quad j=1, \dots, m$$

Subsystem local constraints:

$$g_{ij}(y, x_i) \geq 0 \quad i=1, \dots, s \quad j=1, \dots, n_i$$

where n_i is the number of local constraints for the i -th subsystem.

The solution of the global problem is obtained by a penalty approach solving a series of problems, defined below, for fixed penalty parameters r and r_i such that $r, r_i \rightarrow 0$. This is an extended interior penalty method since the function P below is selected to be an extended interior penalty function. The global problem becomes:

Find y and x_i that

Minimizes:

$$\Phi(y, x_i, r) = f_0(y) + \sum_{i=1}^s f_i(y, x_i) + \sum_{j=1}^m P[g_j(y), r] + \sum_{i=1}^s \sum_{j=1}^{n_i} P[g_{ij}(y, x_i), r_i]$$

The above problem is decomposed into following subproblems:

$$\text{Minimize}_{r>0} \quad \phi(y, r) = \phi_0(y, r) + \sum_{i=1}^s \phi_i(y, r_i)$$

where
$$\phi_0(y, r) = f_0(y) + \sum_{j=1}^m P[g_j(y), r]$$

and
$$\phi_i(y, r_i) = \min_{x_i, r_i} \{f_i(y, x_i) + \sum_{j=1}^{n_i} P[g_{ij}(y, x_i), r_i]\}$$

This multilevel algorithm takes a single Newton iteration to minimize each ϕ_i , followed by a single Newton iteration to minimize ϕ , and then each subsystem ϕ_i is updated. This constitutes one iteration of the multilevel optimization algorithm. Continuing the procedure will converge to the solution of the original problem. The penalty approach given in [H1] successfully exploits the additively separable form of the objective function and the special structure of the constraints. Computational results are given for a portal frame and show substantial time savings over a single-level approach.

In [S1] L. A. Schmit Jr. solves a minimum weight design problem including buckling constraints via a multilevel approach. He decomposes the primary problem into a system level design minimizing total structure weight and a set of uncoupled component level problems with component stiffness change as objective functions. A sequentially unconstrained minimization technique based on the extended interior penalty

function and a modified Newton's method for unconstrained minimization are used in this work.

In [K1] U. Kirsch, M. Reiss and U. Shamir propose a method to solve a structural design problem by partitioning the general optimization problem into a number of small subproblems. The design variable vector is partitioned into a number of vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_{NS}$, where NS is the number of substructures. When optimizing a single substructure i , only the values of the corresponding local design variables $\mathbf{x}_i$ are found, whereas the values of the non-local design variables are kept constant. The optimum value of each $\mathbf{x}_i$ is determined sequentially by an iterative process. Optimization of a substructure is based on transforming the constrained problem into a series of unconstrained problems using exterior penalty functions for the constraints. The partitioning and solution strategies of the proposed approach are problem dependent.

2.3 Strategies in Decomposition of Non-hierarchical Systems

Systems that are not amenable to hierarchical decomposition are known as network or non-hierarchical systems. By network it is meant that a subsystem can send output to and receive input from any other subsystem. Fig. 2 showed a network system example. Few papers deal with problems of this nature in the design optimization literature. In [S3] J. Sobieszcanski-Sobieski suggests a procedure for non-hierarchical system optimization. It includes system analysis, optimum sensitivity analysis, and optimization carried out in design subspaces corresponding to the subsystems. A coordination strategy is used to distribute constraint violations among the subsystems. It is suggested that all subspace optimizations are performed concurrently. No operational experience with the procedure is reported.

Another approach proposed by A. Díaz [D1] decomposes a non-hierarchical system into coupled subsystems. A fuzzy set based approach is used at all levels to model the optimization objective. The contribution of subproblem goals to the overall system goal is modeled using set operations. While solving the subproblems, convex approximations of non-local constraints are included, in addition to the local constraints, to insure global feasibility. The global solution is obtained by solving each subproblem in sequence. The success of this approach depends on the order in which the subproblems are solved. A numerical example reported in the paper shows convergence of the method.

The work cited above is representative of the recent literature in multilevel optimization methods in structural design. Multilevel optimization techniques in other fields are described by D. M. Nachane in [N1]. This is a methodological survey emphasizing the discussion of methods most suitable for applications in other disciplines. Early work in large-scale mathematical programming can also be found in [G1].

CHAPTER 3

BACKGROUND

The derivation of a decomposition algorithm presented later in Chapter 4 is based on well-known results of nonlinear programming. Lagrange multipliers and optimum sensitivity derivatives are two important concepts that will be used extensively in later chapters. This chapter provides the necessary material that will facilitate understanding of the development of the decomposition algorithm.

Given a nonlinear programming problem, named the *primal problem*, there exists a problem that is closely associated with it, named the *Lagrangian dual problem*. These two problems are given below.

Primal Problem PP

Find $\mathbf{x} \in X$, $X \subseteq \mathbb{R}^n$ that

Minimizes $f(\mathbf{x})$

Subject to $g_i(\mathbf{x}) \leq 0 \quad i=1, \dots, m$

where f and g_i are real valued scalar functions.

The *Lagrangian function* of problem PP is defined as:

$$L(\mathbf{x}, \mathbf{u}) \equiv f(\mathbf{x}) + \sum_{i=1}^m u_i g_i(\mathbf{x})$$

Associated with PP is the following problem:

Lagrangian Dual Problem DP

Find $\mathbf{u} \in \mathbb{R}^m$ that

Maximizes $\min_{\mathbf{x}} \{L(\mathbf{x}, \mathbf{u}): \mathbf{x} \in X\}$

Subject to $u_i \geq 0$

The i th component of $\mathbf{u}$, u_i , is referred to as the *dual variable* or *Lagrange multiplier* associated with the constraint $g_i(\mathbf{x}) \leq 0$.

Optimum sensitivity analysis investigates the sensitivity of the solution to an optimization problem to variations in the problem's parameters. It yields derivatives of the optimum values of the objective function and design variables with respect to the parameters. We will see in the next chapter that these derivatives represent the coupling between subsystems in system decomposition of optimal design. The following basic sensitivity theorems from [F1] will be used extensively in the next chapter.

Consider the general problem of determining a local solution $\mathbf{x}^*(\boldsymbol{\varepsilon})$ of problem $P^1(\boldsymbol{\varepsilon})$:

Problem $P^1(\boldsymbol{\varepsilon})$:

Find $\mathbf{x} \in \mathbb{R}^n$ that

Minimizes $f(\mathbf{x}, \boldsymbol{\varepsilon})$

$$\text{Subject to } g_i(\mathbf{x}, \boldsymbol{\varepsilon}) \leq 0 \quad i = 1, \dots, m$$

where $\boldsymbol{\varepsilon}$ is a parameter vector in $\mathbb{R}^k$. The Lagrangian function of $P^1(\boldsymbol{\varepsilon})$ is:

$$L(\mathbf{x}, \mathbf{u}, \boldsymbol{\varepsilon}) \equiv f(\mathbf{x}, \boldsymbol{\varepsilon}) + \sum_{i=1}^m u_i g_i(\mathbf{x}, \boldsymbol{\varepsilon})$$

Let $\mathbf{x}^*(\boldsymbol{\varepsilon})$ be the solution to problem $P^1(\boldsymbol{\varepsilon})$ for $\boldsymbol{\varepsilon}$ near 0 and $\mathbf{u}^*(\boldsymbol{\varepsilon})$ be the associated Lagrange multiplier vector. Then the local *optimal value function* is defined as $f^*(\boldsymbol{\varepsilon}) \equiv f(\mathbf{x}^*(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon})$ and the *optimum value Lagrangian* is defined as $L^*(\boldsymbol{\varepsilon}) = L[\mathbf{x}^*(\boldsymbol{\varepsilon}), \mathbf{u}^*(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon}]$. Throughout the thesis, the gradient operator ∇ is taken with respect to design variables of an optimization problem, unless otherwise specified.

Theorem 3.1 Basic Sensitivity Theorem

If

(i) The functions defining $P^1(\boldsymbol{\varepsilon})$ are twice continuously differentiable in $\mathbf{x}$ and their gradients with respect to $\mathbf{x}$ and the constraints are once continuously differentiable in $\boldsymbol{\varepsilon}$ in a neighborhood of $(\mathbf{x}^*, 0)$.

(ii) The second-order sufficient conditions for a local minimum of $P^1(0)$ hold at $\mathbf{x}^*$, with associated Lagrange multipliers $\mathbf{u}^*$.

(iii) The gradients $\nabla g_i(\mathbf{x}^*, 0)$, for i such that $g_i(\mathbf{x}^*, 0) = 0$, are linearly independent.

(iv) $u_i^* > 0$ when $g_i(\mathbf{x}^*, 0) = 0$, $i = 1, \dots, m$, i.e., strict complementary slackness holds.

Then,

(a) $\mathbf{x}^*$ is a local isolated minimizing point of problem $P^1(0)$ and the associated Lagrange multiplier $\mathbf{u}^*$ is unique.

(b) For $\boldsymbol{\varepsilon}$ in a neighborhood of 0, there exists a unique, once continuously differentiable vector function $\mathbf{y}^*(\boldsymbol{\varepsilon}) = [\mathbf{x}^*(\boldsymbol{\varepsilon}), \mathbf{u}^*(\boldsymbol{\varepsilon})]^T$ satisfying the second-order sufficient conditions for a local minimum of problem $P^1(\boldsymbol{\varepsilon})$ such that $\mathbf{y}^*(0) \equiv [\mathbf{x}^*, \mathbf{u}^*]$, and hence $\mathbf{x}^*(\boldsymbol{\varepsilon})$ is a locally unique minimum of problem $P^1(\boldsymbol{\varepsilon})$ with associated unique Lagrange multipliers $\mathbf{u}^*(\boldsymbol{\varepsilon})$.

(c) For $\boldsymbol{\varepsilon}$ near 0, the set of binding inequalities is unchanged, strict complementary slackness holds, and the binding constraint gradients are linearly independent at $\mathbf{x}^*(\boldsymbol{\varepsilon})$.

Theorem 3.2 Sensitivity of the optimal value function

If the conditions of Theorem 3.1 hold for problem for problem $P^1(\boldsymbol{\varepsilon})$, and if the problem functions are twice continuously differentiable in $(\mathbf{x}, \boldsymbol{\varepsilon})$ near $(\mathbf{x}^*, 0)$ then, in a neighborhood of $\boldsymbol{\varepsilon}=0$,

$$(a) \quad f^*(\boldsymbol{\varepsilon}) = L^*(\boldsymbol{\varepsilon})$$

$$(b) \quad \nabla_{\boldsymbol{\varepsilon}} f^*(\boldsymbol{\varepsilon}) = \nabla_{\boldsymbol{\varepsilon}} L^*(\boldsymbol{\varepsilon})$$

$$= \nabla_{\boldsymbol{\varepsilon}} f(\boldsymbol{\varepsilon}) + \sum_i^m u_i(\boldsymbol{\varepsilon}) \nabla_{\boldsymbol{\varepsilon}} g_i(\boldsymbol{\varepsilon}) = \nabla_{\boldsymbol{\varepsilon}} f(\boldsymbol{\varepsilon}) + \mathbf{u}(\boldsymbol{\varepsilon})^T \nabla_{\boldsymbol{\varepsilon}} \mathbf{g}(\boldsymbol{\varepsilon})$$

CHAPTER 4

NON-HIERARCHICAL SYSTEM DECOMPOSITION

An algorithm for optimal design of non-hierarchical system by decomposition will be developed in this chapter. Its development is based on several propositions derived here. In order to facilitate the understanding of the organization of this chapter, a general flow diagram of the algorithm is given in Fig. 8.

4.1 Problem Statement

Fig. 2 shows an example of the network type system optimization problem that will be considered in this thesis. Let $\mathbf{x} \in X$, $X \subseteq R^n$, be the vector of design variables and let $\mathbf{y}(\mathbf{x}) \in Y$, $Y \subseteq R^r$ be a r -tuple vector of state variables (e.g., stresses, weight or displacement) to be used in the computation of objective and constraint functions. The general optimization problem calls for finding a set of design variables $\mathbf{x} \in X$ that minimizes a scalar objective function $f(\mathbf{x}, \mathbf{y}(\mathbf{x}))$ subject to system constraints $g(\mathbf{x}, \mathbf{y}(\mathbf{x})) \leq 0$. Only inequality constraints are considered in the general

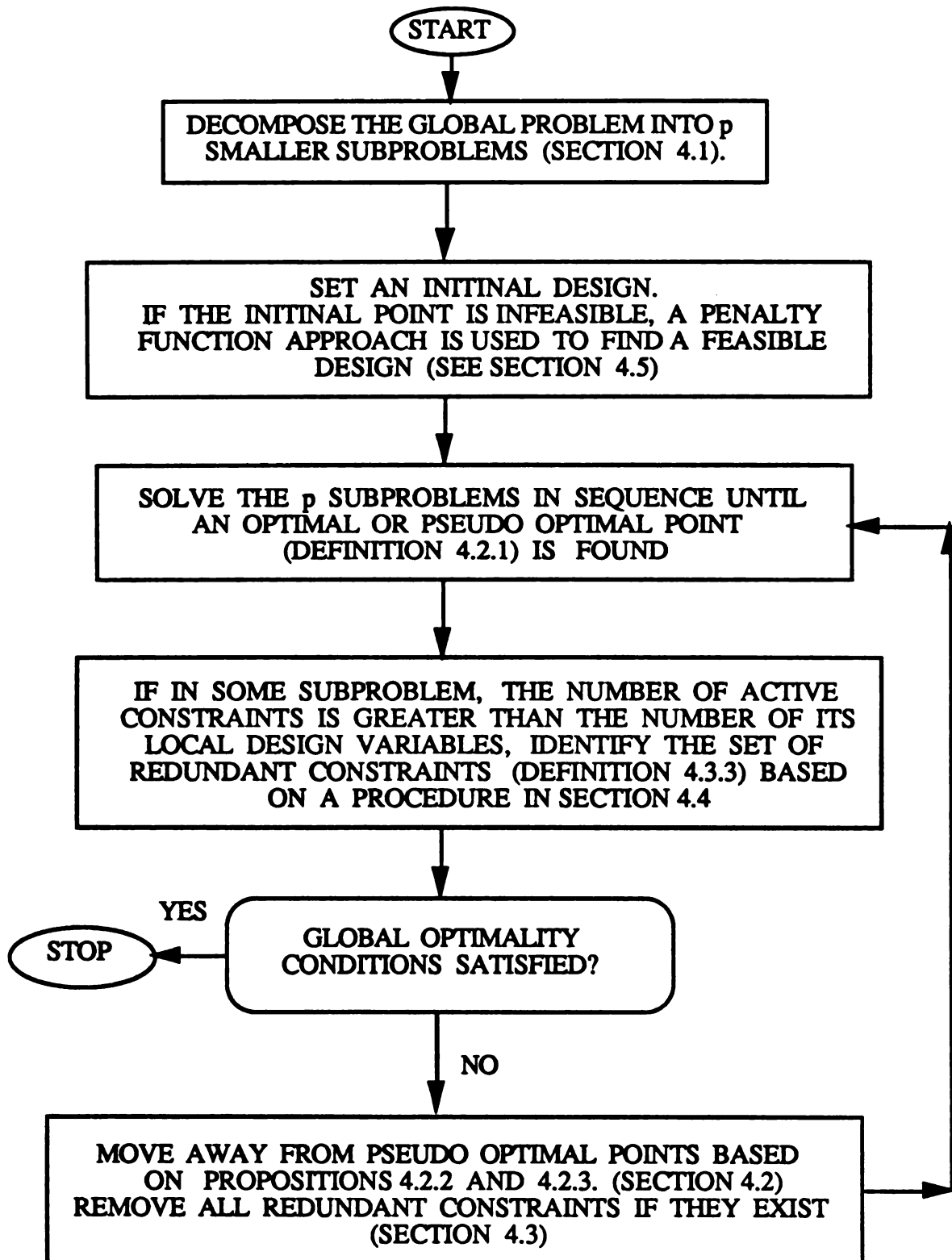


Figure 8. A sequential algorithm for non-hierarchical system decomposition

optimization problem. With this notation, the *global* (not decomposed) problem has the following form:

Global Problem P (no decomposition)

Find $\mathbf{x} \in X \subseteq \mathbb{R}^n$ that

$$\text{Minimizes} \quad f(\mathbf{x}, \mathbf{y}(\mathbf{x})) \quad (1.a)$$

$$\text{Subject to} \quad \mathbf{g}(\mathbf{x}, \mathbf{y}(\mathbf{x})) \leq \mathbf{0}, \quad \mathbf{g} \in \mathbb{R}^m. \quad (1.b)$$

where $\mathbf{g}(\mathbf{x}, \mathbf{y}(\mathbf{x})) = \{g_1(\mathbf{x}, \mathbf{y}(\mathbf{x})), g_2(\mathbf{x}, \mathbf{y}(\mathbf{x})), \dots, g_m(\mathbf{x}, \mathbf{y}(\mathbf{x}))\}^T$.

Problem P is to be decomposed into several smaller subproblems. The decomposition is accomplished by partitioning the vector of the design variables $\mathbf{x}$ into a number of *uncoupled* vectors $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^p$. p is the number of subsystems. In general, each subsystem represents a particular physical aspect of the global system and it is prescribed to be analyzed by experts in the local engineering discipline (e.g., structural design, engine design).

The following features characterize the problem and are fundamental in the construction of the solution strategy:

1. Although the decomposition is carried out on the domain X , the state space Y is partitioned accordingly. As the consequence of the decomposition, each subsystem has only the knowledge necessary to compute *only* its local subset $\{y_1^i, y_2^i, \dots, y_{r_i}^i\}$ of state variables $\mathbf{y}$.

2. The optimization of each subsystem changes only the local design variables $\mathbf{x}^i$ in the design subdomain $X^i \subset X$ and evaluates its own state variables. However, all

subsystems minimize the same global objective function f and must satisfy *all* of the system constraints in g .

Summary of Problem Decomposition

i. The original vector of design variables x is partitioned into p uncoupled vectors $x^1, x^2, \dots, x^p$, where $x^i \in X^i \subseteq R^{n_i}$ and therefore, x is the direct sum of the x^i 's, i.e., $x = x^1 \oplus x^2 \oplus \dots \oplus x^p$. Let R^{n_i} be the usual Euclidean space of dimension n_i and $R^{n_1}, R^{n_2}, \dots, R^{n_p}$ be design subspaces of the global design space R^n . Each design optimization in R^{n_i} changes only *local variables* x^i , and is called the i -th subproblem optimization. Notice that a variable of x that is optimized in one subproblem will not be a design variable in any other subproblems.

ii. The vector of state variables $y(x)$ is decomposed into p vectors $y^1(x), y^2(x), \dots, y^p(x)$ of dimensions $r_1, r_2, \dots, r_p$, respectively, where $y^i(x) \in Y^i, Y^i \subset Y$.

iii. The vector $y^i(x)$ can be evaluated *only* in the i -th subproblem and is called the *local state variable* of the i -th subproblem, $i=1, \dots, p$. One or several entries of y may appear in the objective or in any of the constraint functions. Since y is, potentially, a function of any of the local variables $x^1, x^2, \dots, x^p$, all subsystems are coupled.

iv. Each subproblem optimization includes all constraints.

v. All subproblems seek to optimize the same objective function f .

Subproblems constructed in this way are coupled by non-local state variables and non-local design variables. Let $P^1, P^2, \dots, P^p$ denote the p optimization subproblems. Define $x^0 \equiv \{x^{10}, x^{20}, \dots, x^{p0}\}$ where the $x^{j0} \in X^j, j=1, 2, \dots, p$, are

fixed parameters. $\mathbf{x}^0$ is the *reference design* for the problem P^i . Each subproblem has the following form:

Problem $P^i(\mathbf{x}^0)$

Find $\mathbf{x}^i \in X^i \subseteq \mathbb{R}^{n_i}$ that

$$\text{Minimizes} \quad f(\mathbf{x}^i, \mathbf{y}(\mathbf{x}^i))_{|\mathbf{x}^0} \quad (2.a)$$

$$\text{Subject to} \quad \mathbf{g}(\mathbf{x}^i, \mathbf{y}(\mathbf{x}^i))_{|\mathbf{x}^0} \leq \mathbf{0}, \quad \mathbf{g} \in \mathbb{R}^m. \quad (2.b)$$

while fixing the non-local variables $\mathbf{x}^j$, $j \neq i$ at their respective reference values $\mathbf{x}^{j0}$. The notation

$$f(\mathbf{x}^i, \mathbf{y}(\mathbf{x}^i))_{|\mathbf{x}^0} \equiv f(\{\mathbf{x}^{10}, \mathbf{x}^{20}, \dots, \mathbf{x}^i, \dots, \mathbf{x}^{p0}\}, \mathbf{y}(\{\mathbf{x}^{10}, \mathbf{x}^{20}, \dots, \mathbf{x}^i, \dots, \mathbf{x}^{p0}\}))$$

and

$$\mathbf{g}(\mathbf{x}^i, \mathbf{y}(\mathbf{x}^i))_{|\mathbf{x}^0} \equiv \mathbf{g}(\{\mathbf{x}^{10}, \mathbf{x}^{20}, \dots, \mathbf{x}^i, \dots, \mathbf{x}^{p0}\}, \mathbf{y}(\{\mathbf{x}^{10}, \mathbf{x}^{20}, \dots, \mathbf{x}^i, \dots, \mathbf{x}^{p0}\}))$$

is used for simplicity.

4.2 Derivation of the Algorithm

An algorithm for optimal design of non-hierarchical systems by decomposition developed in this section is based on the following propositions, derived by assuming that,

(a) all functions can be evaluated exactly for each subproblem,

(b) the number of active constraints of each subproblem in the optimization process is always less than that of local design variables.

Later, if there is a function in some subproblem depends on non-local state variables, convex approximations will be introduced to approximate the non-local state variables. Also, if at some point a subproblem violates assumption (b), it will be shown later that there exist redundant constraints in the subproblem, and the propositions stated below are still applicable to the subproblem after removal of all redundant constraints.

Proposition 4.2.1 below states that under what conditions a point $\mathbf{x}$ that satisfy the Karush-Kuhn-Tucker (KKT) conditions of all subproblems P^i is also a KKT point for the global problem P . KKT points of the subproblems that are not optimal points of P are defined later as pseudo optimal points. Propositions 4.2.2 and 4.2.3 will provide a strategy to move away from pseudo optimal points while improving the global objective.

Proposition 4.2.1

Let $\mathbf{x}^* = \{\mathbf{x}^{1*}, \mathbf{x}^{2*}, \dots, \mathbf{x}^{P*}\}$, where $\mathbf{x}^* \in X \subseteq \mathbb{R}^n$, $\mathbf{x}^{i*} \in X^i$, be the vector of global design variables. If

(i). Each $\mathbf{x}^{i*}$ is a KKT point for the subproblem $P^i(\mathbf{x}^*)$, with Lagrange multiplier vector $\boldsymbol{\lambda}^{i*}$.

(ii). Each component of $\boldsymbol{\lambda}^{i*}$ in $P^i(\mathbf{x}^*)$ has the same value for all subproblems, i.e.,

$$\boldsymbol{\lambda}^{1*} = \boldsymbol{\lambda}^{2*} = \dots = \boldsymbol{\lambda}^{P*} = \boldsymbol{\lambda}^* \quad (3)$$

Then the point $\mathbf{x}^*$ is a KKT point of the global problem P with Lagrange multiplier vector $\boldsymbol{\lambda}^*$.

To prove Proposition 4.2.1, note first that at $\mathbf{x}^*$, the stationary conditions of the Lagrangian functions with respect to $\mathbf{x}^i$ for P^i , $i=1, \dots, p$, are the same as the stationary conditions of the Lagrangian function for P with respect to $\mathbf{x}$. In addition, each P^i includes all the constraints, the complementary slackness conditions of each P^i are the same as the global problem P . This completes the proof.

Proposition 4.2.1 can be used to check whether a given point satisfies the global KKT conditions. From Fig. 4 and Fig. 5, it is possible for a point $\mathbf{x}^* = \{\mathbf{x}^{1*}, \mathbf{x}^{2*}, \dots, \mathbf{x}^{p*}\}$ to be a KKT point of all subproblems and yet be a non-optimal point of P if its associated Lagrange multipliers are not all equal. We call this point pseudo optimal.

Definition 4.2.1

A point $\mathbf{x}^* = \{\mathbf{x}^{1*}, \mathbf{x}^{2*}, \dots, \mathbf{x}^{p*}\}$ is said to be a *pseudo optimal point* for problem P if

- (i). Each $\mathbf{x}^{i*}$ satisfies the KKT conditions for problem $P^i(\mathbf{x}^*)$, with multiplier vector λ^{i*} .
- (ii). There exists an active constraint g_k and a pair of problems P^i and P^j , such that $\lambda_k^i \neq \lambda_k^j$.

The operational significance of pseudo optimal points is the following: if subproblems P^i are solved sequentially, using as the reference point the last available solution, the solution process is likely to stop at a pseudo optimal point before reaching a true KKT point. At the pseudo optimal point, no further progress will be possible. A special step must be taken to move away from the pseudo optimal point along a feasible descent direction. The following proposition suggests a procedure to find a feasible descent direction for f in some design subspace $X^i \subseteq R^{n_i}$.

Proposition 4.2.2

(i). Let $\mathbf{x}^0$ be a pseudo optimal point of problem P. Let λ^r , $r=1, 2, \dots, p$, be the Lagrange multiplier vectors of the subproblems $P^r(\mathbf{x}^0)$. Let f^0 be the value of f at $\mathbf{x}^0$.

(ii). Assume that the conditions of Theorem 3.2 hold.

(iii). Select indices i, j and k , so that there exists an active constraint g_k at $\mathbf{x}^0$ in problems P^i and P^j with

$$\lambda_k^j > \lambda_k^i > 0$$

(iv). Let $\mathbf{x}^{i*}(\epsilon)$ be the solution to the perturbed subproblem

Problem $P^i(\mathbf{x}^0(0))$:

Find $\mathbf{x}^i \in X^i \subseteq \mathbb{R}^{n_i}$ that

$$\text{Minimizes} \quad f(\mathbf{x}^i, \mathbf{y}(\mathbf{x}^i))|_{\mathbf{x}^0} \quad (4.a)$$

$$\text{Subject to} \quad g(\mathbf{x}^i, \mathbf{y}(\mathbf{x}^i))|_{\mathbf{x}^0} \leq -\epsilon, \quad (4.b)$$

where $\epsilon = \{\epsilon_1, \epsilon_2, \dots, \epsilon_m\}$.

(v). With $\mathbf{x}^i$ fixed at the value $\mathbf{x}^{i*}(\epsilon)$, let $\mathbf{x}^{j*}$ be the solution to the subproblem

Problem $P^j(\mathbf{x}^0(\epsilon))$:

Find $\mathbf{x}^j \in X^j \subseteq \mathbb{R}^{n_j}$ that

$$\text{Minimizes} \quad f(\mathbf{x}^j, \mathbf{y}(\mathbf{x}^j))|_{\mathbf{x}^0(\epsilon)} \quad (5.a)$$

$$\text{Subject to} \quad g(\mathbf{x}^j, \mathbf{y}(\mathbf{x}^j))|_{\mathbf{x}^0(\epsilon)} \leq 0, \quad (5.b)$$

where $\mathbf{x}^0(\boldsymbol{\varepsilon}) \equiv \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{i^*}(\boldsymbol{\varepsilon}), \dots, \mathbf{x}^{p^0}\}$. Let $f^*(\boldsymbol{\varepsilon})$ be the value of f at the solution to problem $P^j(\mathbf{x}^0(\boldsymbol{\varepsilon}))$ in (5).

If components of the right-hand-side perturbation vector $\boldsymbol{\varepsilon}$ of (4) are chosen by the rule:

$$\varepsilon_s = \begin{cases} \varepsilon & \text{if } s = k \\ 0 & \text{if } s \neq k \end{cases} \quad \text{for } s=1, \dots, m \text{ and } \varepsilon > 0,$$

then

$$f^*(\boldsymbol{\varepsilon}) < f^0$$

Note first that the existence of a constraint g_k in condition (iii) with the desired properties follows from the definition of pseudo optimal point and Proposition 4.2.1. Consider now the problem $P^j(\mathbf{x}^0(\boldsymbol{\varepsilon}))$ in (5). In a neighborhood of $\boldsymbol{\varepsilon}=0$, the solution $\mathbf{x}^{j^*}$ is effectively parametrized by $\boldsymbol{\varepsilon}$ through $\mathbf{x}^{i^*}(\boldsymbol{\varepsilon})$. Applying Theorem 3.2, to P^j , we have,

$$\frac{df^*}{d\boldsymbol{\varepsilon}} = \frac{\partial f}{\partial \boldsymbol{\varepsilon}} + \lambda_k^j \frac{\partial g_k}{\partial \boldsymbol{\varepsilon}} = \left(\frac{\partial f}{\partial \mathbf{x}^i} + \lambda_k^j \frac{\partial g_k}{\partial \mathbf{x}^i} \right) \frac{\partial \mathbf{x}^i}{\partial \boldsymbol{\varepsilon}} \quad (6)$$

where all quantities are evaluated at $\boldsymbol{\varepsilon}=0$. Also, since $\mathbf{x}^0$ is pseudo optimal, $\mathbf{x}^{i^*}(0)$ satisfies the KKT conditions for problem $P^i(\mathbf{x}^0)$

$$\left(\frac{\partial f}{\partial \mathbf{x}^i} + \sum_s \lambda_s^i \frac{\partial g_s}{\partial \mathbf{x}^i} \right) = 0 \quad (7)$$

Combining (6) and (7), obtain

$$\begin{aligned}\frac{df^*}{d\varepsilon} &= \left(- \sum_s \lambda_s^i \frac{\partial g_s}{\partial x^i} + \lambda_k^j \frac{\partial g_k}{\partial x^i} \right) \frac{\partial x^i}{\partial \varepsilon} = \\ &= \left(\lambda_k^j - \lambda_k^i \right) \frac{\partial g_k}{\partial x^i} \frac{\partial x^i}{\partial \varepsilon} - \sum_{s \neq k} \lambda_s^i \left(\frac{\partial g_s}{\partial x^i} \frac{\partial x^i}{\partial \varepsilon} \right)\end{aligned}\quad (8)$$

where the sum is over the active constraints at x^0 . But, from the Theorem 3.1, if g_s is active at $\varepsilon=0$, g_s is identically zero in a neighborhood of $\varepsilon=0$. Therefore

$$\frac{dg_s}{d\varepsilon} = \frac{\partial g_s}{\partial \varepsilon} + \frac{\partial g_s}{\partial x^i} \frac{\partial x^i}{\partial \varepsilon} = 0 \quad (9)$$

and since

$$\frac{\partial g_s}{\partial \varepsilon} = \begin{cases} 1 & \text{if } s = k \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

(8) , (9) and (10) can be used to obtain

$$\frac{df^*}{d\varepsilon} = - \left(\lambda_k^j - \lambda_k^i \right) \quad (11)$$

Finally, from condition (iii), we have

$$\frac{df^*}{d\varepsilon} < 0$$

Since the solution to $P^j(x^0(\varepsilon))$ is feasible and different from x^0 , this implies that $f^*(\varepsilon) < f^0$ for small positive ε . Therefore, the solution to $P^j(x^0(\varepsilon))$ is better than x^0 , i.e., the procedure get 'unstuck' from the pseudo optimal point. This result can be used to move away from pseudo optimal points.

From the previous result, at a pseudo optimal point, the combination of a solution to P^i followed by P^j will produce a new feasible point with reduced objective

provided that P^i has a solution for a given value of ε . Choosing ε too large may make the feasible space of problem (4) empty. To avoid this difficulty it is convenient to replace problem $P^i(x^0(0))$ in (4) by the following equivalent problem:

Problem $P^i(x^0(0))$:

Find $x^i \in X^i \subseteq \mathbb{R}^{n_i}$ that

$$\text{Minimizes} \quad A(x^i, \lambda^0) \quad (12.a)$$

$$\text{Subject to} \quad g_s(x^i, y(x^i))|_{x^0} \leq 0, \quad s=1,2,\dots,m \quad s \neq k \quad (12.b)$$

where $A(x^i, \lambda^0)$ is the augmented Lagrangian function.

$$A(x^i, \lambda^0) = f(x^i, y(x^i))|_{x^0} + \lambda^0 g_k(x^i, y(x^i))|_{x^0} + r^2 (g_k(x^i, y(x^i))|_{x^0})^2 \quad (12.c)$$

λ^0 and $r > 0$ are scalars. The following proposition guarantees that problem (12) is equivalent to problem (4).

Proposition 4.2.3

i. Let x^* be a unique solution to the problem:

Find x that

$$\text{Minimizes} \quad f(x) \quad (13.a)$$

$$\text{Subject to} \quad g(x) \leq 0, \quad g \in \mathbb{R}^m \quad (13.b)$$

with associated Lagrange multiplier vector λ^* . Suppose that $\lambda_k^* > 0$.

ii. Let x^0 be a solution to the problem:

Find x that

$$\text{Minimizes} \quad A(\mathbf{x}, \lambda^0) = f(\mathbf{x}) + \lambda^0 g_k(\mathbf{x}) + r^2 (g_k(\mathbf{x}))^2 \quad (14.a)$$

$$\text{Subject to} \quad g_s(\mathbf{x}) \leq 0, \quad s=1,2,\dots,m; \quad s \neq k \quad (14.b)$$

where $r > 0$ is a given scalar.

If $\lambda^0 > \lambda_k^*$, then

$$g_k(\mathbf{x}^0) = -\varepsilon < 0$$

Proposition 4.2.3 states that, provided that λ^0 is greater than λ_k^* , the solutions to (12) and (4) are equivalent. Problem (12) shifts the control on the step size from ε to r and avoids the possibility of making the feasible space empty by picking ε too large.

To show equivalence of the two problems one must first realize that for r large enough (but bounded), $\mathbf{x}^*$ is also the minimizer of $A(\mathbf{x}, \lambda_k^*)$. Therefore,

$$f(\mathbf{x}^*) + \lambda_k^* g_k(\mathbf{x}^*) + r^2 (g_k(\mathbf{x}^*))^2 \leq f(\mathbf{x}^0) + \lambda_k^* g_k(\mathbf{x}^0) + r^2 (g_k(\mathbf{x}^0))^2 \quad (15)$$

Next, define the Lagrangian function L as

$$L(\lambda) \equiv \min_{\mathbf{x} \in G} f(\mathbf{x}) + r^2 g_k^2(\mathbf{x}) + \lambda g_k(\mathbf{x})$$

where $G \equiv \{\mathbf{x}: g_s(\mathbf{x}) \leq 0; s \neq k\}$.

It follows that

$$L(\lambda^0) \leq L(\lambda_k^*) \quad (16)$$

since L is the Lagrangian associated with

$$\min_{\mathbf{x} \in G} f(\mathbf{x}) + r^2 g_k^2(\mathbf{x})$$

which has solution $(\mathbf{x}^*, \lambda_k^*)$. Combining (15) and (16), obtain

$$(\lambda^0 - \lambda_k^*) g_k(\mathbf{x}^0) \leq 0$$

Since $\mathbf{x}^*$ is unique, (15) implies that $g_k(\mathbf{x}^0) \neq 0$ and hence $\lambda^0 > \lambda_k^*$ will force $g_k(\mathbf{x}^0) < 0$, as desired.

A strategy to solve the global problem in (1) by decomposition based on the propositions derived above can be summarized as follows: the solution of the global problem P in (1) will be obtained by solving each subproblem of (2) in sequence. Once a pseudo optimal point $\mathbf{x}^0$ of problem P is found, proposition 4.2.1 is used to check if $\mathbf{x}^0$ satisfies the global KKT conditions. If not, propositions 4.2.2 and 4.2.3 suggest a procedure such that the global objective can be further improved by sequentially solving subproblems $P^i(\mathbf{x}^0(0))$ and $P^j(\mathbf{x}^0(\mathbf{e}))$, defined in (12) and (5). The main steps of the algorithm are listed as following:

Given a feasible reference point $\mathbf{x}^0$:

Step 1. If $\mathbf{x}^0$ satisfies the conditions in proposition 4.2.1, stop and output the global optimal solution $\mathbf{x}^0$.

Step 2. If $\mathbf{x}^0$ is not pseudo optimal, solve all subproblems in sequential order.

Step 3. If $\mathbf{x}^0$ is pseudo optimal, identify indices i , j , and k such that

$$\lambda_k^j > \lambda_k^i > 0$$

Then solve, in sequence, P^i and P^j defined by equations (12) and (5)

Step 4. Update the reference point with the latest point and restart the procedure at step 1.

To apply this algorithm in practical applications, we must consider the case in which the number of active constraints of some subproblem in the optimization process is greater than that of local design variables, i.e., assumption (b) made at the beginning of the section does not hold. It will be shown in Section 4.3 that if this is the case, there exist *redundant constraints* in the subproblem. A procedure to identify the set of redundant constraints will be given in Section 4.4. Other implementation issues such as infeasible starting point, evaluation of non-local functions via convex approximations, and the use of optimum sensitivity derivatives will be discussed in Section 4.5, 4.6, and 4.7, respectively. In the last section, Section 4.8, a summary of the results derived in this chapter will be given.

4.3 Redundancy

Proposition 4.2.1 checks the global optimality conditions, while propositions 4.2.2 and 4.2.3 are useful in the construction of a strategy that can go past pseudo optimal points. However, they are not sufficient to guarantee convergence of the decomposition algorithm in the presence of *redundant constraints*. Redundancy often occurs in a subproblem if the number of active constraints at the solution is greater than the number of design variables. Even if the solution to the global problem P is a regular point of the constraints, redundancy is inevitable in the decomposed problems because the number of constraints is the same in the global problem and in all subproblems, regardless of the dimension of the local design subspace R^{n_i} . For instance, in the example of Fig. 4 and Fig. 5, there are two active constraints at the optimum point but the local subspaces are only of dimension 1 ($n_1=n_2=1$). Neither in P^1 nor in P^2 will the gradients of the active constraints be linearly independent at the optimum. This feature can cause difficulties in the convergence of the algorithm used to solve the local

subproblems since the existence of redundancy in subproblems will not satisfy the conditions for propositions in Section 4.2. In this case, it will be shown that proposition 4.2.2 and 4.2.3 are still true if redundant constraints are removed from $P^i(x^0(0))$ of (4) and $P^j(x^0(\epsilon))$ of (5).

Definition 4.3.1

A vector $a \in R^n$ is a *linear non-negative combination* of vectors $b_1, \dots, b_N$ where $b_s \in R^n$, if there exist non-negative scalars $\alpha_1, \dots, \alpha_N$, not all of them zero, such that $a = \sum_{s=1}^N \alpha_s b_s$.

Let $x^0 \equiv \{x^{10}, x^{20}, \dots, x^{p0}\}$ be a feasible reference point for the generic i -th subproblem $P^i(x^0)$, and the notation $x^{i*} \equiv \{x^{10}, x^{20}, \dots, x^{i*}, \dots, x^{p0}\}$ is used for simplicity (see equations (2)). Let $A_i = \{j: g_j(x^{i*}) = 0\}$, be the set of indices of active constraints at x^{i*} in $P^i(x^0)$. As usual, the gradient operator ∇ denotes the partial derivatives with respect to local design variables x^i , i.e., $\nabla(\cdot) = \frac{\partial(\cdot)}{\partial x^i}$.

Definition 4.3.2

Let $L(x^i, \lambda) = f(x^i) + \sum_{s=1}^m \lambda_s g_s(x^i)$ be the Lagrangian function of problem $P^i(x^0)$. x^{i*} is said to be a *pseudo optimal irregular* point or simply an *irregular* point of $P^i(x^0)$ if:

- (i). x^{i*} is feasible.
- (ii). There is a multiplier vector $\lambda \geq 0$, $\lambda \in R^m$, such that $\nabla L(x^i, \lambda) = 0$.
- (iii). $\lambda_s g_s(x^{i*}) = 0 \quad s = 1, \dots, m$.

(iv). The vectors $\nabla g_s(\mathbf{x}^{i*})$ for $s \in A_i$ are not linearly independent in $\mathbb{R}^{n_i}$.

Because of (iv), there is an infinite number of multiplier vectors associated with $\mathbf{x}^{i*}$. Note that if all the vectors $\nabla g_s(\mathbf{x}^{i*})$ for $s \in A_i$ were linearly independent in $\mathbb{R}^{n_i}$, an irregular point would be a KKT point.

Definition 4.3.3

A constraint g_t is a *redundant constraint* of $P^i(\mathbf{x}^0)$ at $\mathbf{x}^{i*}$, if

- (i). g_t is active at $\mathbf{x}^{i*}$.
- (ii). $\mathbf{x}^{i*}$ is an irregular point for $P^i(\mathbf{x}^0)$.
- (iii). $\nabla g_t(\mathbf{x}^{i*})$ is a linear non-negative combination of vectors $\nabla g_s(\mathbf{x}^{i*})$, where $s \neq t, s \in A_i$.

An example of redundant constraint is depicted in Fig. 9.

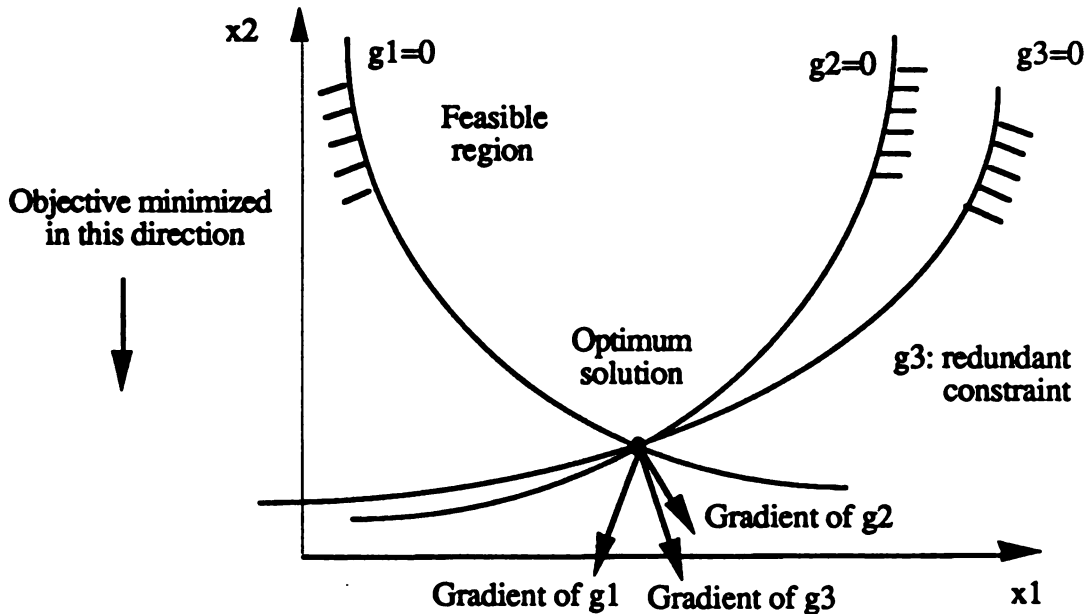


Figure 9. Existence of a redundant constraint

The following proposition states that at an irregular point, there is a redundant constraint. As before, A_i is the set of indices of active constraints at x^{i*} . Let K_A be the cardinality of A_i . Redundancy often occurs when the number of local design variables in $P^i(x^0)$, n_i , is less than the number of active constraints. Thus, we assume $n_i < K_A$ in the following discussion.

Proposition 4.3.1

At an irregular point x^{i*} of $P^i(x^0)$, let F be the set of feasible directions for $P^i(x^0)$ at x^{i*} ,

$$F = \{d \in R^{n_i} : \nabla g_j(x^{i*})d \leq 0 \text{ for all } j \in A_i\}.$$

If x^{i*} is a regular point of constraints in the global problem P , and if F has at least one non-zero element, then,

(i). There is a redundant constraint g_t in $P^i(x^0)$ at x^{i*} .

(ii). Removal of g_t from $P^i(x^0)$ will not change F , the set of feasible directions for $P^i(x^0)$ at x^{i*} , i.e.,

$$F' = \{d \in R^{n_i} : \nabla g_j(x^{i*})d \leq 0 \text{ for all } j \in A_i, j \neq t\},$$

then $F' = F$.

Outline of proof:

Let $\tilde{A} = \{\nabla g_j(x^{i*}) : j \in A_i\}$, be the set of gradient vectors of active constraints in A_i , and $H(\tilde{A})$ be the *convex hull* of $\tilde{A}$ (see Appendix). For an active constraint

g_t at x^{i*} , $t \in A_i$ and $\nabla g_t \in H(\tilde{A})$ from the definition of a convex hull. Using Caratheodory's theorem (Appendix),

$$\nabla g_t = \sum_{s=1}^{n_i+1} \beta_s \nabla g_s, \quad \nabla g_s \in \tilde{A} \quad (17)$$

and $\sum_{s=1}^{n_i+1} \beta_s = 1$ and $\beta_s \geq 0$ for $s = 1, \dots, n_i + 1$.

where n_i is the dimension of R^{n_i} .

Case 1. If there exists an index t , $t \in A_i$, such that ∇g_t does not appear in the right hand side of (17).

Then ∇g_t can be expressed as a linear non-negative combinations of elements of $\tilde{A}$ different from ∇g_t . By definition, g_t is a redundant constraint.

Case 2. For any $t \in A_i$, ∇g_t appears in the right hand side of (17).

Since the maximum number of linearly independent vectors in R^{n_i} is n_i and $n_i < K_A$, then there exist scalars ω_s , $s=1, \dots, n_i + 1$, not all of them zero, such that

$$\sigma \sum_{s=1}^{n_i+1} \omega_s \nabla g_s = 0 \quad (18)$$

for any scalar σ . Combining (17) and (18), for any $t \in A_i$,

$$\begin{aligned}
\nabla g_t &= \sum_{s=1}^{n_i+1} \beta_s \nabla g_s - \sigma \sum_{s=1}^{n_i+1} \omega_s \nabla g_s \\
&= \sum_{s=1}^{n_i+1} (\beta_s - \sigma \omega_s) \nabla g_s
\end{aligned} \tag{19}$$

Select $t \in A_i$, and σ is chosen such that

$$\frac{\beta_t}{\omega_t} = \text{minimum}_{1 \leq s \leq (n_i+1)} \left\{ \frac{\beta_s}{\omega_s} : \omega_s > 0 \right\} = \sigma \tag{20}$$

Note that $\sigma > 0$. By construction, for all $s=1, \dots, n_i + 1$,

$$\beta_s - \sigma \omega_s \geq 0$$

and

$$\beta_t - \sigma \omega_t = 0$$

Define now

$$\alpha_s = \beta_s - \sigma \omega_s \tag{21}$$

Substituting (21) into (19),

$$\nabla g_t = \sum_{s=1}^{n_i+1} \alpha_s \nabla g_s \tag{22}$$

where $\alpha_s \geq 0$ for all $s = 1, \dots, n_i + 1$ and $\alpha_t = 0$ as needed. By definition, g_t is a redundant constraint. This proves (i).

To verify (ii), it is clear that $F \subseteq F'$. Conversely, for each element $d \in F'$, $\nabla g_t d \leq 0$ from (22), hence $d \in F$. Thus, $F=F'$.

It must be emphasized that for proposition 4.3.1 to be true, the conditions

- (i). $\mathbf{x}^{i*}$ is a regular point of constraints in the global problem P,
- (ii). F has at least one non-zero element,

must hold. Examples which violate the above two conditions are given below to show that proposition 4.3.1 is not true.

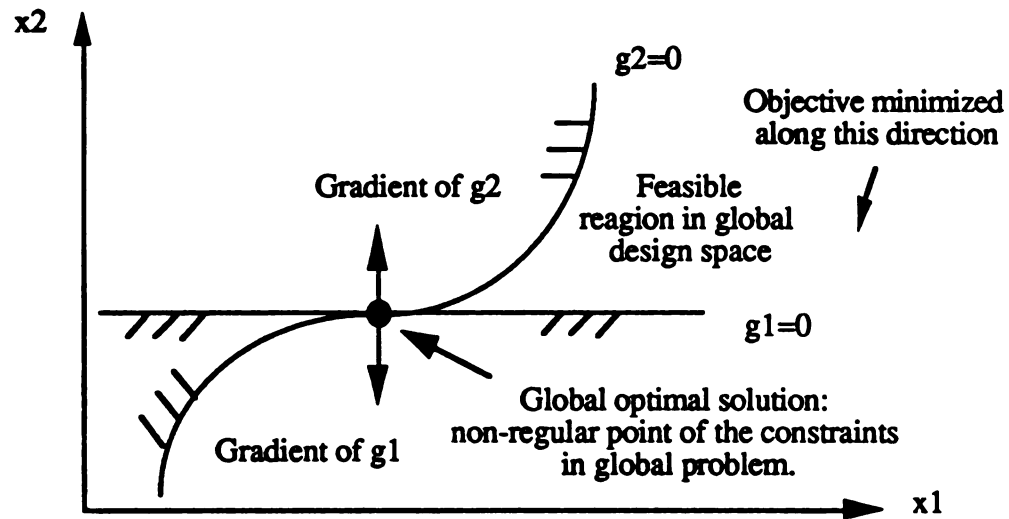


Figure 10. Counterexample 1: proposition 4.3.1 is not true

In Fig. 10, if the two dimensional global problem is decomposed into two, one dimensional subproblems, proposition 4.3.1 does not hold in either subproblems at the global optimal solution. In Fig. 11 below, the feasible region of the subproblem is a point. At this point, proposition 4.3.1 does not hold.

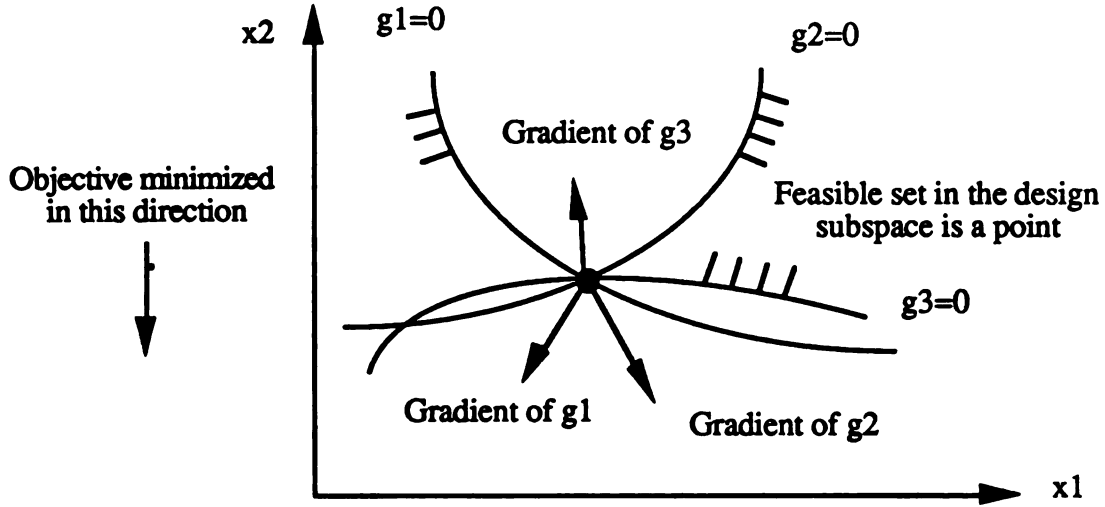


Figure 11. Counterexample 2: proposition 4.3.1 is not true

Using the definitions of redundant constraint and pseudo optimal irregular point we have the following proposition:

Proposition 4.3.2

At an irregular point $\mathbf{x}^{i*}$ of $P^i(\mathbf{x}^0)$, If g_t is a redundant constraint at $\mathbf{x}^{i*}$, then

(i). $\mathbf{x}^{i*}$ is associated with a multiplier vector $\boldsymbol{\lambda} \geq 0$, whose component corresponding to g_t , λ_t , is zero.

(ii). $\mathbf{x}^{i*}$ is either an irregular point or a KKT point for $P^i(\mathbf{x}^0)$ without the constraint g_t .

Proposition 4.3.2 states that if $\mathbf{x}^{i*}$ is an irregular point of $P^i(\mathbf{x}^0)$, it is a KKT point with a *unique* Lagrange multiplier vector satisfying strict complementary slackness condition for $P^i(\mathbf{x}^0)$ with all redundant constraints removed. Equivalently, $\mathbf{x}^{i*}$ is an irregular point for $P^i(\mathbf{x}^0)$ associated with a *unique* Lagrange multiplier vector $\boldsymbol{\lambda} \geq 0$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$, whose components corresponding to redundant constraints are zero, and whose components associated with non-redundant constraints are positive.

This unique Lagrange multiplier vector here is called the *exact Lagrange multiplier vector* for $P^i(x^0)$.

With the introduction of the redundancy concept, we are now ready to expand the definition of pseudo optimal point (definition 4.2.1), propositions 4.2.2 and 4.2.3.

Definition 4.3.4 (with redundancy)

A point $x^* = \{x^{1*}, x^{2*}, \dots, x^{p*}\}$ is said to be a *pseudo optimal point* if

- (i). Each x^{i*} is a KKT point or an irregular point for problem $P^i(x^*)$, with exact Lagrange multiplier vector λ^{i*} ,
- (ii). There exists some non-redundant constraint g_k in a pair of problems P^i and P^j , such that $\lambda_k^i \neq \lambda_k^j$.

In the presence of redundancy, propositions 4.2.2 and 4.2.3 will no longer be true because the conditions of Theorem 3.1 and 3.2 do not hold. Propositions 4.3.3 and 4.3.4 derived below show that after removal of all redundant constraints in $P^i(x^0(0))$ and $P^j(x^0(e))$, defined in (4) and (5), a feasible descent direction can still be found to move away from pseudo optimal points.

In agreement with the notation in proposition 4.2.2, define $x^0 \equiv \{x^{10}, x^{20}, \dots, x^{p0}\}$ to be a pseudo optimal point for the global problem P of (1). Let A_i be the set of indices of active constraints at x^0 in subproblem $P^i(x^0(0))$ of (4). Let R_i and I_i be two subsets of A_i with $A_i = R_i \cup I_i$ and $R_i \cap I_i = \emptyset$, defined by

R_i is the set of indices of redundant constraints of and

I_i is a set of indices of active constraints whose gradient vectors with respect to design variables are linearly independent in the design subspace R^{n_i} .

Let sets A_j , R_j and I_j for $P^j(x^0(\epsilon))$ in (5) be defined similarly. From proposition 4.3.2 and the definition 4.3.3, for a redundant constraint g_t of $P^i(x^0(0))$, $t \in R_i$, it is true that

$$\frac{\partial g_t}{\partial x^i} = \sum_{s \in I_i} \alpha_s \frac{\partial g_s}{\partial x^i}, \text{ where } \alpha_s \geq 0 \quad (23.a)$$

Similarly, for a redundant constraint g_t of $P^j(x^0(\epsilon))$, $t \in R_j$ and

$$\frac{\partial g_t}{\partial x^i} = \sum_{s \in I_j} \alpha_s \frac{\partial g_s}{\partial x^i}, \text{ where } \alpha_s \geq 0 \quad (23.b)$$

Proposition 4.3.3

If problem $P^i(x^0(0))$ of (4) is solved with all redundant constraints removed, and the right hand sides of the constraints are perturbed according to the rule

$$\epsilon_s = \begin{cases} \epsilon & \text{if } s = k \\ 0 & \text{if } s \neq k \end{cases} \quad s \in I_i,$$

then, for any redundant constraint g_t of $P^i(x^0(0))$, $t \in R_i$,

$$\frac{dg_t}{d\epsilon} \leq 0.$$

Proof

Since $t \in R_i$, ϵ does not appear explicitly in constraint g_t .

$$\begin{aligned}
\frac{dg_t}{d\epsilon} &= \frac{\partial g_t}{\partial x^i} \frac{\partial x^i}{\partial \epsilon} = \left(\sum_{s \in I_i} \alpha_s \frac{\partial g_s}{\partial x^i} \right) \frac{\partial x^i}{\partial \epsilon} \\
&= \sum_{s \in I_i} \alpha_s \left(\frac{\partial g_s}{\partial x^i} \frac{\partial x^i}{\partial \epsilon} \right)
\end{aligned} \tag{24}$$

where the second equality is obtained from (23.a). From Theorem 3.1, if g_s is active at $\epsilon=0$, g_s is identically zero in a neighborhood of $\epsilon=0$. Therefore

$$\frac{dg_s}{d\epsilon} = \frac{\partial g_s}{\partial \epsilon} + \frac{\partial g_s}{\partial x^i} \frac{\partial x^i}{\partial \epsilon} = 0 \tag{25}$$

Since the right hand sides of the constraints are perturbed according to the rule

$$\epsilon_s = \begin{cases} \epsilon & \text{if } s = k \\ 0 & \text{if } s \neq k \end{cases} \quad s \in I_i,$$

$$\frac{\partial g_s}{\partial \epsilon} = \begin{cases} 1 & \text{if } s = k \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

Combining (24), (25) and (26), we have

$$\begin{aligned}
\frac{dg_t}{d\epsilon} &= \sum_{s \in I_i} \alpha_s \left(- \frac{\partial g_s}{\partial \epsilon} \right) \\
&= -\alpha_k \leq 0
\end{aligned}$$

Propositions 4.3.3 states that if redundant constraints are removed from problem $P^i(x^0(0))$ defined by equations (4), the solutions to $P^i(x^0(0))$ without including the redundant constraints will not violate the redundant constraints. The next proposition will state a similar result for problem $P^j(x^0(\epsilon))$.

Proposition 4.3.4

If

(i). Problem $P^i(\mathbf{x}^0(0))$ of (4) is solved with all redundant constraints removed, and the right hand sides of the constraints are perturbed according to the rule

$$\varepsilon_s = \begin{cases} \varepsilon & \text{if } s = k \\ 0 & \text{if } s \neq k \end{cases} \quad s \in I_i.$$

(ii). At the solution of $P^i(\mathbf{x}^0(0))$, $\mathbf{x}^{i*}(\varepsilon)$, problem $P^j(\mathbf{x}^0(\varepsilon))$ is solved with all redundant constraints removed.

Then, for any redundant constraint g_t of $P^j(\mathbf{x}^0(\varepsilon))$, $t \in R_j$,

$$\frac{dg_t}{d\varepsilon} \leq 0.$$

Proof:

The solution to problem $P^j(\mathbf{x}^0(\varepsilon))$, $\mathbf{x}^{j*}(\mathbf{x}^{i*}(\varepsilon))$, is implicitly affected by ε through $\mathbf{x}^{i*}(\varepsilon)$, the solution to problem $P^i(\mathbf{x}^0(0))$. Thus,

$$\frac{dg_t}{d\varepsilon} = \frac{\partial g_t}{\partial \mathbf{x}^i} \frac{\partial \mathbf{x}^i}{\partial \varepsilon} + \frac{\partial g_t}{\partial \mathbf{x}^j} \frac{\partial \mathbf{x}^{j*}}{\partial \mathbf{x}^{i*}} \frac{\partial \mathbf{x}^{i*}}{\partial \varepsilon} \quad (27)$$

If $g_t(\mathbf{x}^0)$ is redundant in $P^i(\mathbf{x}^0(0))$, then

$$\frac{dg_t}{d\varepsilon} = \frac{\partial g_t}{\partial \mathbf{x}^i} \frac{\partial \mathbf{x}^i}{\partial \varepsilon} \leq 0$$

from proposition 4.3.3. Otherwise,

$$\frac{\partial g_t}{\partial \mathbf{x}^i} \frac{\partial \mathbf{x}^i}{\partial \varepsilon} = - \frac{\partial g_t}{\partial \varepsilon} \leq 0$$

by equations (25) and (26). Thus, in either case,

$$\frac{\partial g_t}{\partial x^i} \frac{\partial x^i}{\partial \varepsilon} \leq 0 \quad (28)$$

The second term in (27) can be rewritten using (23.b) as

$$\begin{aligned} \frac{\partial g_t}{\partial x^j} \frac{\partial x^{j*}}{\partial x^{i*}} \frac{\partial x^{i*}}{\partial \varepsilon} &= \left(\sum_{s \in I_j} \alpha_s \frac{\partial g_s}{\partial x^j} \right) \frac{\partial x^{j*}}{\partial x^{i*}} \frac{\partial x^{i*}}{\partial \varepsilon} \\ &= \sum_{s \in I_j} \alpha_s \left(\frac{\partial g_s}{\partial x^j} \frac{\partial x^{j*}}{\partial x^{i*}} \frac{\partial x^{i*}}{\partial \varepsilon} \right) \end{aligned} \quad (29)$$

In $P_j(x^0(\varepsilon))$ of (5), g_s is non-redundant. From Theorem 3.1, solving $P_j(x^0(\varepsilon))$ without redundant constraints, the set of active constraints remains unchanged if ε is small enough. Thus,

$$\frac{dg_s}{dx^i} = \frac{\partial g_s}{\partial x^i} + \frac{\partial g_s}{\partial x^j} \frac{\partial x^{j*}}{\partial x^{i*}} = 0 \quad (30)$$

Combining (29) and (30), we have

$$\frac{\partial g_t}{\partial x^j} \frac{\partial x^{j*}}{\partial x^{i*}} \frac{\partial x^{i*}}{\partial \varepsilon} = \sum_{s \in I_j} \alpha_s \left(- \frac{\partial g_s}{\partial x^i} \frac{\partial x^{i*}}{\partial \varepsilon} \right) = \sum_{s \in I_j} \alpha_s \left(- \frac{\partial g_s}{\partial \varepsilon} \right) \quad (31)$$

By equations (25) and (26), if $g_s(x^0)$ is redundant in $P_i(x^0(0))$, then $\frac{\partial g_s}{\partial \varepsilon} = 0$.

Otherwise, $-\frac{\partial g_s}{\partial \varepsilon} \leq 0$. Therefore, (31) becomes

$$\frac{\partial g_t}{\partial x^j} \frac{\partial x^{j*}}{\partial x^{i*}} \frac{\partial x^{i*}}{\partial \varepsilon} \leq 0 \quad (32)$$

Combining (28) and (32), $\frac{dg_t}{d\varepsilon} \leq 0$.

Propositions 4.3.4 states that if redundant constraints are removed from problem $P_j(\mathbf{x}^0(\epsilon))$ of equations (5), the solutions to $P_j(\mathbf{x}^0(\epsilon))$ with all redundant constraints removed will not violate the redundant constraints. Therefore, if there exist redundant constraints at a pseudo optimal point $\mathbf{x}^0$, a procedure based on propositions 4.2.2 and 4.2.3 can still be constructed to move away from $\mathbf{x}^0$ while improving the global objective *provided that redundant constraints are deleted first*.

4.4 A Procedure to Identify the Set of Redundant Constraints

In this section, a strategy to identify redundant constraints among the set of active constraints is described. The following notation is adopted.

Let $\mathbf{x}^0 \equiv \{\mathbf{x}^{10}, \mathbf{x}^{20}, \dots, \mathbf{x}^{p0}\}$ be an irregular point for problem $P^i(\mathbf{x}^0)$ for some i . Let A be the set of active constraints at $\mathbf{x}^0$ and let K_A be the cardinality of A , i.e., the number of active constraints at $\mathbf{x}^0$. Let R and I be two subset of A_i with $A = R \cup I$ and $R \cap I = \emptyset$, where R is the set of redundant constraints and I is the set of active constraints whose gradient vectors with respect to design variables of $P^i(\mathbf{x}^0)$ are linearly independent in the design subspace R^{n_i} . Redundancy often occurs when the number of local design variables in $P^i(\mathbf{x}^0)$, n_i , is less than the number of active constraints. Thus, we assume $n_i < K_A$. Let $\tilde{A}$ be the $n_i \times K_A$ Jacobian matrix associated with A . Column vectors of $\tilde{A}$ are gradient vectors of active constraints in A . The following procedure can be used to construct the subset R , and the subset I .

Procedure of identifying the set of redundant constraints:

Input: The set of active constraints A at $\mathbf{x}^0$ (see equations (2)).

Output: The set of redundant constraints R and the set of active constraints I whose gradient vectors with respect to design variables of $P^i(x^0)$ are linearly independent in the design subspace R^{n_i} .

Step 1. Initially, let $A^0 \equiv A$, $K_0 = K_A$, $\eta = 0$ and $R^0 \equiv \emptyset$.

Step 2. If the columns of $\tilde{A}^\eta$ are linearly independent in R^{n_i} , stop. Set $R \equiv R^\eta$ and $I \equiv A^\eta$ and output R and I .

Step 3. Pick an arbitrary constraint g_s of A^η , set $\eta = \eta + 1$.

Step 4. Let $A^\eta \equiv A^{\eta-1} - \{g_s\}$, $K_\eta = K_{\eta-1} - 1$. Consider the following system of constraints,

$$\tilde{A}^\eta u = \nabla g_s \quad (33)$$

where $u^\eta \geq 0$, $u^\eta \in R^{k_\eta}$.

If (33) has a non-trivial basic feasible solution, go to Step 5. Else, go to Step 6.

Step 5. Set $R^\eta \equiv R^{\eta-1} \cup g_s$, and go to Step 2.

Step 6. Pick an arbitrary constraint g_s of A^η , set $\eta = \eta - 1$, and go to Step 4.

To verify this procedure, note first that at each iteration, if (33) has a basic feasible solution, then ∇g_s is a linear non-negative combination of columns of $\tilde{A}^\eta$. Thus, if the process terminates, the gradient of each element of the output set R is a linear non-negative combination of gradients of elements in the output set I . R is not empty since x^0 is an irregular point for problem $P^i(x^0)$. By definition, R is the set of

redundant constraints. The termination criterion guarantees the gradient vectors of elements in set I are linearly independent in R^{n_i} .

Next, we need to show that the process will terminate. It is known that If columns of $\tilde{A}^\eta$ are not linearly independent in R^{n_i} , there must be at least one redundant constraint. Thus, there exists an active constraint g_s of A^η , such that step 4 has a non-trivial basic solution. Since the cardinality of set A is finite, the procedure will take only a finite number of steps.

In the next three sections, Section 4.5, 4.6 and 4.7, other implementation issues such as infeasible starting point, evaluation of non-local functions via convex approximations, and the use of optimum sensitivity derivatives will be discussed. A summary of the results derived in this chapter will be given in Section 4.8.

4.5 Infeasible Starting Points

The sequential solution scheme to solve the decomposed problem requires the computation of Lagrange multipliers to move away from pseudo optimal points. To obtain meaningful Lagrange multipliers, it is necessary to have a feasible reference point x^0 . If the initial design violates some of the constraints, one needs to find a feasible design first. With this in mind, instead of solving each subproblem $P^i(x^0)$ as defined in equations (2), a sequence of modified subproblems $MP^i(x^0)$ with augmented Lagrangian function as objectives are solved first, in order to find a feasible design.

Problem $MP^i(x^0)$:

Find $x^i \in X^i \subseteq R^{n_i}$ that

Minimizes

$$A(\mathbf{x}^i, \mathbf{u}) = f(\mathbf{x}^i, \mathbf{y}(\mathbf{x}^i)) \Big|_{\mathbf{x}^0} + \sum_{k \in V} (u_k g_k + r_k g_k^2) \quad (34.a)$$

Subject to

$$g_j(\mathbf{x}^i, \mathbf{y}(\mathbf{x}^i)) \Big|_{\mathbf{x}^0} \leq 0 \quad j \notin V \quad (34.b)$$

where V is a set of indices of violated constraints and $u_k \geq 0$ and $r_k > 0$ are positive parameters. Notice that the objective reduces to the exterior penalty function when $u_k = 0$.

Suppose that the feasible set for the global problem P is not empty. If r_k is large enough, $A(\mathbf{x}^i, \mathbf{u}^i)$ will be dominated by the constraint violation. A sequence of solutions of problems $MP^i(\mathbf{x}^0)$ for $i=1, \dots, p$ will move quickly towards a feasible point. As soon as a feasible point is found, the usual subproblem $P^i(\mathbf{x}^0)$ as defined in equations (2) will be solved. The use of an augmented Lagrangian function avoids the necessity of using too large penalty parameter r_k .

4.6 Convex Approximations of Non-local State Variables

The results presented above were all developed under the assumption that all state variables in $\mathbf{y}(\mathbf{x})$ can be computed in all of the subproblems, i.e., all constraints can be evaluated exactly for each subproblem. However, an important application of decomposition involves precisely problems in which only some of the state variables can be computed exactly in each of the subproblems. To deal with this kind of problem, convex approximations of non-local state variables can be used to retain feasibility within iterations.

In [D1] a strategy based on convex approximations was introduced to replace non-local constraints in decomposed problems. Convex approximations to differentiable functions were introduced by Starnes and Haftka in [S9] and have the desirable feature of being 'conservative', in the sense that, if $\tilde{g}(\mathbf{x})$ is the convex approximation of $g(\mathbf{x})$, the set $\tilde{g}(\mathbf{x}) \leq 0$ is often (but unfortunately not always) contained in the set $g(\mathbf{x}) \leq 0$. Because of this feature, convex approximations can be used to evaluate constraints that depend on non-local state variables. For example, if y_k^j is not a local state variable in problem $P^i(\mathbf{x}^0)$ (i.e., the means to evaluate y_k^j exactly are not available in P^i), the following approximation can be used instead:

$$\tilde{y}_k^j(\mathbf{x}^i; \mathbf{x}^0) = y_k^j(\mathbf{x}^0) + \sum_{r \in (+)} \left(\frac{\partial y_k^j}{\partial x_r^i} \right)_{\mathbf{x}^0} (x_r^i - x_r^{i0}) + \sum_{r \in (-)} \left(\frac{\partial y_k^j}{\partial x_r^i} \right)_{\mathbf{x}^0} \frac{x_r^{i0}}{x_r^i} (x_r^i - x_r^{i0}) \quad (35)$$

where the (+) sum is over positive derivatives while the (-) sum is over the negative derivatives of y_k^j . Concave approximations reverse (+) and (-). The notation $\tilde{y}_k^j(\mathbf{x}^i; \mathbf{x}^0)$ indicates that the approximation is defined on X^i and made with reference to the point $\mathbf{x}^0$. Problem $P^i(\mathbf{x}^0)$ in (2) would be replaced by

Problem $P^i(\mathbf{x}^0)$ (with convex approximations)

Define $\mathbf{x}^0 \equiv \{x^{10}, x^{20}, \dots, x^{p0}\}$ where the $x^{j0} \in X^j, j=1,2,\dots,p$ are fixed parameters. $\mathbf{x}^0$ is the *reference design* for the problem P^i . Let $y^i \equiv \{y_1^i, y_2^i, \dots, y_{r_i}^i\}$ be the local and $y^{j(i)} \equiv \{y_1^{j(i)}, y_2^{j(i)}, \dots, y_{r-r_i}^{j(i)}\}$ be the non-local state variables of P^i .

Find $\mathbf{x}^i \in X^i \subseteq \mathbb{R}^{n_i}$ that

$$\text{Minimizes} \quad f(\mathbf{x}^i, \mathbf{y}^i, \bar{\mathbf{y}}^{j(i)})_{|\mathbf{x}}. \quad (36.a)$$

$$\text{Subject to} \quad g(\mathbf{x}^i, \mathbf{y}^i, \bar{\mathbf{y}}^{j(i)})_{|\mathbf{x}} \leq 0, \quad g \in R^m. \quad (36.b)$$

and to the step-size constraint

$$|x_s^i - x_s^{oi}| \leq \Delta x_s^i, \quad s=1,2,\dots,n_i \quad (36.c)$$

The step-size constraint is needed to prevent motion into areas where the convex approximation is no longer valid. In practice this constraint is only needed in regions far away from the optimal point.

In many design problems f and g are monotonic functions of the state variables. The type of approximation used in (36) depends on whether the objective and the constraints are increasing or decreasing functions of $y_k^{j(i)}$. Convex approximations are used for increasing functions of $y_k^{j(i)}$ while concave approximations are preferred if the functions decrease with $y_k^{j(i)}$.

4.7 Use of Optimum Sensitivity Derivatives

In problem $P^i(\mathbf{x}^0)$ described in equations (36), the non-local state variables $y_k^{j(i)}$ are approximated using convex approximation (see (35)). If the reference point $\mathbf{x}^0$ is optimal for some subproblem P^j , then the partial derivatives used in (35) can be obtained via optimum sensitivity analysis:

$$\left(\frac{\partial y_k^j}{\partial x_r^i} \right)_{x^*} = \left(\frac{\partial y_k^j}{\partial x^j} \right)_{x^*} \left(\frac{\partial x^{j0}}{\partial x_r^i} \right)_{x^*}$$

Sensitivity derivatives of optimum solutions with respect to parameters $\left(\frac{\partial x^{j0}}{\partial x_r^i} \right)_{x^*}$ can be found in [F1] and [B2].

The use of optimum sensitivity derivatives is illustrated by a simple example. Suppose that if the objective f in the problem of Fig. 4 were not an explicit function of the local variable x_1 , for instance,

$$f(x_1, x_2) = -x_2$$

it could not be affected by optimization in the space $X^1 = \{x_1\}$. In this case f in problem P^1 must be approximated by

$$\begin{aligned} \tilde{f}(x_1, x_2^0) &= f(x_1^0, x_2^0) + \frac{df^*}{dx_1^0}(x_1 - x_1^0) = \\ &= -x_2^0 + \lambda_2^{2*}(x_1 - x_1^0) \end{aligned}$$

where the sensitivity of f^* to changes in x_1^0 in $P^2(x_1^0)$ is used and g_2 was assumed active. Optimum sensitivity derivatives of state variables $y_k^{j(i)}$ can be computed using formulas such as those developed in [S6].

4.8 Summary of Chapter 4

A general decomposition algorithm has been developed to solve the non-hierarchical problem P defined by equations (1). The procedure first decomposes

problem P into smaller subproblems $P^i(x^0)$ of the form described by equations (2), then solves each of the subproblems in sequence. Once a pseudo optimal point x^0 is found, two subproblems $P^i(x^0(0))$ and $P^j(x^0(e))$, defined by (4) and (5) are selected and solved in an order based on the relative magnitude of the Lagrange multipliers of the two subproblems. To avoid numerical difficulties, $P^i(x^0(0))$ in (4) is replaced by an equivalent problem defined by equations (12) which employs an augmented Lagrangian function. It has been proved that solutions of $P^i(x^0(0))$ and $P^j(x^0(e))$ in (12) and (5) will move away from pseudo optimal point x^0 while improving the global objective. In case that there exist redundant constraints in $P^i(x^0(0))$ and/or in $P^j(x^0(e))$, it has been shown that solutions of $P^i(x^0(0))$ and $P^j(x^0(e))$ with redundant constraints removed will still move away from pseudo optimal point x^0 while improving the global objective. Strategies dealing with infeasible initial designs and the evaluation of non-local state variables have been presented.

CHAPTER 5

NUMERICAL EXAMPLES

5.1 Example 1

The following example from [H2] (Problem 103, pg. 112) illustrates the non-hierarchical decomposition solution procedure. The effectiveness of convex approximations of non-local functions is demonstrated using this example. The augmented Lagrangian function approach to find a feasible design is also tested. The global problem P is:

Problem P: Find $\mathbf{x} \in \mathbb{R}^7$ that

minimizes

$$\begin{aligned} f = & 10 x_1 x_2^{-1} x_4^2 x_6^{-3} x_7^{0.5} + 15 x_1^{-1} x_2^{-2} x_3 x_4 x_5^{-1} x_7^{-0.5} \\ & + 20 x_1^{-2} x_2 x_4^{-1} x_5^{-2} x_6 + 25 x_1^2 x_2^2 x_3^{-1} x_5^{0.5} x_6^{-2} x_7 \end{aligned}$$

subject to

$$\begin{aligned} g_1 = & 0.5 x_1^{0.5} x_3^{-1} x_6^{-2} x_7 + 0.7 x_1^3 x_2 x_3^{-2} x_6 x_7^{0.5} \\ & + 0.2 x_2^{-1} x_3 x_4^{-0.5} x_6^{2/3} x_7^{1/4} - 1 \leq 0 \end{aligned}$$

$$g_2 = 1.3 x_1^{-0.5} x_2 x_3^{-1} x_5^{-1} x_6 + 0.8 x_3 x_4^{-1} x_5^{-1} x_6^2 \\ + 3.1 x_1^{-1} x_2^{0.5} x_4^{-2} x_5^{-1} x_6^{1/3} - 1 \leq 0$$

$$g_3 = 2 x_1 x_3^{-1.5} x_5 x_6^{-1} x_7^{4/3} + 0.1 x_2^{-1} x_3^{-0.5} x_5 x_6^{-1} x_7^{-0.5} \\ + x_1^{-1} x_2 x_3^{0.5} x_5 + 0.65 x_2^{-2} x_3 x_5 x_6^{-1} x_7 - 1 \leq 0$$

$$g_4 = 0.2 x_1^{-2} x_2 x_4^{-1} x_5^{0.5} x_7^{1/3} + 0.3 x_1^{0.5} x_2^2 x_3 x_4^{1/3} x_5^{-2/3} x_7^{1/4} \\ + 0.4 x_1^{-3} x_2^{-2} x_3 x_5 x_7^{3/4} + 0.5 x_3^{-2} x_4 x_7^{0.5} - 1 \leq 0$$

$$0.01 \leq x_i \leq 10$$

This problem was decomposed into the 3 subproblems:

P1: with local variables $x^1 = \{x_1, x_2\}$ and local functions f and g_2 .

P2: with local variables $x^2 = \{x_3, x_4\}$ and local functions f and g_1 .

P3: with local variables $x^3 = \{x_5, x_6, x_7\}$ and local functions f , g_3 and g_4 .

At the solution of the global problem as a whole, the optimum objective value is 127. The problem was solved first assuming that all functions can be evaluated exactly in all subproblems and then using convex approximations for the non-local functions. The starting point is in both cases the (feasible) point

$$x^0 = \{4.0, 0.60, 2.71, 8.74, 1.90, 1.96, 0.01\}$$

Solve P^1 , P^2 and P^3 in sequence until a pseudo optimal point is found. Propositions 4.2.2 and 4.2.3 are used to move away from pseudo optimal points. Iteration histories are shown in Fig. 12 and Fig. 13. The horizontal axis shows which problem subspace was used in the given iteration to find a new point. If two subspaces are shown, it indicates that the reference point was pseudo optimal and a motion away from an active constraint was necessary.

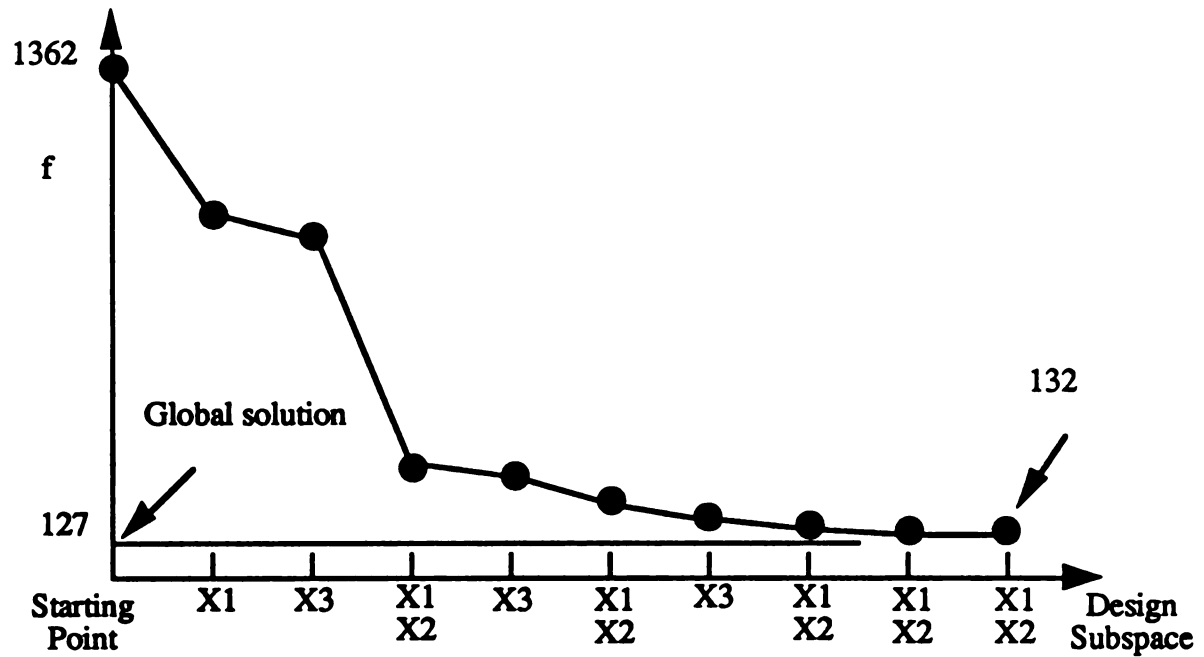


Figure 12. Iteration history for example 1: exact function

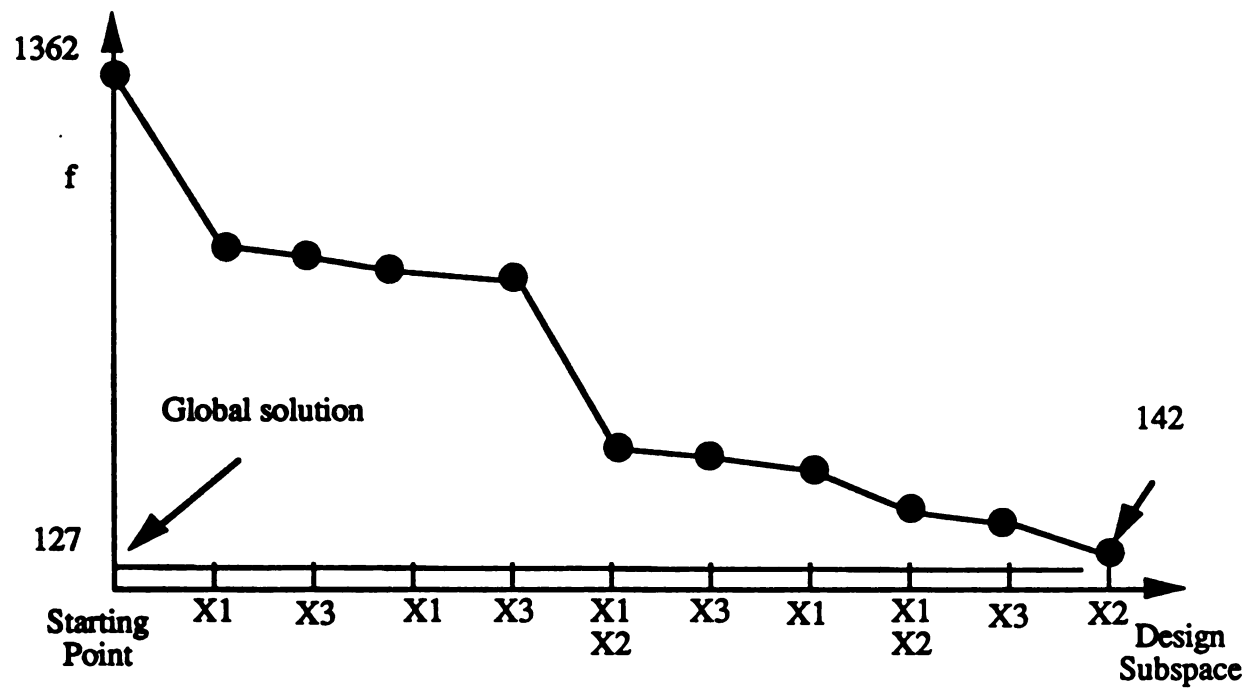


Figure 13. Iteration history for example 1: convex approximations

The effect of the convex approximations was not significant, even though no step size control was used. The algorithm was also started from an infeasible reference point and a feasible point was found using the augmented Lagrangian function in one cycle of solution P^1 - P^2 - P^3 , both using exact and approximate functions. In both cases, the algorithm terminates since the global objective improvement in several cycles of solution P^1 - P^2 - P^3 is very small.

The iteration history shows that the most significant gains are achieved in the step away from a pseudo optimal point. This may indicate that the use of the augmented Lagrangian function more often within iterations may prove fruitful.

The effect of redundancy was not felt. Constraints g_1 , g_2 and g_3 are all active at the optimum point, g_3 is degenerate in problems P^1 and P^2 . A standard generalized reduced gradient algorithm was used to solve the local subproblems.

5.2 Example 2

This example from [B4] (Alkylation problem, pg. 559-560) is used to demonstrate the effectiveness of the non-hierarchical decomposition algorithm in the presence of severe redundancy. It is deliberately chosen since it has 14 constraints and only 7 design variables. At the global optimum, there are 5 active constraints.

Global problem P

Find $x \in \mathbb{R}^7$ that

Minimizes

$$f = 1.715 x_1 + 0.035 x_1 x_6 + 4.0565 x_3 + 10 x_2 \\ - 0.063 x_3 x_5 + 3000$$

Subject to

$$g_1 = 0.005955 x_6^2 + 0.0882929 x_1^{-1} x_3 \\ - 0.117563 x_6 - 1 \leq 0$$

$$g_2 = 1.1088 x_1 x_3^{-1} + 0.130353 x_1 x_3^{-1} \\ - 0.006603 x_1 x_3^{-1} x_6^2 - 1 \leq 0$$

$$g_3 = 0.000662 x_6^2 + 0.017240 x_5 - 0.005660 x_4 \\ - 0.019121 x_6 - 1 \leq 0$$

$$g_4 = 56.7597 x_5^{-1} + 1.08702 x_5^{-1} x_6 + 0.32175 x_4 x_5^{-1} \\ - 0.037620 x_5^{-1} x_6^2 - 1 \leq 0$$

$$g_5 = 0.006198 x_7 + 246.23116 x_2 x_3^{-1} x_4^{-1} \\ - 25.1256 x_2 x_3^{-1} - 1 \leq 0$$

$$g_6 = 161.190 x_7^{-1} + 5000.0 x_2 x_3^{-1} x_7^{-1} \\ - 489510.0 x_2 x_3^{-1} x_4^{-1} x_7^{-1} - 1 \leq 0$$

$$g_7 = 4.4333 x_5^{-1} + 0.330 x_5^{-1} x_7 - 1 \leq 0$$

$$g_8 = 0.022556 x_5 - 0.007595 x_7 - 1 \leq 0$$

$$g_9 = 0.000610 x_3 - 0.0005 x_1 - 1 \leq 0$$

$$g_{10} = 0.819672 x_1 x_3^{-1} + 0.819672 x_3^{-1} - 1 \leq 0$$

$$g_{11} = 24500.0 x_2 x_3^{-1} x_4^{-1} - 250.0 x_2 x_3^{-1} - 1 \leq 0$$

$$g_{12} = 0.010204 x_4 + 0.000012 x_2^{-1} x_3 x_4 - 1 \leq 0$$

$$g_{13} = 0.0000625 x_1 x_6 + 0.0000625 x_1 - 0.00007625 x_3 - 1 \leq 0$$

$$g_{14} = 1.22 x_1^{-1} x_3 + x_1^{-1} - x_6 - 1 \leq 0$$

$$1.0 \leq x_1 \leq 2000.0$$

$$1.0 \leq x_2 \leq 120.0$$

$$1.0 \leq x_3 \leq 5000.0$$

$$85.0 \leq x_4 \leq 93.0$$

$$90.0 \leq x_5 \leq 95.0$$

$$3.0 \leq x_6 \leq 12.0$$

$$145.0 \leq x_7 \leq 162.0$$

The global problem is solved as a whole and the optimal value of objective function is 1049. At the global optimum, five constraints are active, they are: g_1 , g_3 , g_8 , g_9 and g_{12} . Problem P is decomposed into 3 subproblems:

P¹: with local variables $x^1 = \{x_1, x_5\}$ and all constraints.

P²: with local variables $x^2 = \{x_2, x_4, x_6\}$ and all constraints.

P³: with local variables $x^3 = \{x_3, x_7\}$ and all constraints.

Each subproblem is solved by assuming that all functions can be evaluated exactly. Given a feasible starting point $x^0 = \{1745, 110, 3048, 89.2, 92.8, 8, 145\}$, P¹, P² and P³ are solved in sequence. If there exist redundant constraints, they are identified and the algorithm proceeds as usual with the redundant constraints removed. The

algorithm is successfully applied to problem P to deal with redundancy at pseudo optimal points. A standard generalized reduced gradient method is used for optimization within each subspace. The algorithm terminates when the global objective improvement is very small in several cycles of solution P^1 - P^2 - P^3 . The iteration history is given in Fig. 14.

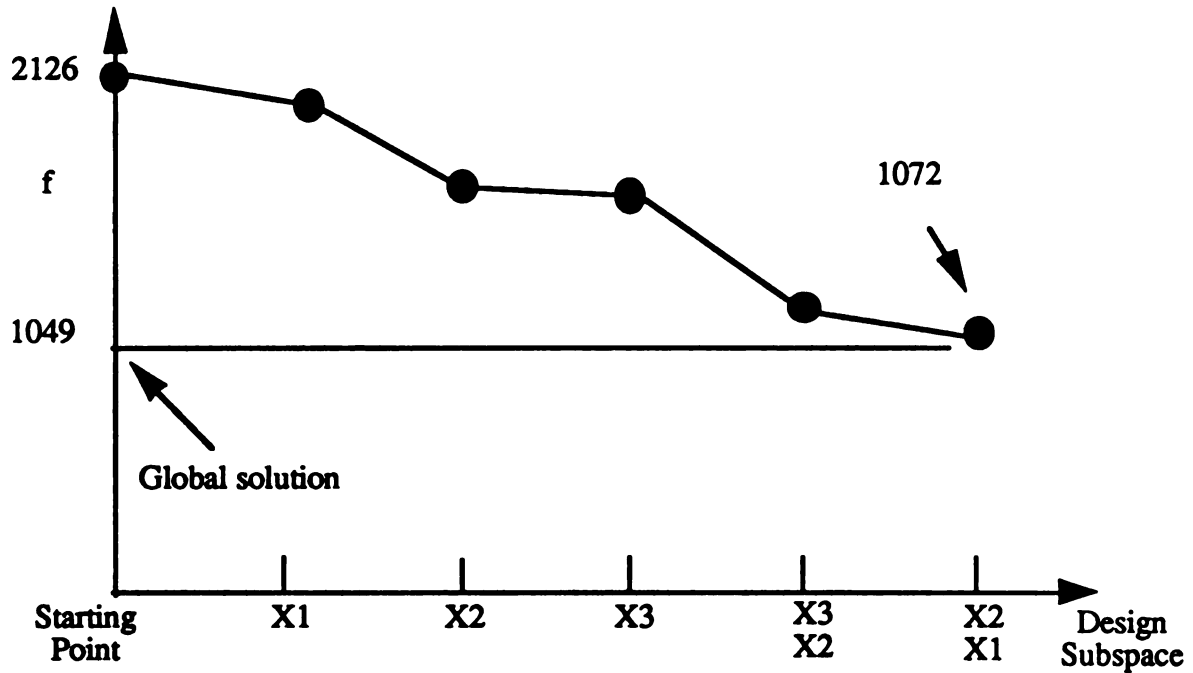


Figure 14. Iteration history for example 2: exact function

5.3 Example 3

In this section, the mathematical model for optimal design of the drive and control system described in Chapter 1 is given and the non-hierarchical optimization algorithm is applied to solved the problem. As stated in Chapter 1, the problem consists of three optimization subproblems:

P^1 : overall system design

P²: clutch detailed design

P³: speed reducer detailed design

Parameters of the electric motor are data, given as follows:

Rotary moment of inertia = 0.178 slug·ft²

Average torque of motor in operating speed range = 60 lb·ft

Minimum allowable motor speed = 151 rad/sec

Motor speed at the time of clutch engagement = 187 rad/sec

Operating speed of motor under load = 183 rad/sec

Optimization models of the three subproblems can now be formulated using the above information.

Design variables of P¹:

x₁: clutch capacity (lb·ft)

x₂: rotary moment of inertia of clutch input section (slug·ft²)

x₃: rotary moment of inertia of clutch output section (slug·ft²)

x₄: speed reducer efficiency

x₅: rotary moment of inertia of the speed reducer input section (slug·ft²)

x₆: speed ratio

Overall system behavior requirements:

(i). Clutch slippage time $\leq$ allowable value.

(ii). Minimum motor speed resulting when clutch is engaged $\leq$ minimum allowable motor speed.

(iii). Time required for the motor to accelerate to the normal operating speed $\leq$ allowable value.

(iv). Heat generated at the clutch face $\leq$ allowable value.

Mathematical expressions of the above constraints are given below.

$$g_1 = \frac{187}{B-A} - 0.25 \leq 0$$

$$g_2 = \frac{187A}{A-B} - 36 \leq 0$$

$$g_3 = \frac{187A}{(A-B)C} - \frac{4}{C} - 0.25 \leq 0$$

$$g_4 = x_1(D-E) - 1000 \leq 0$$

where

$$A = \frac{60 - x_1}{0.178 + x_2}$$

$$B = \frac{x_1 x_4}{x_3 + x_5 + 20.8 x_6^{-2}}$$

$$C = \frac{60 x_4}{0.178 + x_2 + x_3 + x_5 + 20.8 x_6^{-2}}$$

$$D = \frac{34969}{B-A} - \frac{A}{2} \left(\frac{187}{B-A} \right)^2$$

$$E = 0.5B \left(\frac{187}{B-A} \right)^2$$

It is decided to use a cone clutch in the drive and control system. The geometric configuration of the cone clutch is show in Fig. 15, where β is chosen to be 12.5° .

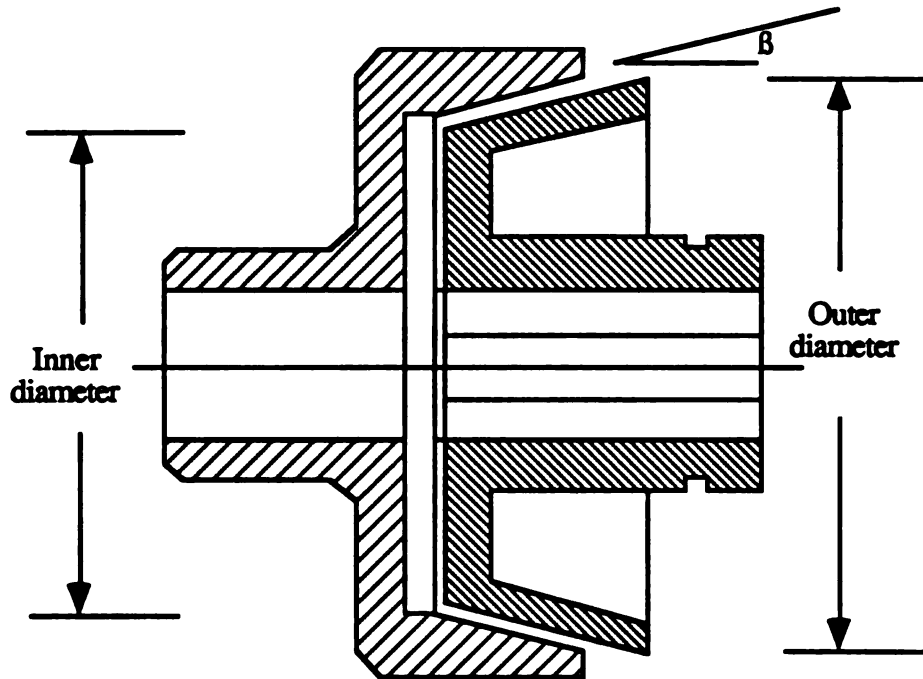


Figure 15. Conical clutch

Design variables of P^2 :

x_7 : clutch inner radius (ft)

x_8 : clutch outer radius (ft)

Clutch design requirements:

(i). Torque transmitted by clutch $\leq$ clutch capacity.

$$g_5 = x_1 - 27280.7 x_7 (x_8^2 - x_7^2) \leq 0$$

(ii). Geometric restrictions based on experience.

$$g_6 = \frac{9.24(x_8 - x_7)}{(x_8 + x_7)} - 0.5 \leq 0$$

$$g_7 = 0.3 - \frac{9.24(x_8 - x_7)}{(x_8 + x_7)} \leq 0$$

(iii). Maximum shear stress of clutch $\leq$ allowable value.

$$g_8 = \frac{17367.43x_8x_7}{(x_8^2 + x_7^2)} - 3.2 \times 10^5 \leq 0$$

(iv). Rotary moment of inertia of clutch required in P^1 equals to that computed in P^2 .

$$g_9 = 7.14(x_8 - x_7)(x_8 + x_7)^4 - x_2 = 0$$

$$g_{10} = 7.14(x_8 - x_7)(x_8 + x_7)^4 - x_3 = 0$$

A worm gear drive is selected as the speed reducer in the drive and control system. The pressure angle normal to the gear teeth is chosen to be 25° .

Design variables of P^3 :

x_9 : worm pitch diameter (ft)

x_{10} : gear pitch diameter (ft)

x_{11} : gear face width (ft)

x_{12} : gear module (ft)

Design requirements for the speed reducer are:

(i). Rotary moment of inertia of speed reducer required in P^1 equals to that computed in P^3 .

$$g_{11} = 19.88x_9^4x_{10}x_6 - x_5 = 0$$

(ii). Speed reducer efficiency required in P^1 equals to that computed in P^3 .

$$g_{12} = x_4 - \frac{0.91 - 0.0286x_{10}x_6^{-1}x_9^{-1}}{0.91 + 0.0286x_6x_9x_{10}^{-1}} = 0$$

(iii). Gear bending stress $\leq$ allowable value.

$$g_{13} = 11.25x_1x_{10}^{-1}x_{11}^{-1}x_{12}^{-1} - 4.26 \times 10^5 \leq 0$$

(iv). gear contact fatigue strength requirement.

$$g_{14} = 3.7258 \times 10^4 \sqrt{\frac{23.5x_1(1+x_6)}{x_6x_{10}^2x_{11}}} - 7.13 \times 10^6 \leq 0$$

(v). Geometric restrictions based on experience.

$$g_{15} = x_9 - \frac{F}{1.7} \leq 0$$

$$g_{16} = \frac{F}{3} - x_9 \leq 0$$

where

$$F = \left(\frac{x_9 + x_{10}}{2} \right)^{0.875}$$

All three subproblems seek to minimize the weight (lb) of the drive and control system.

$$\begin{aligned} \text{Weight} = & 21.45(x_2 + x_3)(x_8 + x_7)^{-2} + 128.7x_5x_9^{-2} + 407.15x_{10}^2x_{11} \\ & + 1836.54(x_8 - x_7)(x_8 + x_7)^2 + 2558.2x_9^4x_{10}x_6 \end{aligned}$$

Bounds on design variables of the three subproblems are given as follows:

$$100 \leq x_1 \leq 1000$$

$$0.01 \leq x_2 \leq 200$$

$$0.01 \leq x_3 \leq 200$$

$$0.01 \leq x_4 \leq 1$$

$$0.01 \leq x_5 \leq 200$$

$$1 \leq x_6 \leq 400$$

$$0.01 \leq x_7 \leq 2$$

$$0.01 \leq x_8 \leq 2$$

$$0.01 \leq x_9 \leq 5$$

$$0.01 \leq x_{10} \leq 5$$

$$0.01 \leq x_{11} \leq 1$$

$$0.01 \leq x_{12} \leq 1$$

Starting from a feasible point $x^0 = \{130.443, 0.0238, 0.0305, 0.6835, 0.01, 73.3351, 0.274, 0.3043, 0.3379, 1.6937\}$, P^1 , P^2 and P^3 are solved in sequence including all constraints in each subproblem. First, each subproblem is solved by assuming that all functions can be computed exactly. The iteration history is shown in Fig.16. At the above starting point, each subproblem is also solved by assuming that only local functions can be evaluated exactly, while non-local functions are replaced by convex approximations. The iteration history is given in Fig.17.

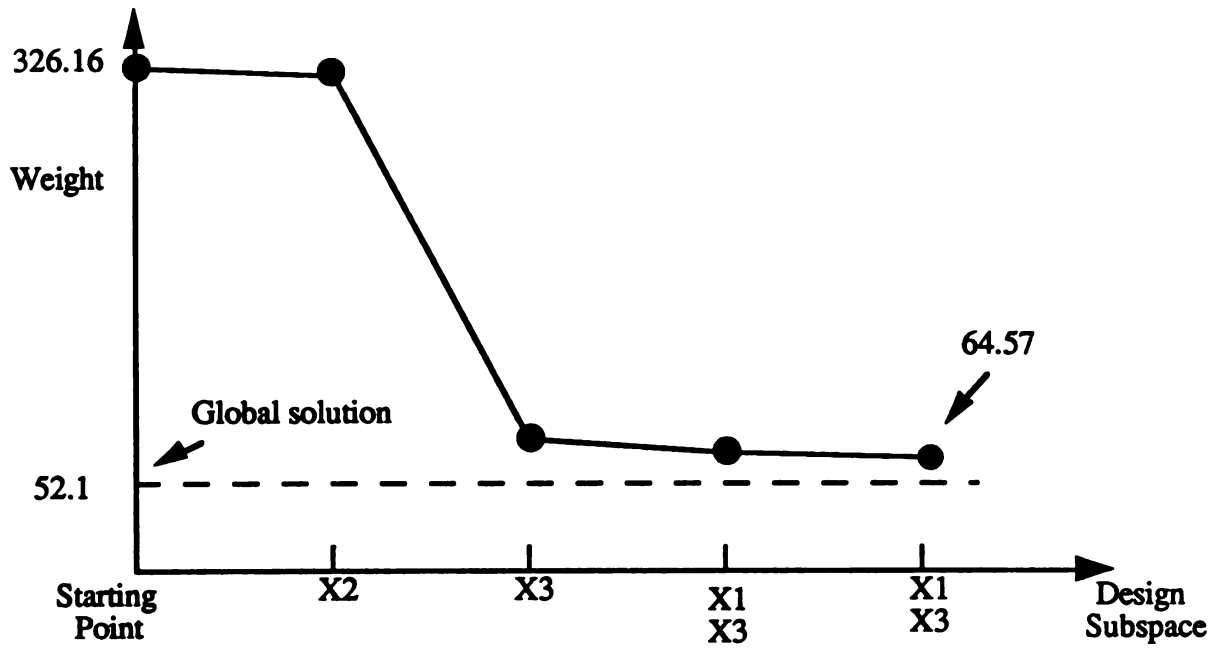


Figure 16. Iteration history for example 3: exact function

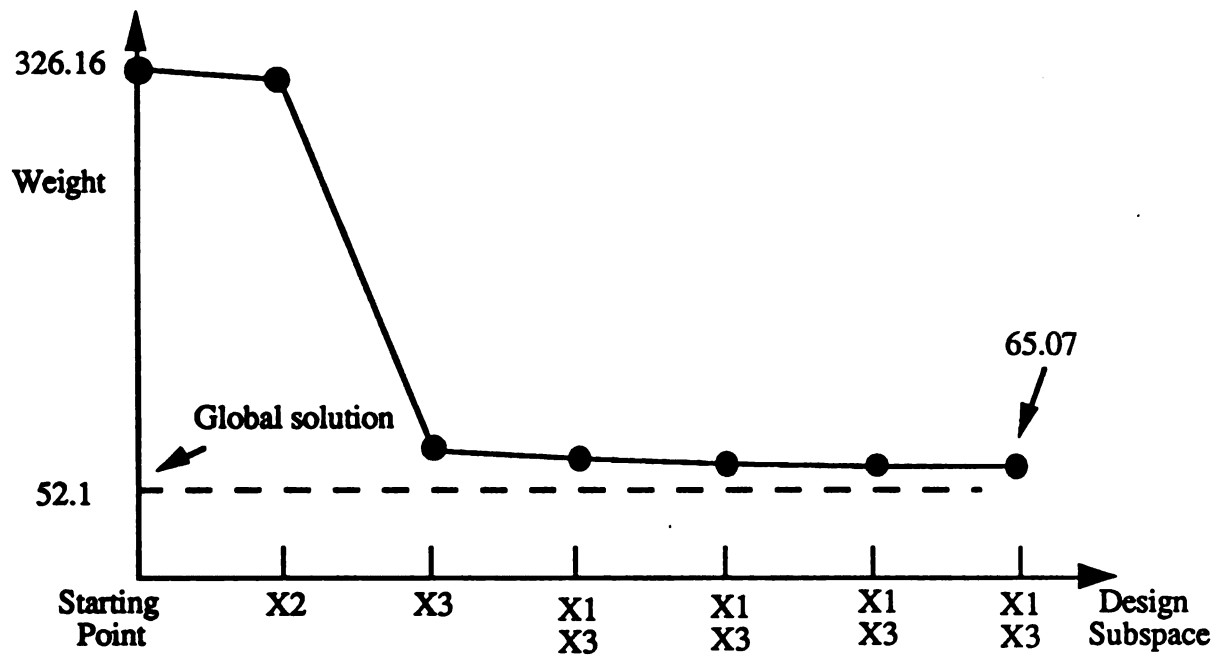


Figure 17. Iteration history for example 3: convex approximations

CHAPTER 6

REMARKS AND CONCLUSIONS

The solution of a global problem P of (1) was obtained in the thesis by changing only a subset of the global design space at each step. An algorithm to solve the decomposed problem was developed based on rigorous analytical results. The algorithm requires the sequential solution of all subproblems and leads to global convergence. The concept of pseudo optimal point has been introduced to represent the fact that a point which satisfies local optimality conditions is not necessarily the global optimum solution. In fact, in a non-hierarchical optimization problem, there are infinite number of pseudo optimal points. A strategy to move away from pseudo optimal points while still improving the global objective was discussed. It should be noted that redundancy often occurs in the global design space decomposition if all constraints are included in each subproblem optimization. In the presence of redundancy, the overall process proceeds as usual after removal of all the redundant constraints in each subproblem optimization. This is verified by the second example. Numerical experience shows that the augmented Lagrangian function method seems to be a good choice in minimization of constraint violations. Convex approximation of non-local state variables works well to insure the global feasibility in the example problems.

It is observed from the numerical experience that the global objective always achieves a large improvement when moving away from pseudo optimal points. This phenomenon suggests that future research direction may seek to explore this feature to speed up convergence.

It should be realized that the sensitivity theorems in Chapter 3 require that the set of active constraints remains unchanged when the parameters vary from their initial values. Therefore, propositions 4.2.2 will theoretically fail if the set of active constraints changes when solving $P^j(x^0(0))$ (see equations (4)), or $P^j(x^0(\epsilon))$ (equations (5)). However, numerical experience shows that the strategy based on proposition 4.2.2 to move away from pseudo optimal points while improving the global objective is still useful *even* in the case when the set of active constraints changes.

From the analytical results derived in Chapter 4, it is clear that optimum sensitivity derivatives and Lagrange multipliers of each subsystem are essential to represent the coupling among subsystems of a non-hierarchical problem. It is conceivable that some interpretation of optimum sensitivity derivatives and Lagrange multipliers of each subsystem may lead to an algorithm that allows the simultaneous solution of all subproblems, which will greatly increase the speed of non-hierarchical system optimization.

The sequential solution of each subproblem will not be as efficient as the solution of the global problem as a whole. Furthermore, the identification of redundant constraints for each subproblem at a pseudo optimal point adds more computational burden to the decomposition algorithm presented in the thesis. However, the algorithm will show its merit in the optimization problems that *must* be solved in their decomposed form. If all subproblems could be solved simultaneously, the speed of the

decomposition algorithm would be greatly increased. Further research is required to realize this ultimate goal.

APPENDIX

APPENDIX

This appendix lists several standard results in optimization theory that are used throughout the thesis, and which are not included in Chapter 3 for brevity.

Consider the general problem of determining a local solution $\mathbf{x}(\boldsymbol{\varepsilon})$ of problem $P^1(\boldsymbol{\varepsilon})$:

Problem $P^1(\boldsymbol{\varepsilon})$:

Find $\mathbf{x} \in \mathbb{R}^n$ that

Minimizes $f(\mathbf{x}, \boldsymbol{\varepsilon})$

Subject to $g_i(\mathbf{x}, \boldsymbol{\varepsilon}) \leq 0 \quad i = 1, \dots, m$

where $\boldsymbol{\varepsilon}$ is a parameter vector in $\mathbb{R}^k$. The Lagrangian function of $P^1(\boldsymbol{\varepsilon})$ is:

$$L(\mathbf{x}, \mathbf{u}, \boldsymbol{\varepsilon}) \equiv f(\mathbf{x}, \boldsymbol{\varepsilon}) + \sum_{i=1}^m u_i g_i(\mathbf{x}, \boldsymbol{\varepsilon})$$

Theorem 1. Second-order sufficient optimality conditions for a strict local solution of problem $P^1(\boldsymbol{\varepsilon})$ [F1]

If the functions defining problem $P^1(0)$ are twice continuously differentiable in a neighborhood of $\mathbf{x}^*$, then $\mathbf{x}^*$ is a strict local minimizing point of problem $P^1(0)$ (i.e., there is a neighborhood of $\mathbf{x}^*$ such that there does not exist any feasible $\mathbf{x} \neq \mathbf{x}^*$ where

$f(\mathbf{x}, 0) \leq f(\mathbf{x}^*, 0)$ if there exist Lagrange multiplier vector $\mathbf{u}^* \in \mathbb{R}^m$ such that the first Karush-Kuhn-Tucker conditions hold, i.e.,

$$g_i(\mathbf{x}^*, 0) \leq 0 \quad i = 1, \dots, m$$

$$u_i^* g_i(\mathbf{x}^*, 0) = 0 \quad i = 1, \dots, m$$

$$u_i^* \geq 0 \quad i = 1, \dots, m$$

$$\nabla L(\mathbf{x}^*, \mathbf{u}^*, 0) \equiv \nabla f(\mathbf{x}^*, 0) + \sum_{i=1}^m u_i^* \nabla g_i(\mathbf{x}^*, 0)$$

and, further, if

$$\mathbf{z}^T \nabla_{\mathbf{x}}^2 L(\mathbf{x}^*, \mathbf{u}^*, 0) \mathbf{z} > 0 \quad \text{for all } \mathbf{z} \neq 0 \text{ and } \mathbf{z} \neq 0 \text{ such that}$$

$$\nabla g_i(\mathbf{x}^*, 0) \mathbf{z} \geq 0 \quad \text{for all } i, \text{ where } g_i(\mathbf{x}^*, 0) = 0$$

$$\nabla g_i(\mathbf{x}^*, 0) \mathbf{z} = 0 \quad \text{for all } i, \text{ where } u_i^* > 0$$

An important special case of problem $P^1(\boldsymbol{\varepsilon})$ is the right-hand-side perturbations of the constraints. For this case problem $P^1(\boldsymbol{\varepsilon})$ is reduced to problem $P^2(\boldsymbol{\varepsilon})$:

Problem $P^2(\boldsymbol{\varepsilon})$:

Find $\mathbf{x} \in \mathbb{R}^n$ that

Minimizes $f(\mathbf{x})$

Subject to $g_i(\mathbf{x}) \leq \varepsilon \quad i = 1, \dots, m$

where $\boldsymbol{\varepsilon}$ is a parameter vector in $\mathbb{R}^k$. The Lagrangian of $P_2(\boldsymbol{\varepsilon})$ is defined as:

$$L(\mathbf{x}, \mathbf{u}, \boldsymbol{\varepsilon}) \equiv f(\mathbf{x}) + \sum_{i=1}^m u_i (g_i(\mathbf{x}) - \varepsilon)$$

Theorem 2. First-order changes in the optimal value function for $P^2(\epsilon)$ [F1]

If

(i) The functions defining $P^2(\epsilon)$ are twice continuously differentiable in x in a neighborhood of x^* .

(ii) The second-order sufficient conditions for a local minimum of $P^2(0)$ hold at x^* , with associated Lagrange multipliers u^* .

(iii) The gradients $\nabla_x g_i(x^*)$ (for i such that $g_i(x^*)=0$) are linearly independent.

(iv) $u_i^* > 0$ when $g_i(x^*)=0$ ($i=1, \dots, m$) (i.e., strict complementary slackness holds).

then,

(a) x^* is a local isolated minimizing point of problem $P^2(0)$ and the associated Lagrange multiplier u^* is unique.

(b) For ϵ in a neighborhood of 0, there exists a unique, once continuously differentiable vector function $y^*(\epsilon)=[x^*(\epsilon), u^*(\epsilon)]^T$ satisfying the second-order sufficient conditions for a local minimum of problem $P^2(0)$ such that $y^*(0)=[x^*, u^*]=y^*$, and hence $x^*(\epsilon)$ is a locally unique minimum of problem $P^2(\epsilon)$ with associated unique Lagrange multipliers $u^*(\epsilon)$.

(c) For ϵ in a neighborhood of 0,

$$f^*(\epsilon) = L[x^*(\epsilon), u^*(\epsilon), \epsilon] = L^*(\epsilon)$$

(d) For ϵ near 0, the set of binding inequalities is unchanged, strict complementary slackness holds, and the binding constraint gradients are linearly independent at $x^*(\epsilon)$.

(e) For ϵ in a neighborhood of 0 the gradient of the optimal value function is $\nabla_{\epsilon} f^*(\epsilon) = u^*(\epsilon)$.

Let S be an arbitrary set in R^n . The *convex hull* of S , denoted by $H(S)$, is the collection of all *convex combinations* of elements of S , i.e., a vector $a \in H(S)$ if and only if a can be expressed as

$$a = \sum_{s=1}^N \beta_s b_s$$

$$\sum_{s=1}^N \beta_s = 1 \text{ and } \beta_s \geq 0 \text{ for } s = 1, \dots, N$$

where N is a positive integer and $b_1, b_2, \dots, b_N \in S$. The following theorem gives an upper bound for the integer N .

Theorem 3. Caratheodory Theorem [B3]

Let S be an arbitrary set in R^n . If $a \in H(S)$, then a can be expressed as

$$a = \sum_{s=1}^{n+1} \beta_s b_s$$

where $\sum_{s=1}^{n+1} \beta_s = 1$ and $\beta_s \geq 0$ for $s = 1, \dots, n+1$ and $b_1, b_2, \dots, b_{n+1} \in S$.

BIBLIOGRAPHY

BIBLIOGRAPHY

[B1] Barthelemy, J. M. and Riley, M.F., "Improved Multilevel Optimization Approach for the Design of complex Engineering Systems," AIAA J., Vol. 26, No. 3, March 1988, pp. 353-360.

[B2] Barthelemy, J. M. and Sobieszczanski-Sobieski, J., "Optimum Sensitivity Derivatives of Objective Functions in Nonlinear programming," AIAA J., Vol. 21, No. 6, June 1983, pp913-915.

[B3] Bazaraa, M.S. and Shetty, C.M., "Nonlinear Programming, Theory and Algorithms," John Wiley and Sons, 1979.

[B4] Beightler, C. S. and Phillips, D. T., "Applied Geometric Programming," John Wiley and Sons, 1976.

[D1] Díaz, A. R., "Fuzzy Set Applications in Engineering Optimization. Multilevel Fuzzy Optimization," in Second NASA/Air Force Symposium On Recent Experiences in Multidisciplinary Analysis and Optimization. Langley Research Center, Hampton, Virginia, September 28-30, 1988.

[F1] Fiacco, A.V., "Introduction to Sensitivity and Stability Analysis in Nonlinear Programming," Academic Press, New York, 1983.

[G1] Geoffrion, A. M., "Elements of Large-scale Mathematical Programming," Part I and II, Management Science, Vol. 16, No. 11, July, 1970, pp652-691.

[H1] Haftka, R. T., "An Improved Computational Approach for Multilevel Optimum Design," J. Struct. Mechanics, Vol 12(2), 1984, pp. 245-261.

[H2] Hock, W. and Schittkowski, K., "Test Examples for Nonlinear Programming Codes," Springer Verlag, New York, 1981.

[K1] Kirsch, U., Reiss, M. and Shamir, U., "Optimum Design by Partitioning into Substructures," J. of Structure Division, ASCE, Jan. 1972, pp. 249-267.

[L1] Lasdon, L. S., "Duality and Decomposition in Mathematical Programming," IEEE Transactions on System and Cybernetics, Vol. SSC-4, No. 2, July 1968, pp. 86-100.

[M1] Minoux, M., "Mathematical Programming Theory and Algorithms," John Wiley and Sons, 1986.

[N1] Nachane, D. M., "Optimization Methods in Multilevel System: A Methodological Survey," European Journal of Operational Research, 21(1984), pp25-38.

[P1] Padula, S.L. and Sobieszczanski-Sobieski, J., "A Computer Simulator for Development of Engineering System Design Methodologies," ICED 87, Boston, MA, August 1987.

[P2] Parkinson, A., Balling, R., Wu, A. and Free, J. , "A General Strategy for Decomposing Large Design Problems Based on Optimization and Statistical Test Plans," ICED 87, Boston, MA, August 1987.

[R1] Rockafellar, R.T., "Convex Analysis," Princeton University Press, Princeton, New Jersey, 1970.

- [S1] Schmit Jr., L.A. and Ramanathan, R.K., "Multilevel Approach to Minimum Weight Design including Bucking Constraints," AIAA J., Vol. 16, No. 2, Feb. 1978, pp97-104.
- [S2] Schrijver, A., "Theory of Linear and Integer Programming," John Wiley and Sons, 1986.
- [S3] Sobieszczanski-Sobieski, J. , "Optimization by Decomposition: A Step from Hierarchic to Non-Hierarchic Systems", in Second NASA/Air Force Symposium On Recent Experiences in Multidisciplinary Analysis and Optimization. Langley Research Center, Hampton, Virginia, September 28-30, 1988.
- [S4] Sobieszczanski-Sobieski, J., "A linear Decomposition Method for Large Optimization Problems- Blueprint for Development," NASA TM-83248, 1982.
- [S5] Sobieszczanski-Sobieski, J., "Sensitivity Analysis and Multidisciplinary Optimization for Aircraft Design: Recent Advances and Results," 16th Congress of the Intn'l Council of the Aeronautical Sciences, Jerusalem, Israel, 1988.
- [S6] Sobieszczanski-Sobieski, J., "On the Sensitivity of Complex, Internally Coupled Systems," AIAA Paper 88-2378, AIAA/ASME/AHS 29th Structures, Structural Dynamics and Materials Conference, Williamsburg, VA, April 18-20, 1988.
- [S7] Sobieszczanski-Sobieski, J., Bloebaum, C.L. and Hajela, P., "Sensitivity of Control-Augmented Structure Obtained by a System Decomposition Method," AIAA Paper 88-2205, AIAA/ASME/AHS 29th Structures, Structural Dynamics and Materials Conference, Williamsburg, VA, April 18-20, 1988.
- [S8] Sobieszczanski-Sobieski, J., James, B.B. and Riley, M.F., "Structural Sizing by Generalized, Multilevel Optimization, AIAA J. Vol. 25, No. 1, January 1987, pp. 139.

[S9] Spotts, M.F., "Design of Machine Elements," 6th Edition, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1985.

[S10] Starnes, J.H. and Haftka, R.T., "Preliminary Design of Composite Wings for Buckling, Strength, and Displacement Constraints," J. Aircraft, 16(8), 1979, pp. 564-570.

[W1] West, P. and Noon, W. D. , "Determine the Operating Characteristics of a Drive and Control System for a Transfer Mechanism," General Motors Engineering Journal, Oct.-Nov.-Dec., 1957.

[W2] West, P. and Noon, W. D., "Solution to the Previous Problem: Determine the Operating Characteristics of a Drive and Control System for a Transfer Mechanism," General Motors Engineering Journal, Jan.-Feb.-Mar., 1958.

[W3] Wrenn, G.A. and Dovi, A.R., "Multilevel Decomposition Approach to the Preliminary Sizing of a Transport Aircraft Wing," AIAA/ASME/ASCE/AHS 28th Structures, Dynamics, and Materials Conference, Monterey, Cal., April 6-8, 1987, AIAA Paper No.87-0714-CP.

MICHIGAN STATE UNIV. LIBRARIES



31293006279610