AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92 NEUILLY-SUR-SEINE FRANCE

AGARDograph No. 149

on

Structural Design Applications of Mathematical Programming Techniques

Edited by

G. G. Pope and L. A. Schmit

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY ON BACK COVER

UNLIMITED

AGARDograph 149-71

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

STRUCTURAL DESIGN APPLICATIONS OF MATHEMATICAL PROGRAMMING TECHNIQUES

edited by

G.G.Pope and L.A.Schmit

2 1971

This AGARDograph was sponsored by the Structures and Materials Panel of AGARD-NATO

The material in this publication has been reproduced directly from copy
supplied by AGARD

Published February 1971

624.07:681.3.06

Printed by Technical Editing and Reproduction Ltd
Harford House, 7-9 Charlotte St. London, W1P 1HD.

FOREWORD

The Structures and Materials Panel of the Advisory Group for Aerospace Research and Development (AGARD) comprises scientists, engineers and technical administrators from government, universities and industry, who are concerned with the advancement of aerospace research and development and with the provision of data necessary for the design and fabrication of the vehicles and equipment which NATO requires. The Panel provides a mechanism for discussion, the exchange of information and for conducting co-operative theoretical and experimental studies in selected areas.

This volume describes the present state of development of the use of mathematical programming techniques in the optimum design of aerospace and similar structures. Although optimization with respect to cost is considered when possible, the main emphasis is on the minimization of weight, due to the overwhelming importance of this parameter in aerospace applications, and also due to the fact that it is one of the few merit functions that can be defined with reasonable precision. The use of mathematical programming techniques in the selection of materials is also discussed to the limited extent meaningful at the present time.

The text is divided into four main sections, the first of which describes basic ideas, reviews the literature, and indicates the relationship of mathematical programming methods both to practical optimization techniques of a more traditional kind, and to relevant aspects of the classical theory of least weight design. Fundamental concepts are introduced first in the context of simple examples for the benefit of newcomers to the field and are subsequently re-expressed in a general form.

The second section consists of three chapters on the algorithmic methods available for the solution of mathematical programming problems, and the third section describes some of the more ambitious applications to date of some of these techniques in the structural design context.

The fourth and final section is devoted to classes of application which are still at a relatively early stage of development but which promise to be fruitful in the future in the design of practical structures. Optimum design based on considerations of reliability - a subject of great importance - is considered in the opening chapter. This is followed by a chapter on optimization in the presence of aeroelastic constraints which includes some material on classical variational methods that is used in simple examples to illustrate a number of subtleties of optimization in that field. The volume concludes with a consideration of the optimum design of aerospace vehicles in a broader context to demonstrate that structural optimization is but one small sub-field of the areas of aerospace design where mathematical programming techniques are potentially useful.

To assist the reader the editors have imposed a degree of uniformity on the notation and conventions employed by the various contributors. They have, however, refrained from enforcing strict conformity when, in their opinion, authors have introduced variations which are unlikely to cause difficulty. Such variations are most frequent in Section 4, which covers ground well outside the confines of the earlier chapters.

The AGARD Structures and Materials Panel first became active in the field of structural optimization early in 1967 and its work in this subject will not be complete for some time yet. In addition to the preparation of this volume, for which the Panel was indeed fortunate to have the services of Prof. Schmit and Dr. Pope as editors, a major symposium was held at Istanbul in the fall of 1969 organised by Dr. R. A. Gellatly. The Panel is also fortunate in being able to delegate the management of its interest in this subject to an expert working group, chaired, first of all, by Mr. A. N. Rhodes (UK) and at present under the chairmanship of Lt. Col. C. K. Grimes (USA).

Anthony J. Barrett
Chairman
AGARD, Structures &
Materials Panel

List of Contributors

Professor H. Ashley	Department of Aeronautics and Astronautics, Stanford University, Palo Alto, California, U.S.A.
Dr. R. L. Fox	School of Engineering, Case Western Reserve University, Cleveland, Ohio, U.S.A.
Dr. J. S. Kowalik	Computer Services Division, The Boeing Company, Seattle, Washington, U.S.A.
Dr. S. C. McIntosh	Department of Aeronautics and Astronautics, Stanford University, Palo Alto, California, U.S.A.
Dr. F. Moses	School of Engineering, Case Western Reserve University, Cleveland, Ohio, U.S.A.
Dr. G. G. Pope	Structures Department, Royal Aircraft Establishment, Farnborough, Hants, U.K.
Professor L. A. Schmit	School of Engineering and Applied Science, University of California, Los Angeles, California, U.S.A.
Mr. B. Silver	Department of Aeronautics and Astronautics, Stanford University, Palo Alto, California, U.S.A.
Dr. W. H. Weatherill	Commercial Airplane Group, The Boeing Company, Renton, Washington, U.S.A.

List of Contents

SECTION I

FUNDAMENTAL CONCEPTS AND LITERATURE REVIEW

List of Contents (Contd.)

SECTION II

ALGORITHMIC TOOLS

List of Contents (Contd.)

List of Contents (Contd.)

SECTION I

FUNDAMENTAL CONCEPTS AND LITERATURE REVIEW

Chapter 1

INTRODUCTION AND BASIC CONCEPTS

by

L. A. Schmit and G. G. Pope

1.1 Introduction

During the last decade, the use of large scale digital computing facilities for structural analysis has become commonplace. This has led rather naturally to a growing interest in the application of digital computers to other quantifiable portions of the structural design process. The combining of computer oriented structural analysis techniques with mathematical programming methods has played a central role in the development of automated procedures for directed redesign. While automated procedures for structural design embrace some form of structural analysis as a subroutine, they must be recognized as only a component part of the overall design process.

It is useful to distinguish between conceptual design, computer aided design and automated procedures for directed redesign. Conceptual design is characterized by ingenuity and creativity and it deals with the overall planning of a system to serve its functional purposes. Computer aided design involves man-machine interactions and it is characterized by qualitative judgments based on externally displayed quantitative information. Automated procedures for directed redesign seek a balanced optimum design in a defined sense and they are characterized by preprogrammed logical decisions based upon internally stored quantitative information. In computer aided design, the use of graphical input-output devices such as oscilloscope display units and light-pens facilitate crossing the man-machine interface. Automated procedures for directed redesign are aimed at keeping the quantifiable portion of the design procedure in the machine and thus avoiding the unnecessary crossing of this interface. These two approaches to the effective use of the large amount of information generated by modern structural analysis methods are not mutually exclusive, but rather they complement and reinforce one another. The portion of the structural design process that can be automated responsibly has moved forward rapidly during the past decade and continued advances are anticipated for the immediate future.

1.2 Basic Concepts

The basic ideas that are fundamental to understanding structural design applications of mathematical programming methods can be introduced by considering two elementary examples. Graphical illustrations will be employed to help fix ideas and mathematical abstraction and the associated generality will be avoided for the present.

1.2.1 Simply Supported Column

Consider a simply supported column with a uniform annular cross section (Fig.1.1) subject to a compressive load of $P = 5000$ lb. Let the length $\ell = 100$ in, the modulus of elasticity $E = 10 \times 10^6$ lb/in^2 and the density $\rho = 0.1$ lb/in^3. The mean diameter is denoted by $D = (D_o + D_i)/2$ where D_o and D_i are respectively the external and internal diameter, and the wall thickness of the tube is denoted by T. Find D, T and the weight W of the minimum weight design such that $D \leq 3.5$ in, $T \geq 0.04$ in; the compressive stress in the member is to be equal to or less than 20000 lb/in^2, and the design must be such that neither Euler buckling nor local buckling can occur.

At the outset, note that the length of the column and the material have been preassigned and that only the mean diameter and the wall thickness are variables to be determined. Note also that only one load condition is given, namely $P = 5000$ lb. Thus, the possibility of various lateral loads acting in combination with P is ignored. The region of all possible positive values of D and T can be viewed geometrically as shown in Fig.1.2. Note that the region is immediately reduced by excluding values of $D > 3.5$ in (line a-a) and excluding values of $T < 0.04$ in (line b-b). It should also be noted that the internal diameter $D_i = D - T$, and since the minimum geometrically realizable value of D_i is zero, the region to the left and above the line $D = T$ (line c-c) is also excluded. The requirement that Euler buckling be precluded is stated as follows:

$$\sigma - \sigma_e \leq 0 \tag{1-1}$$

where σ denotes the stress caused by the applied load P, that is

$$\sigma = \frac{P}{\pi DT} \tag{1-2}$$

and σ_e represents the Euler buckling stress

$$\sigma_e = \frac{\pi^2 E}{8\ell^2}(D^2 + T^2) \quad . \tag{1-3}$$

Assuming $T \ll D$ and substituting the given numerical values $\ell = 100$ in, $P = 5000$ lb, and $E = 10 \times 10^6$ lb/in^2, the curve d-d in Fig.1.2 along which the actual stress equals the Euler buckling stress is defined by the equation

$$\frac{5000}{\pi DT} - 125\,\pi^2 D^2 = 0 \tag{1-4}$$

where T^2 is neglected as small compared with D^2. The region to the left and below the curve d-d in Fig.1.2 is therefore excluded in order to avoid Euler buckling. The requirement that local buckling of the thin walled tube be precluded is stated as follows:

$$\sigma - \sigma_c \leqslant 0 \tag{1-5}$$

where σ_c denotes the local buckling stress which is assumed to be given by the following simple expression

$$\sigma_c = \frac{0.4\,ET}{D} . \tag{1-6}$$

Substituting the given numerical values $P = 5000$ lb and $E = 10 \times 10^6$ lb/in^2 the line along which the actual stress equals the local buckling stress as given by the equation

$$\frac{5000}{\pi DT} - 4 \times 10^6 \frac{T}{D} = 0 \tag{1-7}$$

which is essentially equivalent to

$$T - 0.02 = 0 \tag{1-8}$$

since D and T are necessarily non-zero and positive. The region below the straight horizontal line e-e given by Eq. (1-8) is therefore excluded in order to avoid local buckling. Note that this constraint is in fact less restrictive than the minimum gauge requirement that $T \geqslant 0.04$ in. The requirement that the stress in the member be equal to or less than 20000 lb/in^2 is stated as follows:

$$\sigma - 20000 \leqslant 0 . \tag{1-9}$$

The curve f-f in Fig.1.2 along which the member stress equals 20000 lb/in^2 is given by

$$\frac{5000}{\pi DT} - 20000 = 0 . \tag{1-10}$$

The region below and to the left of the curved line f-f in Fig.1.2 is excluded therefore, in order to prevent stress in excess of 20000 lb/in^2. Note that this constraint is less restrictive than the Euler buckling constraint in the region of interest. The weight of the tubular column member is expressed as follows:

$$W = \rho\,\ell\,\pi\,DT = 10\,\pi\,DT . \tag{1-11}$$

The line g-g in Fig.1.2 along which the weight equals 4 lb, is given by the equation

$$4 - 10\,\pi\,DT = 0 \tag{1-12}$$

and a second contour (h-h) along with the weight equals 6 lb is plotted using the expression

$$6 - 10\,\pi\,DT = 0 . \tag{1-13}$$

It is apparent from Fig.1.2 that the minimum weight design satisfying the various stated limitations lies at point j (D = 3.2 in, T = 0.04 in, W = 4.0 lb). Fig.1.2 is a geometric representation of this simple two variable optimum design problem. By plotting the constraints and contours of constant weight we may scan the entire set of possible designs, points in the (D, T) space, and immediately

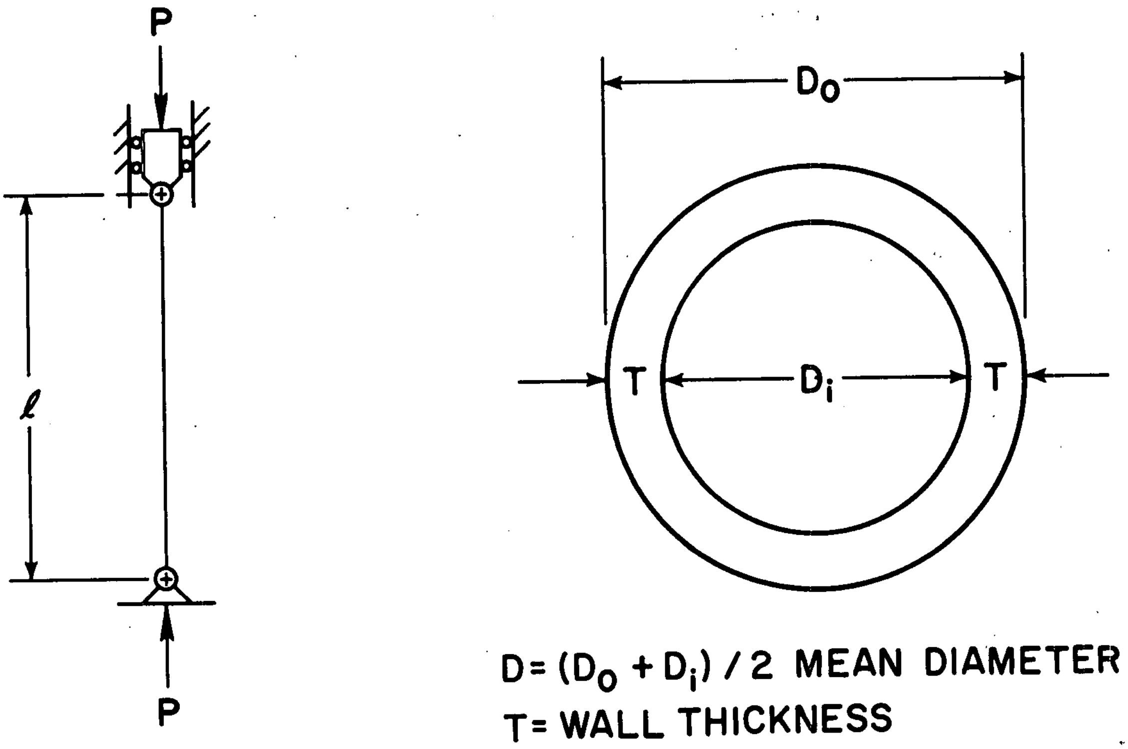


Fig. 1.1 Simple Tubular Column

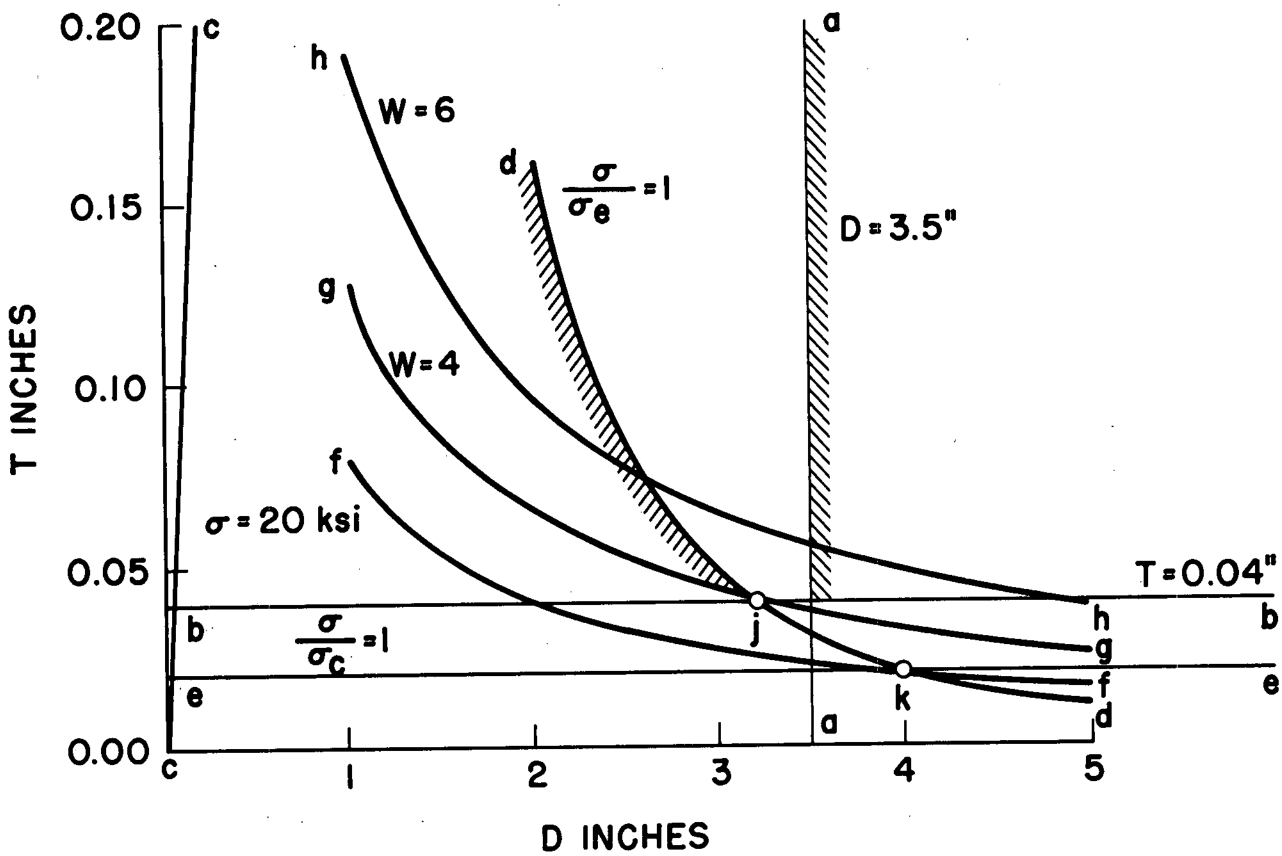


Fig. 1.2 Design Space – Simple Column

seek out the minimum weight design at point j. This design happens to lie at the vertex formed by the Euler buckling stress constraint and the lower limit on the tube wall thickness. It should be noted that if the requirements $D \leq 3.5$ in and $T \geq 0.04$ in were changed to say $D \leq 5$ in and $T \geq 0.015$ in, then the minimum weight design would lie at point k ($D = 4$ in, $T = 0.02$ in, $W = 2.52$ lb). In this case, the optimum design happens to lie at the vertex formed by the Euler buckling stress constraint and the local buckling stress constraint. The fact that both buckling stress limits are equal to 20000 lb/in^2 is fortuitous.

Fig.1.2 is a two-dimensional illustration of what is known as a design space representation; in a design problem involving N design variables such a space has N dimensions. The region corresponding to designs which satisfy all the constraints is known as the feasible region and the surface bounding it is referred to as the constraint surface; for the two-dimensional example shown in Fig.1.2 this surface degenerates into a collection of lines. In a two-dimensional space a vertex is formed by the intersection of two lines while in an N-dimensional space a vertex represents the intersection of N surfaces.

1.2.2 Two Bar Truss*

In the foregoing example, it was seen that various combinations of constraints could be critical at the optimum design depending upon the limitations specified. However, the optimum designs at point j in Fig.1.2 (for the case when $D \leq 3.5$ in and $T \geq 0.04$ in) and at point k (for the case when $D \leq 5$ in and $T \geq 0.015$ in) are both vertices. The second simple example illustrates that an optimum design need not necessarily lie at a vertex point in the design space.

Consider a symmetric two member truss (see Fig.1.3) subject to a load $2P = 66000$ lb. Let the two identical members have uniform annular cross section with a preassigned wall thickness $T = 0.1$ in. The horizontal distance between the support points is $2B = 60$ in and the pertinent material properties are given as follows; modulus of elasticity $E = 30 \times 10^6$ lb/in^2, density $\rho = 0.3$ lb/in^3, and yield stress $\sigma_y = 60000$ lb/in^2. The problem is to find the mean tube diameter D, the height H of the truss and the minimum weight W such that the compressive stress in the members is equal to or less than the Euler buckling stress σ_e and the yield stress σ_y. In this example, the wall thickness T, the support spacing B, and the structural material have been preassigned and only the mean diameter D of the tubes and the height of the truss H are variables to be determined. It should be noted that only one load condition is considered. The problem takes the following algebraic form:

$$\text{Minimize } W = 2\rho\,\pi\,DT\,(B^2 + H^2)^{\frac{1}{2}} \tag{1-14}$$

subject to the inequality constraints:

(1) Euler buckling

$$\frac{P\,(B^2 + H^2)^{\frac{1}{2}}}{\pi\,TDH} - \frac{\pi^2 E\,(D^2 + T^2)}{8\,(B^2 + H^2)} \leq 0 \tag{1-15}$$

(2) Yield stress

$$\frac{P\,(B^2 + H^2)^{\frac{1}{2}}}{\pi\,TDH} - \sigma_y \leq 0 \quad . \tag{1-16}$$

Introduction of the given numerical values into Eq. (1-14) through (1-16) makes it possible to construct the design space representation of this example shown in Fig.1.4.

It is apparent from the design space depicted in Fig.1.4 that the minimum weight design satisfying the various stated limitations lies at point p ($D = 2.47$ in, $H = 30$ in, $W = 19.8$ lb). In this case, the optimum design does not lie at the vertex, rather it is seen that the only critical constraint at point p in Fig.1.4 is the yield stress limitation. It is interesting and important to note that if the yield stress limit is raised to $\sigma_y = 100\,000$ lb/in^2, and the rest of the problem statement remains unchanged, then the design space is modified to that shown in Fig.1.5. Examining the design space shown it is apparent that the minimum weight design lies at point p ($D = 1.87$ in, $H = 20.2$ in, $W = 12.8$ lb); in this instance the optimum design happens to lie at a vertex formed by the intersection of the Euler buckling and the yield stress constraints.

1.2.3 Relationship to Traditional Approaches

Early contributors to the literature of the least weight design of aircraft structures such as Farrar [1.2], Shanley [1.3] and Gerard [1.4] almost always formulated the structural optimization problem in terms of equations. That is to say, the solution of a given problem was sought by preselecting the set of critical constraints that were thought to characterize the optimum design. This approach yields least weight designs in certain useful classes of application where the required number of constraints

*This example is due to R. L. Fox, see [1.1].

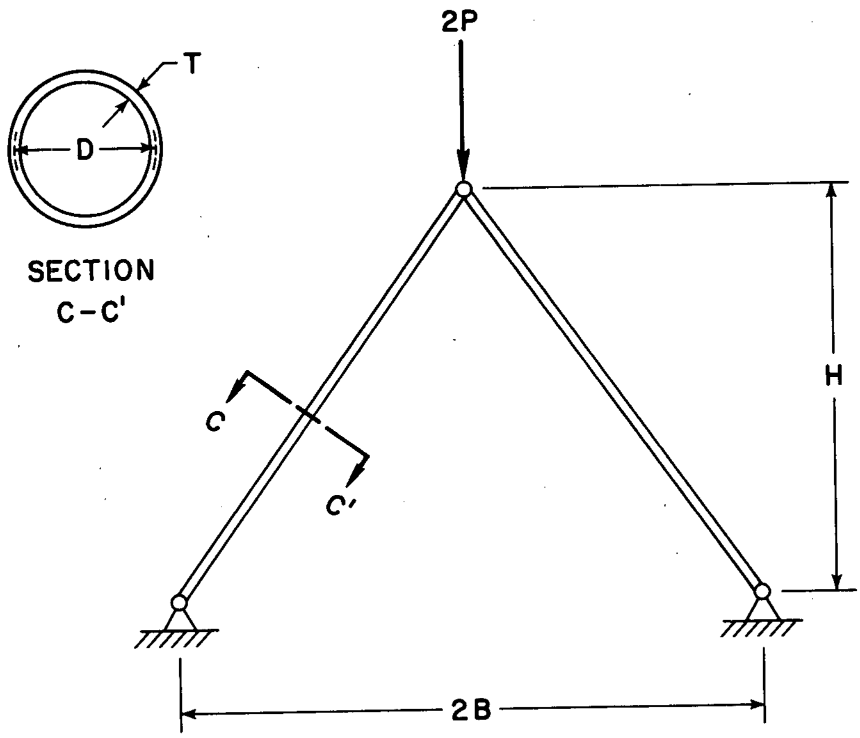


Fig.1.3 Two Bar Truss

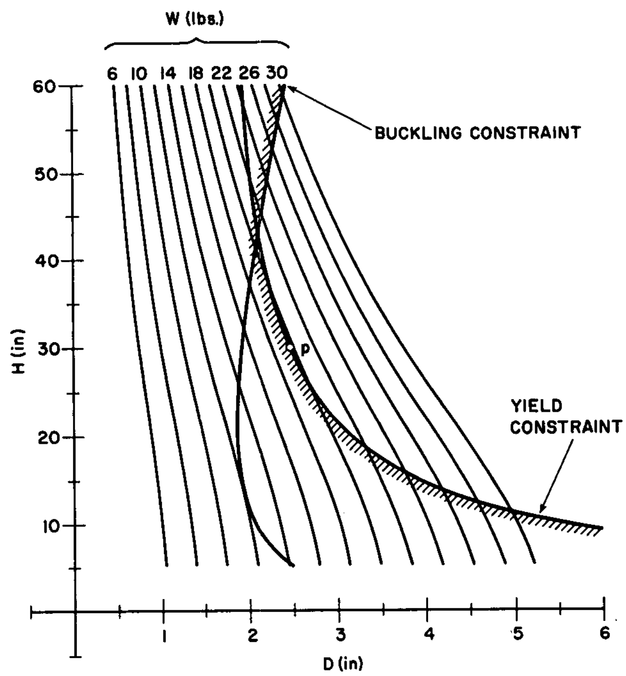


Fig.1.4 Design Space – Two Bar Truss (σ_y = 60 000 lb/in^2)

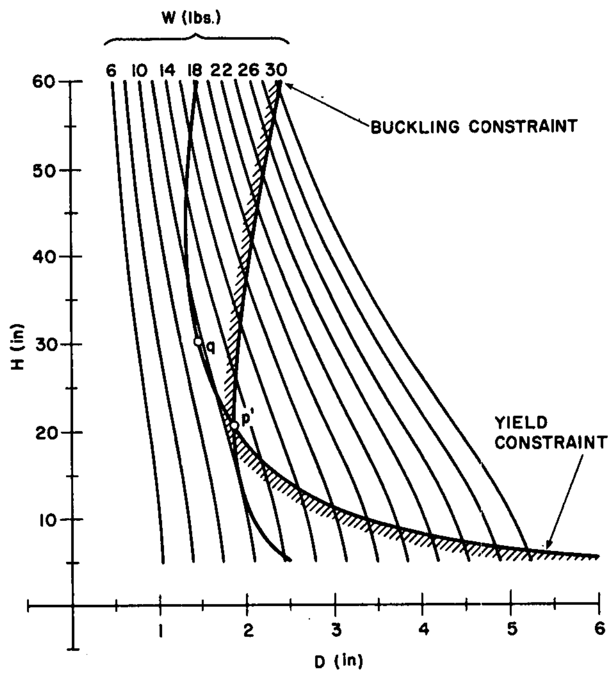


Fig.1.5 Design Space – Two Bar Truss (σ_y = 100 000 lb/in^2)

are critical and where the critical constraints are easily identified. A wrong choice of critical constraint can, however, lead to a wrong solution which may violate a constraint that was assumed not to be critical. For example, it might be assumed in the column application described in the preceding section that local buckling and overall buckling are both critical. Equating the failure stresses in these two modes, a design would be obtained in which D = 4 in and T = 0.02 in; the minimum gauge constraint that T = 0.04 in would thus be violated. This approach also falls down, of course, in applications where the optimum design involves less than the necessary number of critical constraints and consequently does not lie at a vertex of the constraint surface in design space.

A common variation of this traditional approach is the reduction of the objective function to a function of a single variable by preselecting an appropriate set of critical constraints. In the case of the two bar truss, it might be assumed for example that the yield stress constraint is critical at the optimum design, then it would follow from Eq. (1-16) that

$$D = \frac{P\,(B^2 + H^2)^{\frac{1}{2}}}{\pi\, TH\, \sigma_y} \quad . \tag{1-19}$$

Using Eq. (1-19) to eliminate D from the weight expression given by Eq. (1-14) would yield

$$W = \frac{2\rho\, P}{\sigma_y} \frac{(B^2 + H^2)}{H} \quad . \tag{1-20}$$

Setting the derivative of W with respect to H to zero gives

$$\frac{dW}{dH} = \frac{2\rho\, P}{\sigma_y} \left[- \frac{(B^2 + H^2)}{H^2} + 2 \right] = 0 \tag{1-21}$$

which would indicate that W is a minimum at H = B = 30 in. The corresponding values of D and W could then be computed from Eq. (1-19) and (1-20) respectively, and they would be found to be D = 2.47 in and W = 19.8 lb (see point p, Fig.1.4). Using this approach for the case where σ_y = 100 000 lb/in^2 would lead to the design H = 30 in, D = 1.48 in, W = 11.9 lb that is clearly in violation of the Euler buckling constraint (see point q, Fig.1.5).

The feature to be emphasized here is that, in general, it cannot be anticipated how many or which constraints will be critical at the optimum design. Thus, the use of inequality constraints becomes essential to a proper treatment of the structural design optimization problem.

1.2.4 Terminology and General Problem Statement

The application of mathematical programming techniques to structural design problems will be facilitated by introducing the following terminology. An idealized structural system can be described by a finite set of quantities that specify the materials, the arrangement, and the dimensions of the structure. Preassigned parameters are those quantities defining a structural system that are fixed at the outset of the automated design procedure. They are not varied by the directed redesign algorithm. Design variables are those quantities defining a structural system that are varied by the automated design procedure. The term load condition refers to one of several distinct sets of mechanical and thermal loads that approximately represent the effect on the structure of the environment to which it is exposed. A failure mode is defined as any structural behaviour characteristic subject to limitation by the responsible engineer. A rather broad class of failure modes which includes limitations on stress, deflection, buckling, natural frequency, and other behavioral characteristics can be formulated using inequality constraints. An objective function is defined as a function of the design variables the value of which provides a basis for choice between alternative acceptable designs

A rather general and very significant class of structural design problems can be stated concisely as problems in mathematical programming using the foregoing terminology. Given the preassigned parameters and a set of distinct load conditions, find the vector of design variables $(\vec{D})$ such that the objective function $M(\vec{D})$ is minimized (or maximized) subject to a collection of inequality constraints on the design variables,

$$h_j(\vec{D}) \leqslant 0 \quad ; \quad j = 1,2,\ldots J$$

where the functions $h_j(\vec{D})$ are such that

(1) unsatisfactory behaviour with respect to each failure mode under each load condition is precluded and

(2) the design variables are subject to further restrictions based upon considerations such as fabrication limitations, geometric realizability, and analysis validity.

Example

The usefulness of the terminology and the general problem statement is illustrated by discussing a simple example problem that is indeterminate and involves two distinct load conditions. Consider the three bar symmetric planar truss shown in Fig.1.6*. The configuration and the truss material are assumed to be fixed, i.e. the preassigned parameters are $N = 10$ in, $\beta_1 = 135^\circ$, $\beta_2 = 90^\circ$, $\beta_3 = 45^\circ$, $\rho = 0.1$ lb/in^3, and $E = 10 \times 10^6$ lb/in^2. Since the truss is to be symmetric, it is required that $A_1 = A_3$ and, therefore, the two independent design variables are A_1 and A_2. There are two distinct load conditions, the first specified by $P_1 = 20000$ lb acting at an angle of 45° to the X axis and the second specified by $P_2 = 20000$ lb acting at an angle of 135° to the X axis. The failure modes to be guarded against are simple upper and lower limits on the stress in each member in each load condition. Also, since negative areas must obviously be excluded, the range of admissible values for the design variables A_1 and A_2 have lower limits, i.e. $A_1 \geqslant 0$ and $A_2 \geqslant 0$. Minimization of the total weight is the goal of the optimization and, therefore, the objective function can be expressed in terms of the design variables as follows:

$$W(\vec{D}) = \rho N [2\sqrt{2}A_1 + A_2] \tag{1-22}$$

where it is understood that a point in the design space A_1, A_2 is defined by the vector $\vec{D}$, that is

$$\vec{D}^T = \lfloor A_1, A_2 \rfloor \quad . \tag{1-23}$$

Let σ_{ij} refer to the stress in the ith member in the jth load condition. From symmetry, it is obvious that $\sigma_{11} = \sigma_{32}$, $\sigma_{21} = \sigma_{22}$ and $\sigma_{31} = \sigma_{12}$. Therefore, it is only necessary to consider σ_{11}, σ_{21} and σ_{31}. The tensile stress limits can be written in standard form as follows:

$$h_1(\vec{D}) = \sigma_{11} - 20000 \leqslant 0 \tag{1-24a}$$

$$h_2(\vec{D}) = \sigma_{21} - 20000 \leqslant 0 \tag{1-24b}$$

$$h_3(\vec{D}) = \sigma_{31} - 20000 \leqslant 0 \tag{1-24c}$$

where the maximum permissible tensile stress is 20000 lb/in^2. The compression stress limits are expressed in the form,

$$h_4(\vec{D}) = -\sigma_{11} - 15000 \leqslant 0 \tag{1-25a}$$

$$h_5(\vec{D}) = -\sigma_{21} - 15000 \leqslant 0 \tag{1-25b}$$

$$h_6(\vec{D}) = -\sigma_{31} - 15000 \leqslant 0 \tag{1-25c}$$

where the maximum permissible compressive stress is 15000 lb/in^2. The constraints precluding negative areas can be put in the standard form,

$$h_7(\vec{D}) = -A_1 \leqslant 0 \tag{1-26a}$$

$$h_8(\vec{D}) = -A_2 \leqslant 0 \quad . \tag{1-26b}$$

From elementary structural analysis the following expressions may be substituted in Eq. (1-24) and (1-25):

$$\sigma_{11} = 20000 \left(\frac{1}{A_1} - \frac{A_2}{2A_1 A_2 + \sqrt{2}A_1^2} \right) \tag{1-27a}$$

$$\sigma_{21} = \frac{20000\sqrt{2}A_1}{2A_1 A_2 + \sqrt{2}A_1^2} \tag{1-27b}$$

*This example was first presented in [1,5].

$$\sigma_{31} = -\frac{20000A_2}{2A_1 A_2 + \sqrt{2}A_1^2} . \tag{1-27c}$$

The significant portion of the design space for this example is shown in Fig.1.7. The constraints separating the region of acceptable designs from the unacceptable domain are $h_1(\vec{D}) \leq 0$ (the tension stress limit in member 1 under load condition 1) and $h_6(\vec{D}) \leq 0$ (the compressive stress limit in member 3 under load condition 1). Note that the constraint $h_2(\vec{D}) \leq 0$ (the tension stress limit in member 2 under load condition 1) is always satisfied for designs in the positive quadrant $[h_7(\vec{D}) \leq 0,\ h_8(\vec{D}) \leq 0]$ provided the tension stress limit in member 1 in load condition 1 is satisfied $[\text{i.e. } h_1(\vec{D}) \leq 0]$. Selected contours of constant weight $\left(\frac{W}{\rho N} = W \text{ since } \rho N = 0.1 \times 10 = 1\right)$ are also shown in Fig.1.7. Scanning this design space, it is apparent that the minimum weight design lies at point 1 [i.e. $A_1 = A_3 = 0.788 \text{ in}^2$, $A_2 = 0.41 \text{ in}^2$ and $W = 2.64$ lb]. It should be noted that this optimum design does not lie at a vertex and it represents an indeterminate structure in which member 2 is not fully stressed in either load condition. The design represented by point 2 [i.e. $A_1 = A_3 = 1.0$ and $W = 2.83$ lb] is not the minimum weight optimum design in this case even though it is (a) at a vertex, (b) determinate, and (c) fully stressed in the sense that each member is fully stressed in at least one load condition*. It may be observed that the design represented by point 3 in Fig.1.7 is (a) at a vertex, (b) indeterminate, and (c) not fully stressed. This example illustrates again that the intuitive substitution of what is thought to be an equivalent problem for an inequality constrained minimum weight design problem can lead to incorrect results.

1.2.5 Features of the Mathematical Programming Approach

The application of mathematical programming techniques to structural design problems may be viewed as a generalization of conventional methods for structural optimization based on the realization that inequality constraint concepts are, in general, essential to proper formulation of these problems. When the structural design optimization problem is viewed as a mathematical programming problem:

(a) it is possible to consider the design of a structural system rather than the design of individual elements; allowance can be made where appropriate for quantities such as the weight of structural connections using, perhaps, statistical information,

(b) the behavioral characteristics of the optimum design need not be presumed, rather they emerge as a consequence of the design procedure,

(c) a variety of failure modes in each of several load conditions may be guarded against,

(d) restrictions on the design variables arising from fabrication considerations and limitations of the analysis employed can be treated,

(e) a wide variety of restrictions on structural behavior including stress, displacement, buckling, dynamic and thermal response can be dealt with,

(f) the approach is not inherently linked to weight minimization; that is to say, objective functions other than structural weight may be readily employed.

While reviewing the potential of mathematical programming techniques in the structural design field, it is well to point out a fundamental property of these techniques which can sometimes be a cause of difficulty. In any optimization problem of the form illustrated in Fig.1.8a, standard mathematical programming methods will yield the optimum solution; such problems are referred to as convex problems. Many structural applications are, however, of a more general form as, for example, illustrated in Fig.1.8b where local optima exist as well as the global optimum which is sought. Now mathematical programming techniques look, in effect, for conditions which are satisfied by a local optimum, so the solution obtained is liable to depend on the initial design from which the search procedure is started. This difficulty can be alleviated by repeating computations from radically different starting points and comparing results until reasonable confidence is built up that the global optimum has been achieved. A single application remains a powerful tool, however, as a means of improving a design which is the best that can be achieved by traditional means; in many problems single applications of mathematical programming techniques have yielded significantly more efficient designs than can be achieved without their aid.

1.2.6 Relationship to Materials Selection

The formulation of the structural design problems as a mathematical programming problem is in principle general enough to embrace both the design of the structural configuration and the structural material. Most applications of mathematical programming techniques have assumed that the design variables are continuous variables. However, the materials selection problem is usually characterized by a discrete set of available materials from which a choice is to be made. Such discrete

*The assumption that a fully utilized design is equivalent to a minimum weight design is frequently but not always valid. This topic has been examined in some depth and the interested reader is referred to [1.6], [1.7] and [1.8].

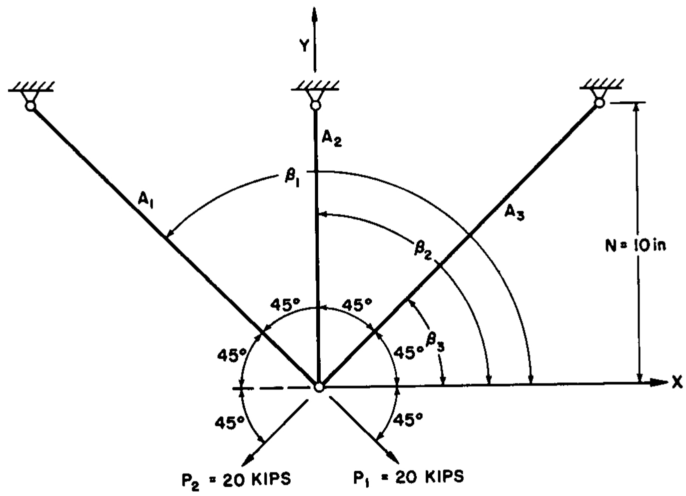


Fig.1.6 Symmetric Three Bar Truss

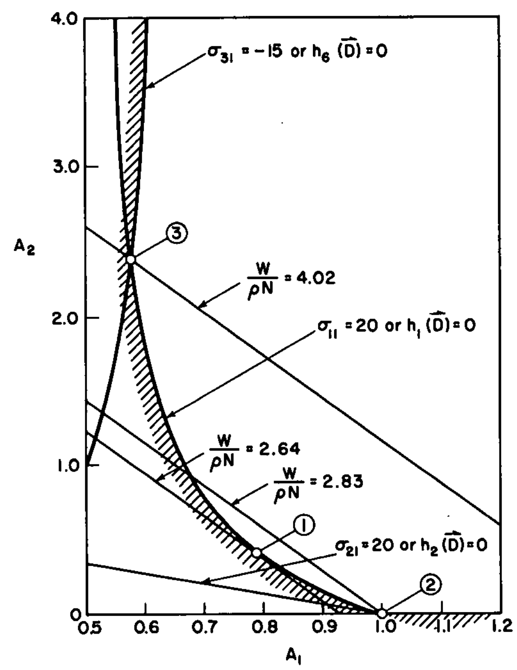


Fig.1.7 Design Space – Three Bar Truss

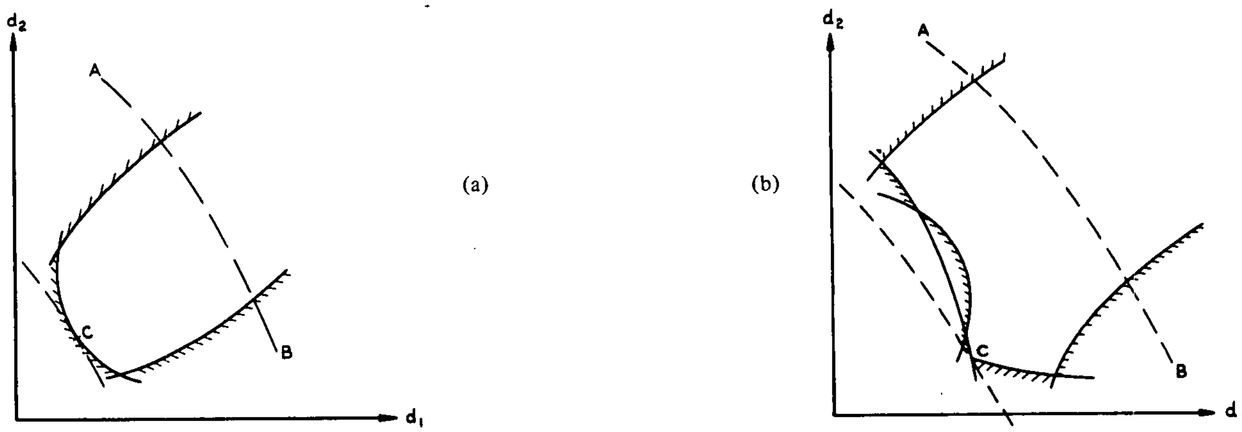


Fig.1.8 Convexity and Local Minima

variables might in theory be incorporated in the optimization process at the expense of a considerable increase in complexity and computational time, but when there are only two or three candidate materials and when the same material is to be used throughout the structure, it would probably be more efficient to perform the optimization on the basis of each material in turn and to compare the results at the end. Even if it is hypothesized that material variables can be treated as continuous, serious practical problems arise because most of the engineering material properties that are important in structural design depend upon experimental characterization. Furthermore, the dependence of engineering material properties, including cost, upon processing, fabrication and composition variables currently defies description.

The idea of applying mathematical programming techniques to simultaneous selection of a structural configuration and materials can be illustrated by the following simple but admittedly rather impractical example [1.9]. Consider the problem of designing the lightest weight three bar planar truss to transmit to a fixed support line represented by r-r in Fig.1.9, various concentrated loads (P_k) applied at point (s) and oriented at angles α_k to the X axis. Stress, displacement, and buckling failure modes are to be guarded against in each of several distinct loading conditions (mechanical and thermal). It is assumed that the pertinent engineering material properties may be expressed as continuous functions of the density. For this example the modulus of elasticity, thermal expansion coefficient and the yield stress of a representative class of structural alloys were plotted versus density and then curve-fitted, see [1.9] for details. The cross section of each truss member is assumed to be annular, with a preassigned mean-diameter to wall thickness ratio selected to preclude local buckling. The preassigned parameters for this example are: N, the normal distance from point s to the support line r-r in Fig.1.9, the mean diameter to thickness ratio $\left(\frac{D}{T}\right)$, for each thin walled tubular member, the modulus of elasticity as a function of density $E(\rho)$, the coefficient of thermal expansion as a function of density $\tilde{\alpha}(\rho)$, and the yield stress as a function of density $\sigma_y(\rho)$. The design variables are the density (ρ_p), the orientation angle (β_p) and the cross sectional area (A_p) for each of the members ($p = 1, 2, 3$).

Constraints are placed on the range of values that can be assumed by the various design variables as follows:-

$$0.05 \leqslant \rho_p \leqslant 0.32 \;;\; p = 1, 2, 3 \tag{1-28}$$

$$\beta_2 \leqslant \beta_1 < \pi \tag{1-29a}$$

$$\beta_3 \leqslant \beta_2 \leqslant \beta_1 \tag{1-29b}$$

$$0 < \beta_3 \leqslant \beta_2 \tag{1-29c}$$

and

$$0 < A_p \leqslant (A_p)_{max} \tag{1-30}$$

where the $(A_p)_{max}$ represent upper limits on the cross sectional areas. From an examination of Fig.1.9 it can be seen that the constraints stated in Eq. (1-29) serve to preclude the possibility of members of infinite length and they also order the position of the members. The load conditions are specified by giving the magnitude P_k and the orientation α_k of the mechanical load applied at joint s for each load condition k as well as the corresponding temperature changes ΔT_{pk}. Inequality constraints are easily generated to guard against unsatisfactory behavior with respect to the several failure modes. The stress in each member p in each load condition k is required to be equal to or less than the tensile yield stress and equal to or greater than the compressive yield stress or buckling stress whichever is critical (assuming tensile stress is positive and compressive stress is negative). The x and y displacement components of the point s are subject to upper and lower limits in each load condition. The structural weight which is seen to be the non-linear function of the nine design variables,

$$W(A_p, \beta_p, \rho_p) = \sum_{p=1}^{3} \rho_p \frac{N}{\sin \beta_p} A_p \tag{1-31}$$

is taken as the objective function.

The analysis used to predict the behavior of any particular trial design follows from a straight forward application of elementary structural mechanics. The directed redesign procedure used to obtain numerical results is described in [1.9]. Results for several numerical examples* are given there and it is shown that mathematical programming techniques can be used to carry out simultaneous selection of structural material and configuration within the context of this rather highly idealized example.

---

*Another interesting aspect of these results was that when displacement constraints governed the design, it was often found that many optimum designs all having the same minimum weight existed.

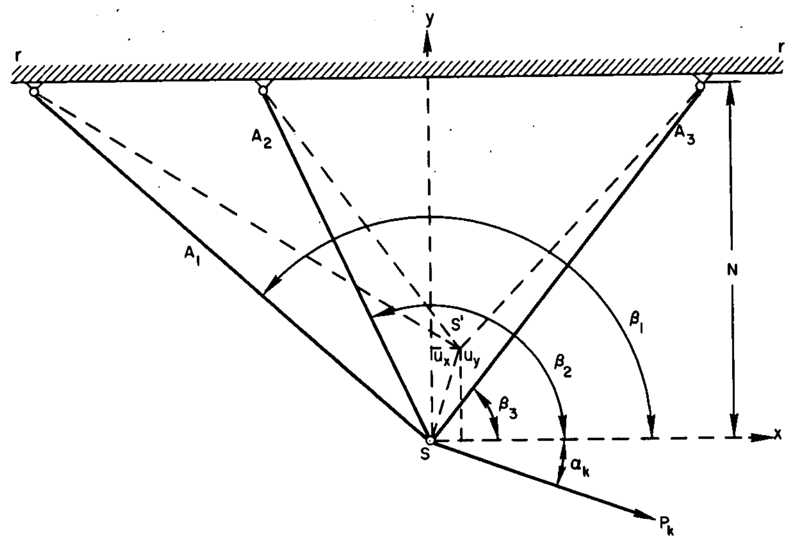


Fig.1.9 Three Bar Truss

While this example is admittedly impractical, the basic approach it illustrates may have long range potential.

The emergence of high performance composite materials has encouraged some further consideration of the idea of simultaneous design of structural configuration and structural material. For example, in fiber composites the volume fraction of fibers could be considered as a design variable. With carbon fibers the modulus of elasticity in the longitudinal direction may be treated in principle as a continuous design variable over a very wide range.

It should be noted that ply orientation angles are not viewed here as material design variables but rather they are thought of as laminate configuration design variables. In the area of ceramic materials and particulate composites, it is possible in principle to represent both the composition and the density of the material using continuous design variables. Here again difficulties are experienced due to the dependence of engineering material properties, including cost, on the material design variables which can in general only be obtained by an extensive experimental characterization program. Even if one imagines carrying out such a program for a sample set of material design variable values, there is no assurance that interpolation between such data points is valid.

For the foregoing reasons, the materials selection problem even for composite materials tends in practice to be discrete. While the simultaneous design of materials and structures remains a desirable long range goal, major advances are needed in the prediction of engineering material properties from material design variables to make this possible. For the present, the application of mathematical programming techniques in structural design can aid in the materials selection process by making it possible to compare optimum designs based upon alternative discrete materials. It may be noted that these existing methods can also be used to generate optimum designs based on hypothetical material properties that are judged to be realizable in the future. In this way, methods for seeking optimum structural designs for alternative hypothetical and existing materials can be used to help guide materials development effort into areas of high payoff.

List of References

Ref.

1.1 Fox, R. L., *An Introduction to Optimization Methods for Engineers*, to be published by Addison-Wesley, Reading, Massachusetts, 1970

1.2 Farrar, D. J., 'The Design of Compression Structures for Minimum Weight', Journal of the R.Ae.S., Vol.53, 1949, pp.1041-1052

1.3 Shanley, F. R., *Weight Strength Analysis of Aircraft Structures*, McGraw-Hill Book Co., Inc., New York, 1952

1.4 Gerard, G., *Minimum Weight Analysis of Compressive Structures*, New York University Press, New York, 1956

1.5 Schmit, L. A., 'Structural Design by Systematic Synthesis', Proc. of the Second National Conference on Electronic Computation, Structural Division, *ASCE*, Pittsburgh, Pa., September 1960, pp.105-132

1.6 Kicher, T. P., 'Optimum Design - Minimum Weight Versus Fully Stressed', J. of the Structural Division, *ASCE*, Vol.92, No.ST6, December 1966, pp.265-279

1.7 Razani, Reza, 'The Behavior of Fully-Stressed Design of Structures and its Relationship to Minimum Weight Design,' *AIAA Journal*, Vol.3, No.12, December 1965, pp.2262-2268

1.8 Dayaratnam, P. and Patnaik, S., 'Feasibility of Full Stress Design', *AIAA Journal*, Vol.7, No.4, 1969, pp.773-774

1.9 Schmit, L. A. and Mallett, R. A., 'Structural Synthesis and the Design Parameter Hierarchy', J. of the Structural Division, *ASCE*, Vol.89, No.ST4, August 1963, pp.269-299

Chapter 2

A BASIS FOR ASSESSING THE STATE-OF-THE-ART

by

L. A. Schmit

2.1 Introduction

The growing awareness that a significant class of structural design problems may be attacked by combining computer oriented structural analysis with mathematical programming methods to generate automated directed redesign procedures has led to a diverse and increasing body of knowledge. In order to provide a basis for reviewing some of the recent literature and to help achieve an organized and coordinated overview of the subject, the following philosophical framework is set forth. Most of the successful structural design applications of mathematical programming techniques deal with the problem in design variable space. The method of attack corresponds to that illustrated by the various simple examples of Chapter 1 (see Sections 1.2.1, 1.2.2, 1.2.4). This general class of problems can be concisely stated as follows:

$$\text{Find } \vec{D} \text{ such that } h_j(\vec{D}) \leq 0 \;;\quad j = 1,2,\ldots J \tag{2-1a}$$

$$\text{and } \quad M(\vec{D}) \rightarrow \text{Min} \;. \tag{2-1b}$$

The vector of N design variables $\vec{D}$ locates a point in an N-dimensional space, while the J inequality constraints (2-1a) must be satisfied for a design to be acceptable; $M(\vec{D})$ is the objective function. Structural design problems of this form are finite in the sense that the design vector $\vec{D}$ contains a finite number of components. It is assumed that the assignment of numerical values to these components specifies a unique structure. It should be noted in passing that many problems in structural analysis and integrated analysis-design may be viewed as mathematical programming problems. Some of these will be discussed briefly in Section 2.7. However, structural design applications having the form of Eq. (2-1) are of primary interest in this volume.

2.2 Finite or Analytic

It will be useful to distinguish finite optimization problems, to which mathematical programming techniques may be applied directly, from analytic optimization problems in which the goal is to find the form of one or more functions. In the case of analytic optimization problems, the structural design is represented by one or more unknown functions and the form of these functions is sought such that the objective functional is minimized subject to various equality and inequality constraints. Analytical solutions of structural design optimization problems* when they can be found, provide valuable insight and benchmark solutions against which finite solutions can be evaluated. However, it is likely that the design optimization of practical structures exhibiting realistic complexity will continue to be accomplished mainly by the use of finite formulations. This viewpoint is supported by the well established widespread use of finite formulations in structural analysis. It should be noted that essentially this same opinion was expressed by Sheu and Prager in the concluding remarks section of their recent literature review [2.3].

2.3 Design Philosophy

Characterization of a structural design philosophy involves many considerations. Three of the more important bases for characterization are:

(a) classification of the design philosophy as deterministic or probability based,

(b) identification of the kinds of failure modes to be guarded against,

(c) classification with respect to consideration of service load conditions and/or overload conditions.

Structural systems are usually subjected to environments that are complex and continuously changing with time. In design practice, the environment is usually replaced by a multiplicity of distinct loading conditions and this idealization is a critical step requiring professional judgement and experience. Both deterministic and probability based design philosophies are possible within the idealized context in which a discrete set of load conditions is presumed to replace the actual environment. If any of the quantities involved in a structural design problem are treated as random variables, the formulation will be classified as probability based (PB). On the other hand, if all of the quantities involved in a structural design problem are treated as deterministic (DET), then the formulation will be so classified. Although the elastic deterministic design philosophy is still common practice today, it can be argued that in view of uncertainties with respect to load levels and strengths, it would be more rational to treat these quantities (and others) as random variables, see for example [2.4], [2.5] and [2.6]. Recent developments in the area of probability based structural design optimization are discussed in Chapter 10.

*For some recent examples, see [2.1] and [2.2].

There are various ways of seeking to assure that a structural system will perform its specified functional purposes. These consist of striving to avoid the occurrence of various kinds of failure modes. What constitutes failure must be carefully defined and this can be expected to vary from one design task to another. Furthermore, the kinds of failure modes to be guarded against under service load conditions will usually differ markedly from those considered under overload conditions. Service load conditions will be defined as design load conditions representative of normal use. Overload conditions will be defined as load conditions representative of certain anticipated extraordinary or emergency situations. It is useful to distinguish overload conditions that stem from scaling up a service load condition (by multiplying by a 'safety factor') from overload conditions, such as earthquake and nuclear weapons effects, that do not correspond to any service load condition. In aircraft structural engineering practice static service load conditions are generally called 'limit load' conditions and overload conditions are generally called 'ultimate load' conditions. In civil engineering, however, overload conditions obtained by scaling up service load conditions are often called limit or ultimate load conditions. Adequate performance of a structural system may be sought by trying to avoid failure modes such as initial yielding, excessive deflection, and local instability under service load conditions and/or by striving to prevent failure modes such as rupture, collapse, and general instability under overload conditions.

One well known approach is to design the structure so that initial yielding under service load conditions is avoided. For example in civil engineering practice the elastic deterministic design philosophy consists of applying a 'safety factor' to the material yield stress in order to establish allowable working stresses and then designing the structure deterministically so that these allowable working stresses are not exceeded under service load conditions. The objective of this approach is to make it highly unlikely that the yield stress will ever be exceeded under service load conditions. In an elastic probability based design philosophy the structure is designed so that the probability of exceeding the yield stress under service load conditions is less than a specified minimum. In other words, failure is assumed to have taken place if the yield stress in any member is exceeded in any service load condition, and the structure is designed to insure that the probability of failure is less than a specified minimum.

A second well known approach, which may be used as an alternative, or in addition to the foregoing, is to design so as to prevent collapse under service load conditions. For example, in civil engineering practice deterministic limit design philosophy consists of applying a 'safety factor' to the service load conditions in order to establish the overload conditions and then designing the structure deterministically so as to preclude plastic collapse under the overload conditions. The objective of this approach is to make it highly unlikely that plastic collapse will occur under service load conditions. In a probability based limit design philosophy the structure is designed so that the probability of plastic collapse is less than a specified minimum when the structure is subject to a set of service load conditions. It should be noted that from a probability based viewpoint, design against plastic collapse and design against initial yield are both service load oriented design philosophies. Deterministic design to preclude plastic collapse under overload conditions scaled up from service load conditions, may be viewed as an artificial device for trying to keep the probability of plastic collapse under service load conditions small.

Another approach to seeking assurance that a structural design will perform its specified functional purposes is to design the structure to avoid permanent damage under service load conditions and catastrophic failure under overload conditions. For example, aircraft structural engineering practice often consists of designing the structure deterministically so as to preclude damage under static service load conditions (limit loads) as well as prevent catastrophic failure under overload conditions (ultimate loads). The corresponding probability based design philosophy would seek to limit the probability of permanent damage under service load conditions as well as the probability of catastrophic failure under overload conditions. Within the spirit of this design philosophy it would also be appropriate to strictly limit the probability of catastrophic failure under service load conditions.

In examining a particular application of mathematical programming to structural design, it will be useful to

(a) classify the design philosophy as deterministic or probability based,

(b) identify the kinds of failure modes considered,

(c) know if service load conditions and/or overload conditions are considered.

2.4 Kinds of Design Variables

The design variables used to describe structural systems can be categorized from a mathematical and physical viewpoint. From a mathematical point of view, it is important to distinguish between continuous and discrete design variables. In practical design problems, many of the design variables are strictly speaking, discrete. For example, sheet thicknesses may only be selected from commercially available gauges. However, if a large number of discrete values exists uniformly distributed over a limited interval, use of a continuous variable representation is often satisfactory, followed by selection of the nearest available discrete value. When a strictly discrete design variable is handled in this way, it will be categorized as pseudo-discrete. While a significant class of structural synthesis problems can be adequately formulated using continuous or pseudo-discrete design variables, it should be recognized that situations arise where it will be essential to employ discrete or integer variables. Integer variables can play an important role in describing a structural system. The number of major rings in a stiffened cylindrical shell, the number of plies in a laminated ply construction, the number of flange splices in a continuous welded girder are all examples of important integer variables. Problems involving integer variables are often further complicated by the fact that the number of continuous, pseudo-discrete, or discrete design variables describing the structure often depends upon the value of the integer variable(s). Declaration of the existence (1) or absence (0) of a structural element, such as a truss member joining two nodes, may be thought of as an important special case of an integer variable limited to the values 0 and 1. Sved and Ginos [2.7] have pointed out that optimization problems with inequality constraints can have singular global minima that cannot be reached from an arbitrary point through the continuous set of variables involved. They illustrate this point with a three bar truss example. This suggests that it may be necessary to represent some design variables D_i as the product of a 0-1 integer δ_i and a scalar $\tilde{D}_i$ [i.e. $D_i = \delta_i \tilde{D}_i$].

From a physical point of view, it may be helpful to consider that a design variable hierarchy exists which facilitates classification of the various quantities describing structural systems. Thus section properties or cross sectional dimensions of structural elements are said to describe the sizing or proportioning of a structure. The coordinates locating joints in trusses and frames may be viewed as configuration or geometric layout variables. In fiber composite materials ply orientation angles can be thought of as configuration variables while the number of plies at each angle is an integer sizing variable. Furthermore, in such materials the fiber volume fraction and the longitudinal modulus of elasticity of the fibers may in some cases, be viewed as a class of quantities that can be called material design variables. Another level in the hierarchy is represented by the possibility of using integer variables in a connectivity matrix to describe whether or not a member exists (1) or is absent (0). Design variables of this type will be referred to as topological variables.

2.5 Objective Function

When considering the application of mathematical programming to the structural design problem, it is necessary that a basis for choice between alternate acceptable design be selected. The nature of the structural design problem is such that there will usually be many designs that perform the specified functional purposes adequately provided that limitations on weight and/or cost are ignored. The objective of structural design optimization is frequently taken to be weight minimization. It can often be argued that weight minimization tends toward an economical structure since cost is intimately related to the amount of material required. Perhaps another reason that weight is so often used as the objective function in this field is because it is readily quantifiable. This is soon appreciated when one attempts to gather information for constructing a cost function including in addition to material cost, fabrication costs, tooling costs, etc. Indeed the cost of initially designing and constructing a structure is only a part of the overall cost picture which would usually include factors such as operating and/or maintenance costs, repair costs, insurance costs, etc. Ignoring the difficulties of quantification, an approach that appears rational would be to seek a structure of minimum total cost subject to constraints that limit the probability of failure during a specified lifetime. It is even possible to imagine carrying this thought one step further to minimization of total cost, including failure costs which depend upon the probabilities of failure. Contributions to the total cost, charged against failure, could be given by the damage cost associated with a particular failure multiplied by its probability of occurrence. It is, however, recognized that answering the moral question of what constitutes an appropriate failure damage cost is likely to be as difficult as selecting an acceptable probability of failure.

The selection of an objective function that is quantifiable and which effectively relates a structural system (or subsystem) to the larger system of which it is a part calls for mature professional judgement, experience, and deep insight. One guide to selecting an objective function may be stated as follows: the design should be optimized with respect to the 'most important' design property that can be 'meaningfully quantified' and that is not constrained in advance. In this connection it may be noted that if weight or cost are severely constrained in addition to the structural behavior, the set of acceptable designs may be extremely small or even null.

It addition to being readily quantifiable, weight is often the most important design property in aerospace applications as well as in other vehicle systems, including ships, trains and trucks. Structural weight saved can be converted directly into increased payload or indirectly into increased range, etc. The demand for high performance aerospace structures has provided a major impetus to the development of tools for minimum weight design. It must, however, be emphasized that the application of mathematical programming to the structural design problem is not inherently committed to the exclusive use of weight as the objective function.

2.6 Formulations and Algorithmic Tools

Once a structural design problem has been formulated and cast in the form of a mathematical programming problem, selection of a solution procedure remains. The basic non-linear programming problem of Eq. (2-1) may be attacked directly employing various feasible direction methods (see Chapter 7) or the problem may be transformed into an alternative form such as a sequence of linear programs (see Chapter 5) or a sequence of unconstrained minimizations (see Chapter 6). It should be noted that the classical formulation of the inequality constrained minimization problem, using Lagrange multipliers and slack variables, may be viewed as a way of transforming the basic problem, Eq. (2-1), into a set of non-linear simultaneous equations. Replacement of the basic problem statement with an equivalent substitute problem is a formulative device leading to an alternative casting of the basic problem. This step precedes the selection of an algorithm for obtaining numerical results. It is useful to distinguish between various alternative formulations because for each casting, a different collection of algorithmic tools may be drawn upon. The relationship between the four alternate formulations and the corresponding collection of algorithmic tools is summarized as follows:

Formulation	Algorithmic Tools	Relevant portion of this Volume
Sequence of Linear Programs SLP	Simplex and other LP Algorithms	Section 2.6.1 and Chapter 5
Sequence of Unconstrained Minimizations Techniques SUMT	Unconstrained Minimization Algorithms	Section 2.6.2 and Chapter 6
Basic Non-linear Programming Approach NLP	Feasible Direction Methods	Section 2.6.3 and Chapter 7
Classical Formulation with Slack Variables and Lagrange Multipliers	Methods for Solving Non-linear Simultaneous Equation	Section 2.6.4

Note that a guide to portions of this volume dealing with each formulation and the corresponding algorithms is given in the foregoing outline.

2.6.1 Sequence of Linear Programs (SLP) Formulation

Transformation into a sequence of linear programming problems can be accomplished by replacing the functions $h_j(\vec{D})$ and $M(\vec{D})$ (see Eq. (2-1)) by linear approximations obtained from Taylor series expansions about a point $\vec{D}_p$. Let $\vec{D}_o$ denote the initial trial design, then the sequence $\vec{D}_p$, $p = 1, 2, 3, \ldots$ represents successive solutions of the following linear programming problem:

$$\text{Find } \vec{D} \text{ such that } \tilde{h}_j^{(p)}(\vec{D}) \leqslant 0 \;; \quad j = 1,2,\ldots \tag{2-2}$$

$$\text{and} \qquad \tilde{M}^{(p)}(\vec{D}) \to \text{Min} \tag{2-3}$$

$$\text{where} \qquad \tilde{h}_j^{(p)}(\vec{D}) = h_j(\vec{D}_p) + \nabla h_j(\vec{D}_p) \cdot [\vec{D} - \vec{D}_p] \tag{2-4}$$

$$\tilde{M}^{(p)} = M(\vec{D}_p) + \nabla M(\vec{D}_p) \cdot [\vec{D} - \vec{D}_p] \quad \text{for } p = 0,1,2,\ldots \tag{2-5}$$

This alternative formulation, as a sequence of linear programs, makes it possible to bring existing linear programming algorithms to bear on the basic non-linear programming problem. The basic ideas involved in this approach are illustrated graphically in Fig.2.1 which depicts a sequence of three linear programs for the two member truss problem previously discussed (see Figs.1.3 and 1.5, also Section 1.2.2); it will be seen that additional constraints known as move limits have been introduced to prevent undesirably large changes in the variables in a given linearized problem.

In this example (Fig.2.1), the actual solution lies at a vertex point in design space. If the solution of the original problem does not lie at a vertex, additional constraints have to be introduced to achieve convergence. It is best in problems that are not known to be convex to use move limits for this purpose. This and other techniques for achieving convergence have been studied by Reinschmidt, Cornell and Brotchie [2.8] and by Moses [2.9] and the subject is discussed more fully in Chapter 5 which is devoted to the sequence of linear programs formulation.

2.6.2 Sequence of Unconstrained Minimizations Techniques (SUMT)

There are several alternative castings of the basic problem (see Eq. (2-1)) that can be classified as penalty function formulations. Penalty function methods transform the basic problem into alternative formulations such that numerical solutions are sought by solving a sequence of unconstrained minimization problems. For example, the Fiacco-McCormick formulation [2.10], [2.11], [2.12] can be stated as follows:

Given an initial value of the scalar $r_p = r_1$ and an initial value of $\vec{D} = \vec{D}_o$ such that

$$h_j(\vec{D}_o) < 0 \quad , \quad j = 1,2,\ldots J$$

generate a sequence of vectors $\vec{D}_p$, $p = 1,2,\ldots$

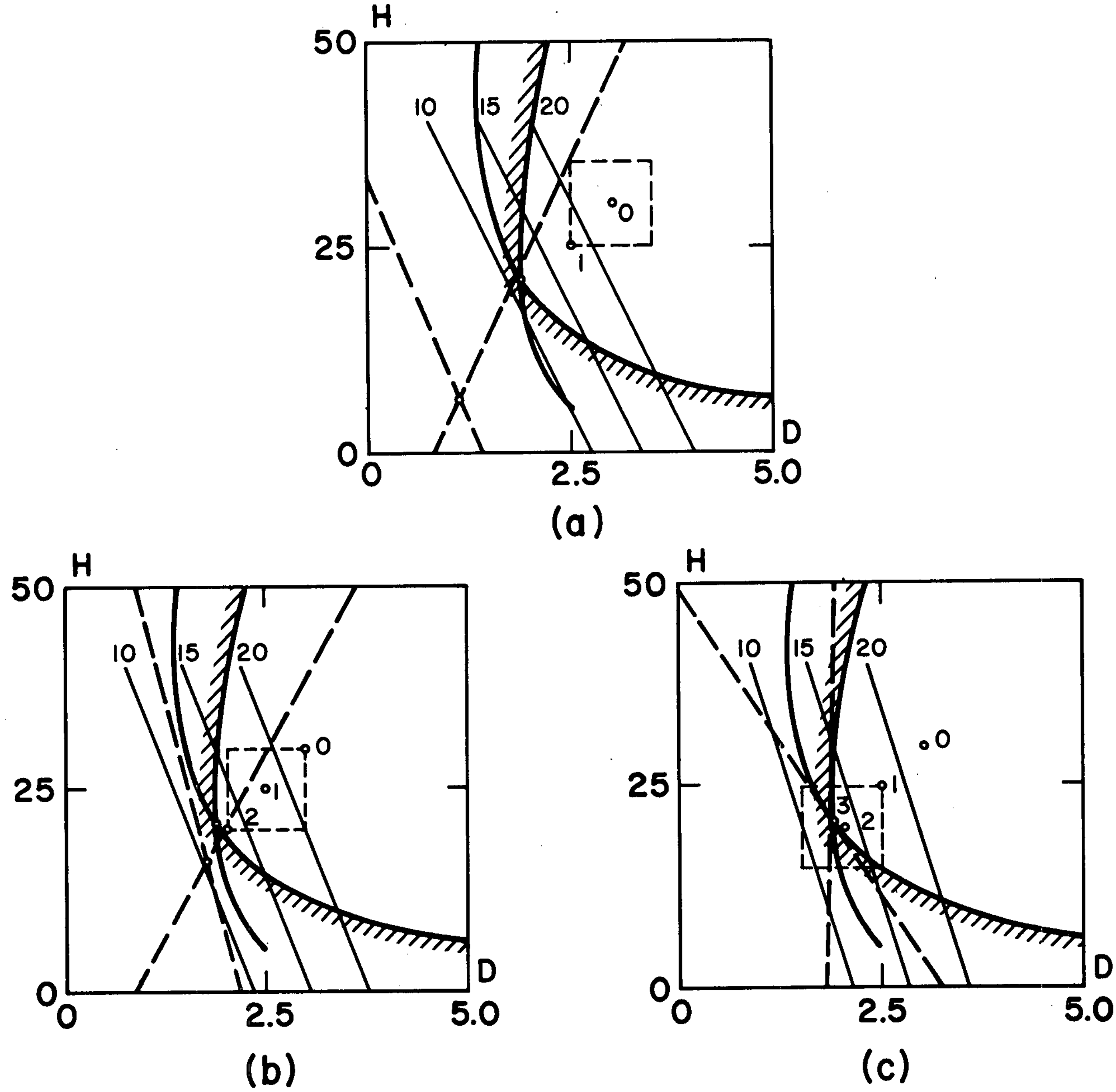


Fig.2.1 Sequence of LP's for Two Bar Truss

such that $$\phi(\vec{D},r_p) \rightarrow \text{Min} \tag{2-6}$$

and $$h_j(\vec{D}) \leq 0 \; ; \quad j = 1,2,\ldots J \tag{2-7}$$

where

$$\phi(\vec{D},r_p) = M(\vec{D}) - r_p \sum_{j=1}^{J} \frac{1}{h_j(\vec{D})} \tag{2-8}$$

and $$r_{p+1} < r_p \quad . \tag{2-9}$$

The first term on the right hand side of Eq. (2-8) is the objective function and the second term is a constraint repulsion function that serves to keep $\vec{D}$ inside the acceptable region defined by the J inequality constraints. For large values of r_p, the penalty function

$$- r_p \sum_{j=1}^{J} \frac{1}{h_j(\vec{D})}$$

interferes with the true minimum of $M(\vec{D})$. However, small values of r_p lead to functions $\phi(\vec{D},r_p)$ that are difficult to minimize, so a sequence of tractable unconstrained minimization problems is generated by reducing r_p gradually. Considerable insight into the nature of the Fiacco-McCormick penalty function formulation can be gleaned from the sequence of three contour plots shown in Fig.2.2 which are based on the two bar truss problem previously discussed (see Eq. (1-14) through (1-16) and Chapter 6). The unconstrained minimum of $\phi(\vec{D},r_1)$ shown in Fig.2.2a is at point 1. The second unconstrained minimization stage [i.e. find $\vec{D}$ such that $\phi(\vec{D}, r_2) \rightarrow$ Min] terminates at point 2 in Fig.2.2b, and the third stage terminates at point 3 in Fig.2.2c. Note that as r decreases $[r_1 > r_2 > r_3]$ the function becomes more eccentric. It is seen that the method generates a sequence of designs that approach the constraints gradually. The solution of the initial unconstrained minimization problem begins from a given starting design D_o which satisfies the inequality constraints, (2-7). Each subsequent stage can use the solution of the previous stage as a starting point. However, it is possible, in many applications, to accelerate the overall procedure by employing extrapolation techniques to determine starting points for subsequent unconstrained minimization cycles (after two or more minimization stages have been completed). Starting points obtained by extrapolation must be checked to be sure that they satisfy the constraints, (2-7), because at each stage, it is necessary to start the unconstrained minimization of $\phi(\vec{D},r_p)$ from an acceptable design point.

Since each of the designs generated by the foregoing penalty function approach lies inside the acceptable region of the design space, the method is classified as an <u>interior penalty function</u> formulation. This constraint repulsion feature has important engineering implications. The method tends to generate a sequence of designs which decrease the value of the objective function such that none of the designs in the sequence is critical with respect to the set of inequality constraints, (2-7). Qualitatively speaking, it can be said that the method tends to 'funnel' the sequence of trial designs down the middle of the acceptable region. This characteristic makes it possible to consider the use of approximate analysis methods during major portions of the optimization procedure, see [2.13] and [2.14]. Marcal and Gellatly [2.15] have suggested that this type of formulation can be extended to embrace discrete variables.

As suggested by Zoutendijk [2.16], this formulation can also be extended to deal with parametric inequality constraints of the form

$$h_j(z,\vec{D}) \leq 0 \; ; \quad z_1 \leq z \leq z_2 \quad ; \quad j = 1,2,\ldots J \tag{2-10}$$

by redefining the function $\phi(\vec{D},r_p)$ in Eq. (2-8) as follows:

$$\phi(\vec{D},r_p) = M(\vec{D}) - r_p \sum_{j=1}^{J} \left[\frac{1}{(z_2 - z_1)} \int_{z_1}^{z_2} \frac{dz}{h_j(z,\vec{D})} \right] . \tag{2-11}$$

The effect of this extension is to introduce into the penalty function the influence of each inequality constraint over the entire specified range of values for the parameter z, rather than just the influence of each constraint at the z value for which it is most critical. When using the integral penalty function formulation (Eq. (2-11)) care must be exercised to ensure that the parametric inequality constraints represented by Eq. (2-10) are not violated at any value of z in the range between z_1 and z_2 during any stage of the solution process. This approach can be further extended

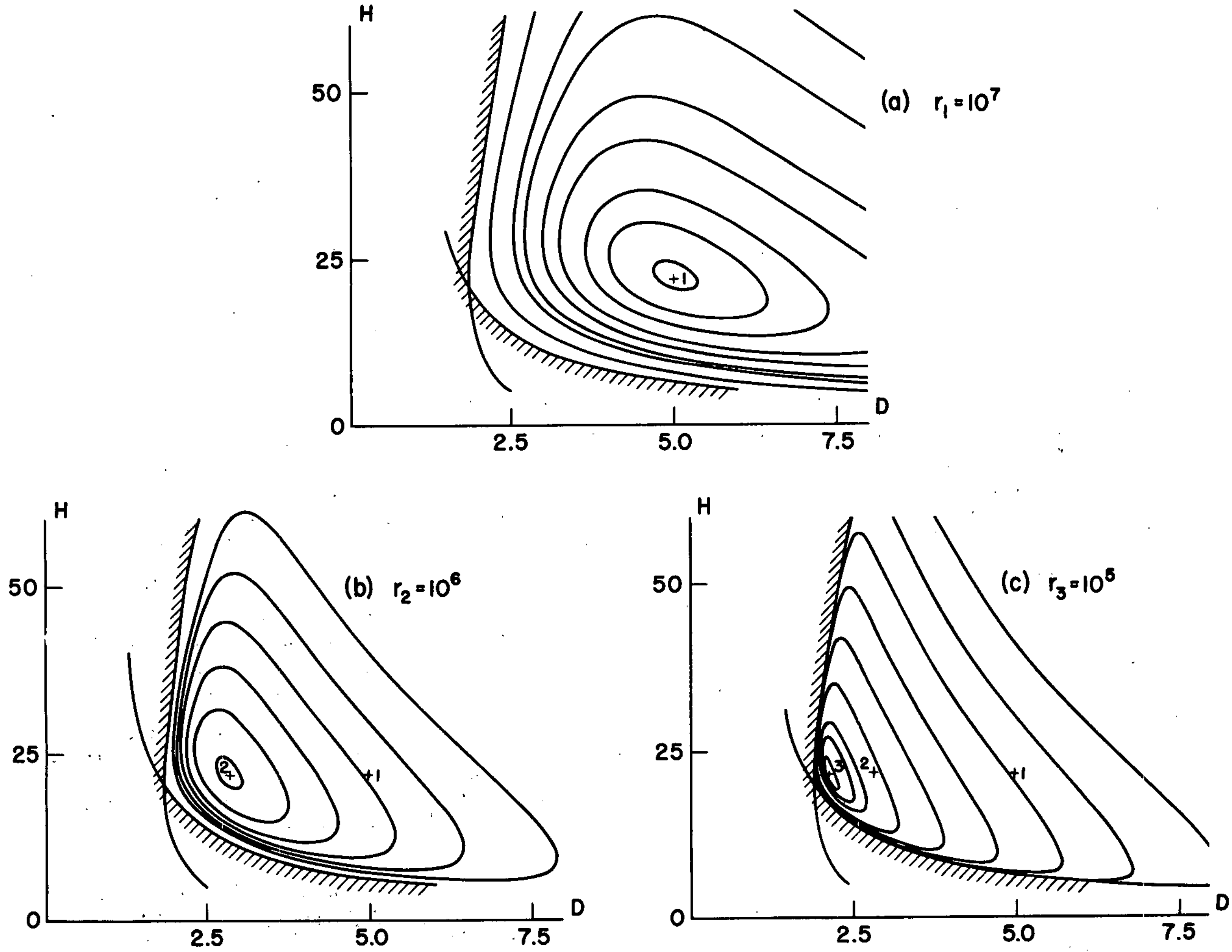


Fig.2.2 SUMT for Two Bar Truss

to deal with inequality constraints that are dependent on several parameters such as time and/or spatial parameters, see for example [2.17]. It should also be noted that envelope type loadings (see Fig.2.3) or moving load situations (see Fig.2.4) can be dealt with using integral penalty functions to introduce the influence of parametric inequality constraints.

As a second example of a penalty function formulation, consider the following transformation of the basic problem:

Given a small initial value of $r_p = r_1$ and an initial value of $\vec{D} = \vec{D}_o$ generate a sequence of vectors $\vec{D}_p$, $p = 1,2,\ldots$ such that

$$\phi(\vec{D},r_p) \rightarrow \text{Min} \tag{2-12}$$

where

$$\phi(\vec{D},r_p) = M(\vec{D}) + r_p \sum_{j=1}^{J} \langle h_j(\vec{D})\rangle^2 \tag{2-13}$$

$$\langle h_j(\vec{D})\rangle = \begin{cases} h_j(\vec{D}) & ; \quad h_j(\vec{D}) \geqslant 0 \\ 0 & ; \quad h_j(\vec{D}) < 0 \end{cases} \tag{2-14}$$

and

$$r_{p+1} > r_p \ . \tag{2-15}$$

The first term on the right hand side of Eq. (2-13) is the objective function and the second term is the penalty function. Note that each contribution to this penalty function has the property that it is zero in the acceptable region. Therefore, in this formulation there is no penalty for approaching the constraints from the acceptable region, rather a penalty is incurred only if an inequality constraint is violated. As the scalar r_p is increased ($r_{p+1} > r_p$) the sequence of solutions is driven toward the acceptable region of the design space where the inequality constraints are satisfied. In this formulation, large values of r_p lead to functions $\phi(\vec{D},r_p)$ that are difficult to minimize; therefore, by increasing r_p gradually, a sequence of tractable unconstrained minimization problems is generated. The unconstrained minima in the sequence of designs generated lie outside the acceptable region of the design space and therefore this formulation may be classified as an exterior penalty function method. From an engineering design point of view, exterior penalty function methods have the disadvantage that intermediate designs obtained prior to the optimum design are not acceptable (i.e. they violate one or more of the inequality constraints). On the other hand, exterior penalty function methods do not require a starting point $\vec{D}_o$ that satisfies the inequality constraints, (2-1a).

It should be pointed out that penalty function formulations can be subject to operational difficulties because the functions generated are sometimes difficult to minimize. Relative minima present in the basic problem statement do not vanish and in some cases additional relative minima are created by the formulation. Usually, the convexity of the functions involved in the basic problem statement cannot be assured and strict equivalence of the substitute problem cannot be guaranteed. Unconstrained minimization algorithms and penalty function formulations are dealt with further in Chapter 6.

2.6.3 Basic Non-linear Programming (NLP) Approach

Most of the large scale applications of mathematical programming to structural design optimization problems have attacked the problem directly using one of the various feasible direction methods.

To begin, assume that an acceptable design $\vec{D}_q$ is available, that is, let $\vec{D}_q$ be a design such that

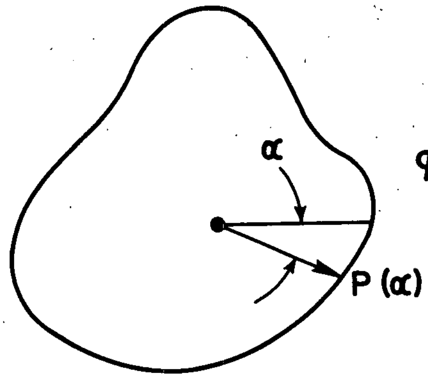
$$h_j(\vec{D}_q) < 0 \quad ; \quad j = 1,2,\ldots J \ . \tag{2-16}$$

The next design in the sequence $\vec{D}_{q+1}$ can be generated by moving in the direction of steepest descent, that is let $\vec{D}_{q+1}$ be determined as follows:

$$\vec{D}_{q+1} = \vec{D}_q + \alpha_q \vec{S}_q \tag{2-17}$$

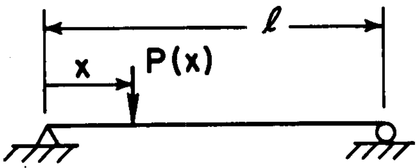
where

$$\vec{S}_q = -\nabla M(\vec{D}_q) \tag{2-18}$$



$$\phi(\vec{D}, r_p) = M(\vec{D}) - r_p \sum_{j=1}^{J}\left[\frac{1}{2\pi}\int_0^{2\pi}\frac{d\alpha}{h_j(\alpha, \vec{D})}\right]$$

Fig.2.3 Envelope Loading



$$\phi(\vec{D}, r_p) = M(\vec{D}) - r_p \sum_{j=1}^{J}\left[\frac{1}{\ell}\int_0^{\ell}\frac{dx}{h_j(x, \vec{D})}\right]$$

Fig.2.4 Moving Load

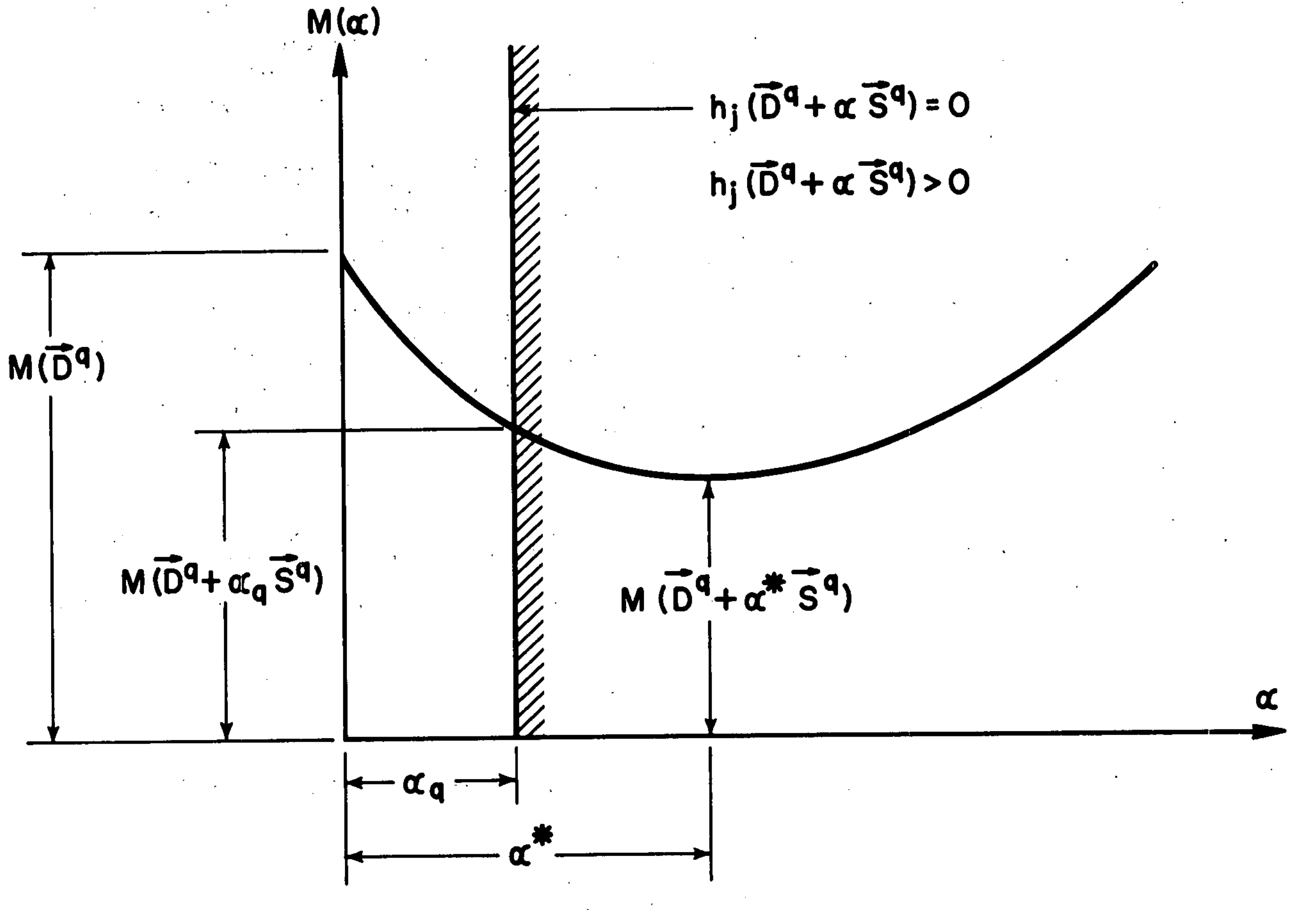


Fig.2.5 Constrained One-Dimensional Minimization

and α_q is the solution of a one-dimensional inequality constrained minimization problem. This one-dimensional problem is depicted graphically in Fig.2.5 and it can be stated concisely as follows: find α such that

$$h_j(\vec{D}_q + \alpha\, \vec{S}_q) = h_j(\alpha) \leq 0 \quad ; \quad j = 1,2,\ldots J \tag{2-19}$$

and

$$M(\vec{D}_q + \alpha\, \vec{S}_q) = M(\alpha) \rightarrow \text{Min} \quad . \tag{2-20}$$

In almost all structural design optimization problems, modification in the direction of steepest descent leads to a design $\vec{D}_{q+1}$ with one or more constraints critical, that is

$$h_j(\vec{D}_{q+1}) = 0 \quad \text{for } j \in J_c \quad . \tag{2-21}$$

The point $q+1$ in the two-dimensional design space shown in Fig.2.6 represents such a design.

The next design in the sequence is determined from the expression

$$\vec{D}_{q+2} = \vec{D}_{q+1} + \alpha_{q+1}\, \vec{S}_{q+1} \tag{2-22}$$

where the direction of modification $\vec{S}_{q+1}$ must satisfy the following inequality constraints

$$\vec{S}^T_{q+1}\, \nabla h_j(\vec{D}_{q+1}) \leq 0 \quad ; \quad j \in J_c \tag{2-23}$$

and

$$\vec{S}^T_{q+1}\, \nabla M(\vec{D}_{q+1}) \leq 0 \quad . \tag{2-24}$$

Directions $\vec{S}_{q+1}$ that satisfy Eq. (2-23) are feasible in the sense that design modification in such a direction is possible without violating the currently critical constraints. Directions $\vec{S}_{q+1}$ that satisfy Eq. (2-24) are called usable because they are directions such that the objective function is reduced or at least held invariant. Any direction $\vec{S}_{q+1}$ that satisfies Eq. (2-23) and (2-24) is called a usable-feasible direction. Design modification in such a direction does not violate the active constraints and does not increase the value of the objective function $M(\vec{D})$ locally. Three particular methods for determining usable-feasible directions $\vec{S}_{q+1}$ that have found application in structural design optimization are presented in detail in Chaper 7. Once a usable-feasible direction $\vec{S}_{q+1}$ has been determined the scalar α_{q+1} in Eq. (2-22) that determines how far to go can again be determined as the solution of a one-dimensional inequality constrained minimization problem. Note also that this one-dimensional minimization problem (α_{q+1}) may be unconstrained as shown in Fig.2.6 or it may be constrained as depicted in Fig.2.7. In the case shown in Fig.2.6 the design procedure can continue by making another move in the direction of steepest descent $\vec{S}_{q+2} = -\nabla M(\vec{D}_{q+2})$ while in the case illustrated in Fig.2.7 the design procedure is continued by generating another usable-feasible direction considering the new set of critical constraints at $\vec{D}_{q+2}$.

2.6.4 Classical Formulation

It is interesting to observe that the classical formulation of the inequality constrained minimization problem may be viewed as a device for transforming the basic problem (Eq. (2-1)) into a set of non-linear simultaneous equations. Using slack variables β_j (i.e. variables to convert inequalities into equations) and Lagrange multipliers μ_j the classical formulation can be cast in terms of a set of non-linear simultaneous equations as follows:

Find $(\vec{D},\vec{\beta},\vec{\mu})$ such that $\Omega(\vec{D},\vec{\beta},\vec{\mu})$ is stationary

where

$$\Omega(\vec{D},\vec{\beta},\vec{\mu}) = M(\vec{D}) + \sum_{j=1}^{J} \mu_j\, [\beta_j^2 + h_j(\vec{D})] \tag{2-25}$$

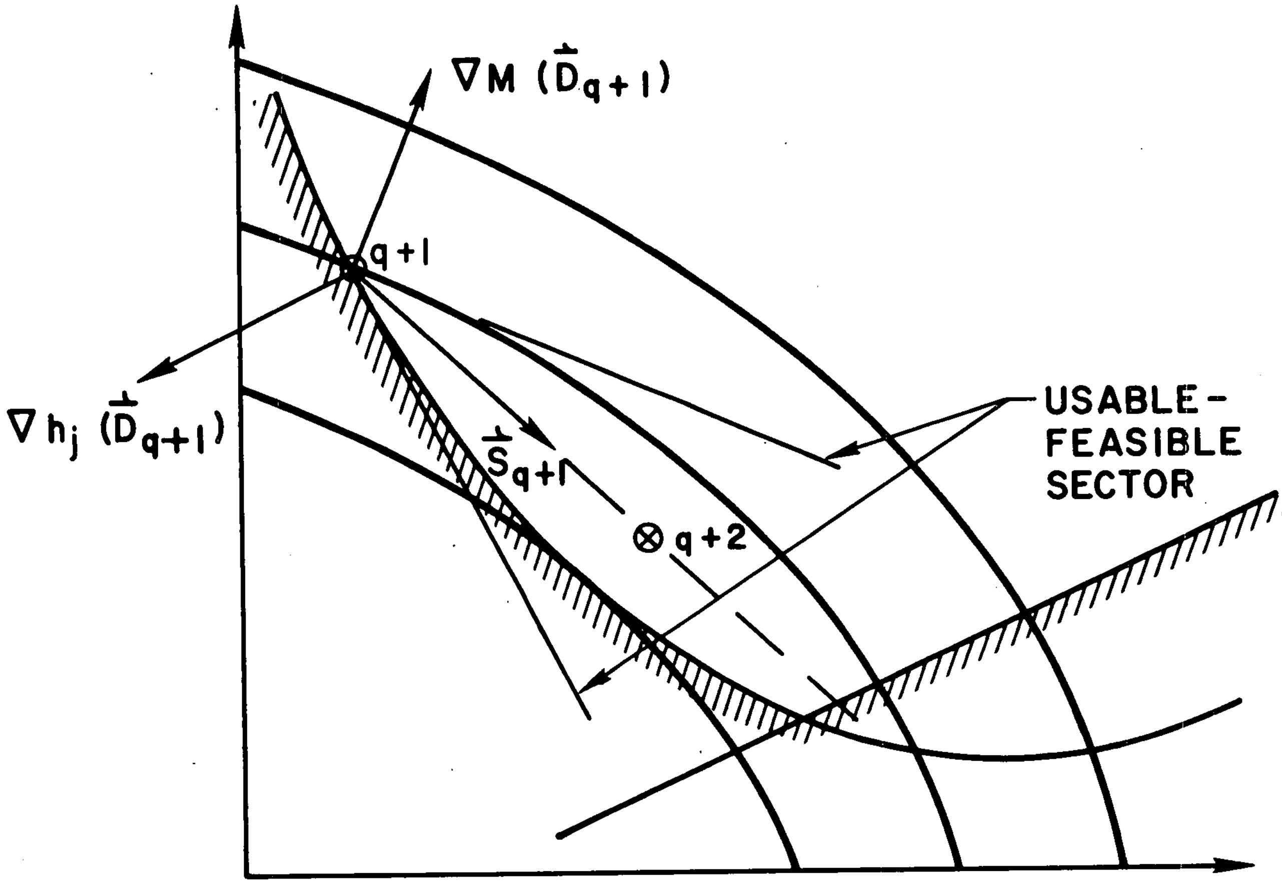


Fig.2.6 Schematic of Usable Feasible Vector

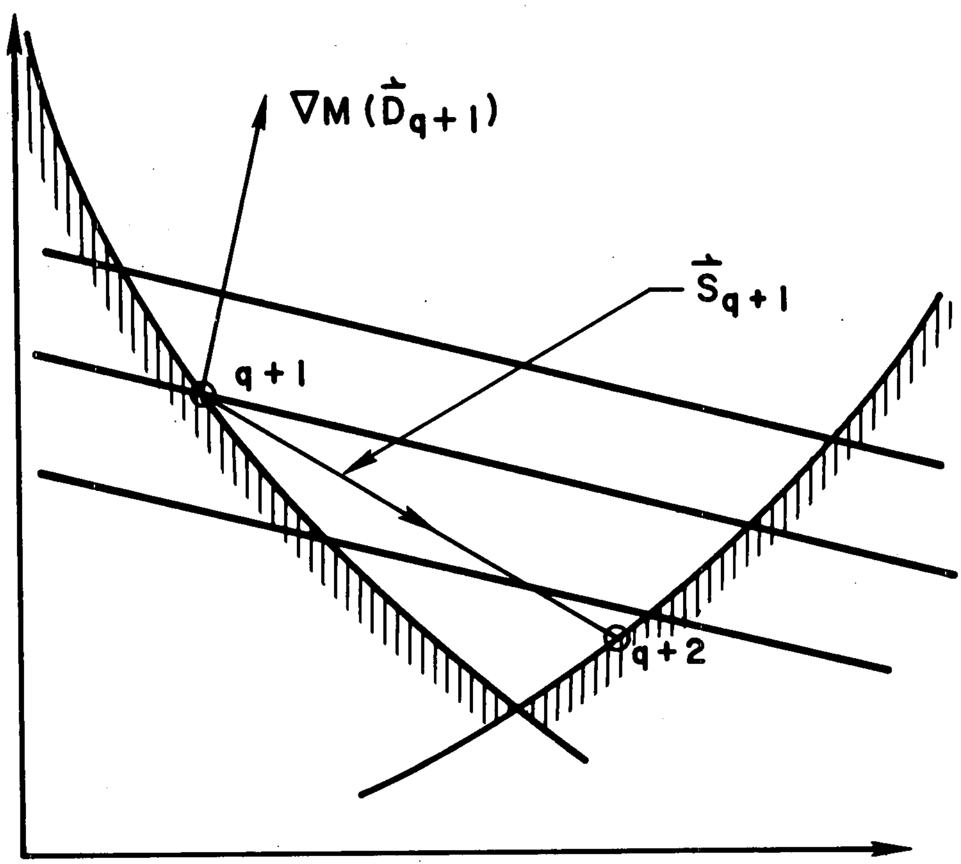


Fig.2.7 Two Critical Designs in Sequence

which implies

$$\frac{\partial\Omega}{\partial D_i} = \frac{\partial M}{\partial D_i} + \sum_{j=1}^{J} \mu_j \frac{\partial h_j}{\partial D_i} = 0 \quad ; \quad i = 1,2,\ldots I \tag{2-26}$$

$$\frac{\partial\Omega}{\partial \mu_j} = \beta_j^2 + h_j(\vec{D}) = 0 \quad ; \quad j = 1,2,\ldots j \tag{2-27}$$

$$\frac{\partial\Omega}{\partial \beta_j} = 2\mu_j\beta_j = 0 \quad ; \quad j = 1,2,\ldots J \tag{2-28}$$

These simultaneous non-linear equations (Eq. (2-26), (2-27) and (2-28)) are only necessary conditions for a minimum of $M(\vec{D})$ subject to the inequality constraints, (2-1a), and in general they admit multiple solutions. It is observed that this formulation increases the number of unknowns from I to (I+2J). Finding all of the solutions and then sorting out which of these represents the best solution of the basic problem is usually an exhaustive task. The classical formulation was applied to inequality constrained minimization problems in the context of structural design by Klein [2.18]. It should be pointed out that while the classical formulation has serious practical limitations, it can be useful, particularly when some foreknowledge is available as to how many and which of the inequality constraints are critical.

Since so much of the structural optimization literature tends to assume that the responsible engineer can often anticipate how many and which inequality constraints will be active for the optimum design, it may be well to briefly elaborate on the relationship of this view to the classical formulation of the inequality constrained minimization problem. Consider an example with 3 inequality constraints (J = 3). The possible combinations of critical constraints can be listed as follows in terms of the set of integers denoted J_c: null; (1); (2); (3); (1, 2); (2, 3); (3, 1); (1, 2, 3). The slack variables and the Lagrange multipliers for each of these eight combinations can be tabulated as follows:

Combination Number	J_c	β_1	β_2	β_3	μ_1	μ_2	μ_3
1	Null	*	*	*	0	0	0
2	(1)	0	*	*	*	0	0
3	(2)	*	0	*	0	*	0
4	(3)	*	*	0	0	0	*
5	(1, 2)	0	0	*	*	*	0
6	(2, 3)	*	0	0	0	*	*
7	(3, 1)	0	*	0	*	0	*
8	(1, 2, 3)	0	0	0	*	*	*

where * indicates an unknown to be determined from the solution of the equations

$$\frac{\partial M}{\partial D_i} + \sum_{j \in J_c} \mu_j \frac{\partial h_j}{\partial D_i} = 0 \quad ; \quad i = 1,2,\ldots I \tag{2-29}$$

$$\beta_j^2 + h_j(\vec{D}) = 0 \quad ; \quad j \in J_c \tag{2-30}$$

which follow from Eq. (2-26), (2-27) and (2-28). Note that Eq. (2-29) can be written in an alternative form as

$$\nabla M + \sum_{j \in J_c} \mu_j \nabla h_j = 0 \quad . \tag{2-31}$$

For any particular assumed combination of critical constraints (1 through 8), the value of μ_j obtained from the solution of Eq. (2-29) and (2-30) can be examined to determine whether or not the Kuhn-Tucker conditions [2.19] is satisfied. This <u>necessary</u> condition for any constrained optimum is that the negative gradient $(-\nabla M)$ of the objective function be a non-negative linear combination of the gradients to the critical constraints $(\nabla h_j; j \in J_c)$. Therefore, if the Lagrange multipliers μ_j in Eq. (2-31) are non-negative the Kuhn-Tucker condition is satisfied. If the constraint functions are convex and the objective function is at least locally convex, then satisfaction of the Kuhn-Tucker condition is sufficient to establish the constrained optimum under examination as a local optimum. If both the constraint functions and the objective function are convex, then satisfaction of the foregoing condition is sufficient to establish the constrained optimum being tested as a global optimum. Thus, it becomes apparent that for problems where insight or prior experience suggest which combination of constraints are likely to be critical at the optimum, it may not be necessary to solve Eq. (2-29) and (2-30) for all possible combinations of critical constraints. This discussion is offered to show that many of the traditional methods of structural optimization may be viewed as special cases of the more general viewpoint represented by the application of mathematical programming techniques. It is suggested that whenever design optimization is sought by assuming that a certain set of critical constraints characterize the optimum, an effort should be made to determine whether or not the result obtained at least satisfies the necessary condition represented by the Kuhn-Tucker test.

2.7 A More General View

While most applications of mathematical programming techniques to structural design optimization have attacked the problem as an inequality constrained minimization problem having the form of Eq. (2-1), it should be recognized that a more general class of problems in structural engineering can be viewed in the context of mathematical programming. The general mathematical programming problem can be stated concisely as follows:

$$\text{Find } \vec{X} \text{ such that } f_k(\vec{X}) = 0 \; ; \quad k = 1,2,\ldots K \tag{2-32a}$$

$$h_j(\vec{X}) \leq 0 \; ; \quad j = 1,2,\ldots J \tag{2-32b}$$

and

$$M(\vec{X}) \rightarrow \text{Min} \; . \tag{2-32c}$$

It is understood that the vector $\vec{X}$ locates a point in an N-dimensional space, the functions $f_k(\vec{X}) = 0$ denote equality constraints, the functions $h_j(\vec{X}) \leq 0$ represent inequality constraints and $M(\vec{X})$ is an objective function. The previously discussed class of structural design optimization problems (see Eq. (2-1)) are clearly a special case of Eq. (2-32) in which $\vec{X}$ is replaced by the design variables $\vec{D}$ and equality constraints are not present.

The more general formulation given by Eq. (2-32) embraces a wide variety of structural engineering problems including design optimization problems, analysis problems and integrated analysis-design optimization.

Design problems involving equality constraints between the design variables are easily imagined. The three bar truss discussed in Section 1.2.4 is a simple example. Symmetry of the final design can be imposed using an equality constraint, namely $A_1 = A_3$ and then dealing with the problem as a three variable problem (A_1, A_2, A_3). Alternatively, in the case of simple equality constraints the number of independent design variables can be reduced. When this approach is taken, the number of variables for the three bar truss example is reduced to two (A_1 and A_2) and the design problem is of the form given by Eq. (2-1). In situations where the equality constraints between design variables are complicated, it may not be possible to use equality constraints to reduce the number of independent design variables. When this situation exists the structural design optimization problem has the form of a general mathematical programming problem (i.e. Eq. (2-32)).

Structural analysis problems can be viewed as special cases of the formulation given by Eq. (2.32). For example, the analysis of a structural system based upon minimizing the total potential energy may be viewed as an equality constrained minimization problem. Let $\vec{X}$ be replaced by $\vec{u}$, the vector of generalized displacement variables and let the objective function $M(\vec{X})$ be replaced by $\pi_p(\vec{u})$ the total potential energy. Then the structural analysis problem can be stated as follows:

$$\text{Find } \vec{u} \text{ such that } \pi_p(\vec{u}) \rightarrow \text{Min} \tag{2-33}$$

subject to a set of equality constraints

$$f_k(\vec{u}) = 0 \; ; \quad k = 1,2,\ldots K \tag{2-34}$$

that impose the geometric admissibility conditions on the displacement variables.

The total potential energy $\pi_p(\vec{u})$ is quadratic in the generalized displacement variables for linear analysis problems. Extension to include geometric non-linearities is easily accomplished using non-linear strain-displacement relations representing various levels of refinement. For instance, the use of finite displacement theory strain-displacement relations leads to a total potential energy function $\pi_p(\vec{u})$ that is quartic in the generalized displacement variables. Extension to include material non-linearity is also straightforward in principle, provided the non-linear stress-strain relations can be adequately represented by a strain energy density type of potential function; however, most plastic stress-strain relations do not satisfy this requirement.

If the geometric admissibility conditions (Eq. (2-34)) are used to reduce the number of displacement variables, then the structural analysis problem can be viewed as an unconstrained minimization problem expressed in terms of the kinematically independent displacement variables $\vec{u}_c$, that is

$$\text{Find } \vec{u}_c \text{ such that } \tilde{\pi}_p(\vec{u}_c) \rightarrow \text{Min} \quad . \tag{2-35}$$

Some example applications of finite element structural analyses based on this mathematical programming viewpoint will be found in [2.20] and [2.21].

The analysis of a structure based upon minimizing the total complementary energy may also be viewed as an equality constrained minimization problem. Let $\vec{X}$ be replaced by $\vec{F}$, the vector of generalized force variables, and let the objective function $M(\vec{X})$ be replaced by $\pi_c(\vec{F})$ the total complementary energy. Then the structural analysis problem can be stated as follows:

$$\text{Find } \vec{F} \text{ such that } \pi_c(\vec{F}) \rightarrow \text{Min} \tag{2-36}$$

subject to a set of equality constraints

$$f_k(\vec{F}) = 0 \quad ; \quad k = 1,2,\ldots K \tag{2-37}$$

that impose the static admissibility conditions on the force variables. The total complementary energy $\pi_c(\vec{F})$ is quadratic in the force variables for linear analysis problems. Extension to include material non-linearities is easily accomplished provided the non-linear strain-stress relations can be adequately represented by a complementary energy density type of potential function. Extensions to include geometric non-linearities are generally unsuccessful because the nonlinear force displacement relations are such that the total complementary energy cannot be expressed solely in terms of force variables.

If static admissibility conditions (Eq. (2-37)) are used to reduce the number of force variables, then the structural analysis problem can again be viewed as an unconstrained problem expressed in terms of the statically independent force variables $\vec{R}$, that is

$$\text{Find } \vec{R} \text{ such that } \tilde{\pi}_c(\vec{R}) \rightarrow \text{Min} \quad . \tag{2-38}$$

Limit analysis offers another example of the applicability of the general mathematical programming formulation Eq. (2-32) in the context of structural analysis. The limit analysis of a structure, from the statical point of view, has as its goal determination of the maximum load carrying capacity of the structure subject to the requirements that the force distribution satisfies the equilibrium conditions and the yield conditions. In the case of a truss [2.22] the problem of the determination of the maximum load carrying capacity has the following form:

Find $\vec{F}$ and λ such that

$$\sum_{j=1}^{J} a_{ij} F_j = \lambda P_i \quad ; \quad i = 1,2,\ldots I \tag{2-39}$$

$$L_j - F_j \leq 0 \quad ; \quad j = 1,2,\ldots J \tag{2-40a}$$

$$F_j - U_j \leq 0 \quad ; \quad j = 1,2,\ldots J \tag{2-40b}$$

and

$$-\lambda \rightarrow \text{Min} \tag{2-41}$$

where F_j represents the force in the jth member,
P_i represents the contribution of the applied load condition to the ith equilibrium equation,
a_{ij} represents the contribution to the ith equilibrium equation of a unit value of the force in the jth member,
L_j represents the force required to yield the jth member in compression,
U_j represents the force required to yield the jth member in tension,
λ is a positive scalar factor which determines the magnitude of the applied load condition.

The I equations embodied in Eq. (2-39) are the equilibrium equations, the 2J inequalities stated by Eq. (2-40a) and (2-40b) are the yield conditions, and the objective function (Eq. (2-41)) is $-\lambda$ since the maximum load carrying capacity is sought. It is apparent that the limit analysis of trusses from the statical point of view has the form of a linear programming problem in terms of the force variables ($\vec{F}$) and the load factor (λ). It should be clearly recognized that this limit analysis can only be carried out for a structure of specified design, that is the geometric layout, the member areas and the yield stresses of the member materials must be given.

The combined analysis and design optimization of a structural system can often be stated as a general mathematical programming problem having the form of Eq. (2-32). In the case of combined analysis and design optimization, it is useful to view the vector $\vec{X}$ in Eq. (2-32) as the concatenation of two vectors $\vec{D}$ and $\vec{Y}$ where $\vec{D}$ is the vector of design variables and $\vec{Y}$ is the vector of analysis variables. This vector $\vec{Y}$ should be understood to contain an independent component for each analysis unknown for each load condition. Depending upon the analysis method adopted, the analysis unknowns may characterize the displacement state, the force distribution or a combination of both.

An interesting example of a combined analysis-design optimization formulation can be generated by considering the minimum weight sizing of trusses based upon limit analysis as described in Chapter 3.

List of References

Ref.

2.1 Prager, W. and Taylor, J. E., 'Problems in Optimal Structural Design', *J. of Applied Mechanics,* Vol.35, 1968, pp.102-106

2.2 Prager, W. and Shield, R. T., 'Optimal Design of Multipurpose Structures', *Int'l J. of Solids and Structures,* Vol.4, 1968, pp.469-475

2.3 Sheu, C. Y. and Prager, W., 'Recent Developments in Optimal Structural Design', *Applied Mechanics Reviews,* Vol.21, No.10, October 1968, pp.985-992

2.4 Moses, F. and Kinser, D. E., 'Optimum Structural Design with Failure Probability Constraints', *AIAA Journal,* Vol.6, No.6, 1967, pp.1152-1158

2.5 Ghista, D. N., 'Structural Optimization with Probability of Failure Constraints', NASA TN D-3777, December 1966

2.6 Moses, F. and Stevenson, J. D., 'Reliability Based Structural Design', Case Western Reserve University, DSMSMD Report No.16, January 1968

2.7 Sved, G. and Ginos, Z., 'Structural Optimization under Multiple Loading', *Int. J. Mech. Sci.,* Vol.10, 1968, pp.803-805

2.8 Reinschmidt, K. F., Cornell, A. C. and Brotchie, J. F., 'Iterative Design and Structural Optimization', J. of the Structural Division, ASCE, Vol.92, No.ST6, December 1966, pp.281-318

2.9 Moses, F., 'Some Notes and Ideas on Mathematical Programming Methods for Structural Optimization', Meddelelse SKB II/M8, Norges Tekniske Høgskole, Trondheim, Norway, January 1967

2.10 Fiacco, A. and McCormick, G. P., 'Programming under Non-linear Constraints by Unconstrained Minimization: A Primal-Dual Method', RAC-TP-96, Bethesda, Maryland, 1963

2.11 Fiacco, A. and McCormick, G. P., 'The Sequential Unconstrained Minimization Technique for Non-linear Programming: A Primal-Dual Method', *Manag. Sci.* 10, No.2, 1964, pp.360-365

2.12 Fiacco, A. and McCormick, G. P., 'Computational Algorithm for the Sequential Unconstrained Minimization Technique for Non-linear Programming', *Manag. Sci.* 10, No.4, 1964, pp.601-617

2.13 Schmit, L. A., Morrow, W. M. and Kicher, T. P., 'A Structural Synthesis Capability for Integrally Stiffened Cylindrical Shells', AIAA/ASME 9th Structures, Structural Dynamics and Materials Conference, Palm Springs, California, April 1-3, 1968, AIAA Pre-print No.68-327

2.14 Morrow II, W. M., and Schmit, L. A., 'Structural Synthesis of a Stiffened Cylinder', NASA CR-1217, December 1968

2.15 Marcal, P. V. and Gellatly, R. A., 'Application of the Created Response Surface Technique to Structural Optimization', Proc. of the Second Conference on Matrix Methods in Structural Mechanics, WPAFB, Ohio, October 1968, AFFDL-TR-68-150, December 1969, pp.83-110

2.16 Zoutendijk, G., 'Non-linear Programming: A Numerical Survey', *J. SIAM Control,* Vol.4, No.1, 1966, pp.194-210

2.17 Thornton, W. A. and Schmit, L. A., 'The Structural Synthesis of an Ablating Thermostructural Panel', NASA CR-1215, December 1968

2.18 Klein, B., 'Direct Use of Extremal Principles in Solving Certain Optimization Problems Involving Inequalities', Operations Research, Vol.3, 1955, pp.168-175

List of References (Contd.)

Ref.

2.19 Kuhn, H. W. and Tucker, A. W., 'Non-linear Programming', Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, California, 1950, pp.481-492

2.20 Schmit, L. A., et al., 'Developments in Discrete Element Finite Deflection Structural Analysis by Function Minimizations', USAF, AFFDL-TR-68-126, September 1968

2.21 Fox, R. L. and Stanton, E., 'Developments in Structural Analysis by Direct Energy Minimization', *AIAA Journal*, Vol.6, No.6, June 1968, pp.1036-1042

2.22 Dorn, W. S. and Greenberg, H. J., 'Linear Programming and Plastic Limit Analysis of Structures', *Quarterly of Applied Math.*, Vol.XV, No.2, 1957, pp.155-167

Chapter 3

CLASSICAL OPTIMIZATION THEORY RELEVANT TO THE DESIGN OF AEROSPACE STRUCTURES

by

G. G. Pope

3.1 Introduction

Special classes of structural design problems which can be solved advantageously by analytical as opposed to numerical techniques have been studied widely, and comprehensive references on the subject will be found in the review papers by Sheu and Prager [3.1] and by Wasiutynski and Brandt [3.2]. Since attention is concentrated in this present volume on the use of mathematical programming techniques in the design of aerospace structures, most of which must behave elastically under service conditions, it is appropriate here to restrict our attention to those aspects of analytical work on structural optimization which are relevant in this narrower context.

This Chapter is concerned mainly with the classical theorem due to Michell which is applicable directly to the least weight design of highly idealised frameworks. Apart from the obvious value of this theorem in the derivation of exact solutions for use as yardsticks in the assessment of the efficiency of practical structures, the general results which may be derived from it can also provide useful guidance in the choice of the layout not only of frameworks but also of stressed-skin and plate type structures. For example, an appreciation of the properties of the optimum types of strain field derived by Michell can reduce significantly the range of geometries which need to be considered in the laborious numerical studies that are often necessary to obtain an optimum structural layout. A useful indication may, moreover, sometimes be obtained of circumstances where the least weight design is non-unique and where consequently the designer may be able to impose geometrical restrictions to suit requirements not included in the idealised design problem, without increasing the structural weight.

The analysis given in this Chapter starts from the assumption that the structure is fabricated from a material with elastic/perfectly plastic properties. It is demonstrated, however, that the least weight design obtained on this basis when one load condition only is applied is identical with the least weight design for purely elastic deformation provided stress limits only are considered. Michell's theorem of minimum weight design is deduced for a framework consisting of a finite number of members, by formulating the search for the least weight design as a problem in linear programming, and by using the duality properties of problems of this class, following arguments given previously by Hemp [3.3], [3.4] who along with Pearson [3.5] and with Dorn, Gomory and Greenberg [3.6] has employed linear programming techniques in the least weight design of ideal frameworks of this type.

3.2 Basic Theory for Elastic/Perfectly Plastic Frameworks

3.2.1 Single Load Condition

Consider the minimum weight design of a pin-jointed framework which is supported in such a way that all the external reactions may be evaluated directly from the overall equilibrium conditions. No restrictions are imposed on the permissible displacements and buckling effects are neglected; the members are all fabricated from the same material and the weights of the connections between them are assumed negligible. The basic geometry is specified, and the cross-sectional areas of the M members that constitute the framework are treated as design variables and are denoted by a column vector $\vec{D}$. Loads are applied at the nodal points joining adjacent members and the single load condition which is considered initially is specified by a column vector $\vec{P}$; this has an element corresponding to each of the K equations required to establish equilibrium. These equations may be expressed in the form

$$G\vec{F} = \vec{P} \tag{3-1}$$

where $\vec{F}$ is a vector of M terms defining the loads in the members and G is an appropriate transformation matrix.

If the yield stresses in tension and compression are given by σ^+ and σ^- respectively and are the same for the entire framework, the loads in the members must satisfy the following conditions:

$$\sigma^-\vec{D} \leqslant \vec{F} \leqslant \sigma^+\vec{D} \ . \tag{3-2}$$

Note that σ^- is so defined that it will in practice have a negative value.

The total volume V of the members constituting the framework is given by

$$V = \vec{\ell}^T\vec{D}$$

where $\vec{\ell}$ is a vector containing the lengths of the members. The problem of finding the least weight design reduces therefore to minimizing V subject to the constraints (3-1) and (3-2). This is a linear programming problem which may be expressed purely in terms of positive variables by substituting

$$\vec{F} = \vec{F}' - \vec{F}'' \ .$$

Expressing each of Eq. (3-1) as a pair of inequalities, this problem may be expressed in the following form:

minimize

$$V = \{\vec{\sigma}\ \vec{\sigma}\ \vec{\ell}\}^T \{\vec{F}'\ \vec{F}''\ \vec{D}\}$$

where

$$\{\vec{F}'\ \vec{F}''\ \vec{D}\} \geq 0 \qquad (3\text{-}3)$$

and

$$\begin{bmatrix} -I & I & \sigma^+ I \\ I & -I & -\sigma^- I \\ G & -G & 0 \\ -G & G & 0 \end{bmatrix} \begin{bmatrix} \vec{F}' \\ \vec{F}'' \\ \vec{D} \end{bmatrix} \geq \begin{bmatrix} \vec{\sigma} \\ \vec{\sigma} \\ \vec{P} \\ -\vec{P} \end{bmatrix}$$

The optimum solution is necessarily one in which all the members are fully-stressed, since a reduction in the cross-section of any member which is not fully-stressed would reduce V without violating any of the governing equations. It is however possible for the cross-sectional area of unnecessary members to vanish completely.

The dual of the above problem may be expressed as follows:

maximize

$$W = \frac{1}{\sigma^+ \varepsilon^-} \{\vec{\sigma}\ \vec{\sigma}\ \vec{P}\ -\vec{P}\}^T \{\vec{\gamma}'\ \vec{\gamma}''\ \vec{u}'\ \vec{u}''\}$$

where

$$\{\vec{\gamma}'\ \vec{\gamma}''\ \vec{u}'\ \vec{u}''\} \geq 0 \qquad (3\text{-}4)$$

and

$$\frac{1}{\sigma^+ \varepsilon^-} \begin{bmatrix} -I & I & G^t & -G^t \\ I & -I & -G^t & G^t \\ \sigma^+ I & -\sigma^- I & 0 & 0 \end{bmatrix} \begin{bmatrix} \vec{\gamma}' \\ \vec{\gamma}'' \\ \vec{u}' \\ \vec{u}'' \end{bmatrix} \leq \begin{bmatrix} \vec{\sigma} \\ \vec{\sigma} \\ \vec{\ell} \end{bmatrix} ;$$

the product $\sigma^+ \varepsilon^-$, where ε^- is the yield strain in compression, is introduced so that the dual variables may be interpreted as extensions and displacements.

Since the optimum framework is necessarily fully-stressed, half the constraints in the primal problem derived from the inequalities (3-2) must be satisfied as equalities in the optimum solution, i.e. one for each member of the framework. It follows therefore from the second of the properties of dual problems described in Chapter 5 that the corresponding components of $\vec{\gamma}'$ and $\vec{\gamma}''$ must be zero in the optimal solution to the dual problem. Consequently the latter problem may be re-expressed as follows:

maximize

$$W = \frac{1}{\sigma^+ \varepsilon^-} \vec{P}^T \vec{u}_v \qquad (3\text{-}5)$$

where

$$\vec{\gamma} = G^T \vec{u}_v \qquad (3\text{-}6)$$

$$\varepsilon^- \vec{\ell} \leq \vec{\gamma} \leq \varepsilon^+ \vec{\ell} \qquad (3\text{-}7)$$

and where

$$\vec{u}_v = \vec{u}' - \vec{u}'' \ , \quad \vec{\gamma} = \vec{\gamma}' - \vec{\gamma}'' \quad ;$$

ε^+ represents the yield strain in tension. If now the variables $\vec{u}_v$ are interpreted as virtual displacements of the nodes of the framework, the resulting work done by the applied forces is proportional to the value of the merit function W. Substituting Eq. (3-1) and (3-6) in Eq. (3-5) we obtain

$$W = \frac{1}{\sigma^+ \varepsilon^-} \vec{F}^T \vec{\gamma} \quad .$$

Since $\vec{F}^T \vec{\gamma}$ represents the increment of strain energy associated with the virtual displacements $\vec{u}_v$, it is clear that the variables $\vec{\gamma}$ represent the corresponding deformations of the individual members. The dual problem seeks therefore to maximize the virtual work done by the external forces when the strains in all members are restricted to being less in absolute value than ε^+ in tension and ε^- in compression. Using again the second of the duality properties described in Chapter 5, it may further be deduced that the following conditions are necessary and sufficient to ensure that a pin-jointed framework has the least possible weight:

(1) The stresses in all the members due to the applied loading are either σ^+ (tension) or σ^- (compression).

(2) The framework must permit a virtual displacement of all its possible nodes which produces a strain of ε^+ in its tension members, a strain of ε^- in its compression members and no tensile strain greater than ε^+ or compressive strain greater in absolute value than ε^- in any segment along which a potential member could lie.

In the special case when ε^+ and $-\varepsilon^-$ are equal, the above conditions reduce to those shown by A. G. M. Michell [3.7] to be sufficient to establish a least weight design; in the more general case when ε^+ and $-\varepsilon^-$ are not equal, it may be demonstrated that these conditions are equivalent to Michell's conditions by considering a virtual dilatational strain in addition to the strain system considered in the present analysis.

The arguments, based on duality properties, which have been used here to show that the above conditions are necessarily satisfied by a minimum weight design are due to Hemp [3.4]; they are only strictly applicable when the number of potential members is finite.

The virtual strain system defined in Eq. (3-6) becomes identical with the actual strains when a minimum weight design is achieved. It follows that the minimum weight design is necessarily an elastic design and also that a statically determinate least weight design must always be possible, although there may be other designs of the same weight.

It should be noted that the linear programming technique described here sometimes yields an array of members which is a mechanism rather than a structure; additional members are then necessary to carry even the most trivial alternative loading. Under such circumstances it is, of course, advantageous when possible to deduce an alternative minimum weight design.

It may readily be shown that the least weight design for a framework to carry a single load condition is also the stiffest framework which will carry the loading at the same level of stress; a concise proof of this result is given by Hegemier and Prager [3.8] in a paper which is concerned primarily with the introduction of constraints on natural frequency into the design of idealised frameworks.

3.2.2 Multiple Load Conditions

If the equilibrium equation (3-1) and the inequalities (3-2) are increased in number to include several load conditions applied in turn to the framework, the search for a minimum weight design remains a problem in linear programming. The strain criteria deduced in Section 3.2.1 are, however, no longer valid and consequently the optimum design experiences, in general, plastic deformation under at least one of the design load conditions.

Hemp [3.4] has shown, with the aid of the dual problem, that in the special case where two loadings only are considered, the least weight design may be obtained by superposing the least weight designs for the single load conditions $\frac{1}{2}(\vec{P}_1 + \vec{P}_2)$ and $\frac{1}{2}(\vec{P}_1 - \vec{P}_2)$ where $\vec{P}_1$ and $\vec{P}_2$ represent the applied load systems. Some general results for least weight elastic/perfectly plastic structures under multiple load conditions are given by Shield [3.9].

3.3 Optimum Layout of Elastic Frameworks

A. G. M. Michell used conditions equivalent to those deduced in Section 3.2.1 to evolve, for a single loading, least weight frameworks in which no restrictions are imposed on the number and position of the nodal points. Such structures usually involve an indefinitely large number of infinitesimal members so they are seldom suitable for direct use in engineering design; they are, nevertheless, of significant value for the reasons indicated in Section 3.1, and they have two general properties that are worthy of note:

(1) Tension and compression members necessarily meet orthogonally to satisfy the conditions imposed on the strains.

(2) Any fully-stressed design in which all the member loads are of the same sign necessarily satisfies the optimality conditions; an infinite number of optimum configurations exists therefore when such designs are possible.

The latter result may also be deduced directly from a theorem due to Clark Maxwell [3.10] which preceded Michell's contribution to this field.

Least weight frameworks of the type evolved by Michell are considered in detail by Cox [3.11] and close approximations to them have been obtained by H. S. Y. Chan [3.12] using the linear programming approach and assuming that member intersections only occur at a finite number of points; members are permitted to run between any pair of the assumed intersection points. The design of Michell frameworks is analogous to the analysis of the slip line fields associated with the flow of rigid/perfectly plastic materials. A graphical technique developed for use in the latter context has been employed by A. S. L. Chan [3.13] to obtain framework designs of least weight.

The optimum configuration of frameworks in which the layout is more severely restricted may, of course, also be obtained,by specifying the points at which member intersections can occur and, if necessary, by restricting the pairs of intersections between which members may lie. It is likely that some of the possible members will vanish completely in the optimization process; this is permissible because no compatibility conditions are involved directly in the primal analysis. It should be noted that it is much more difficult to permit members to vanish in the more complex problem, considered elsewhere in this volume, of the design of an optimum structure to carry several load systems in turn without yielding, since the analysis equations would there impose artificial constraints on the strains in the non-existent members [3.14].

Acknowledgement - This Chapter is British Crown Copyright, reproduced with the permission of the Controller, Her Majesty's Stationery Office.

List of References

Ref.

3.1 Sheu, C. Y. and Prager, W., "Recent Developments in Optimal Structural Design", *Applied Mechanics Reviews*, Vol.21, No.10, October 1968, pp.985-992

3.2 Wasiutynski, Z. and Brandt, A., "The Present State of Knowledge in the Field of Optimum Design of Structures", *Applied Mechanics Reviews*, Vol.16, No.5, May 1963, pp.341-359

3.3 Hemp, W. S., "Studies in the Theory of Michell Structures", *Proc. of the 11th Int. Cong. Appl. Mech., 1964*, Springer, Berlin, 1966, pp.621-628

3.4 Hemp, W. S., Abstract of lecture course "Optimum Structures", 2nd ed., Engineering Laboratory, University of Oxford, 1968

3.5 Pearson, C. E., "Structural Design by High Speed Computing Machines", *Proc. of the 1st Conference on Electronic Computation, ASCE*, New York, 1958, pp.417-436

3.6 Dorn, W. S., Gomory, R. E. and Greenberg, H. J., "Automatic Design of Optimal Structures", *Journal de Mécanique*, Vol.3, No.1, 1964, pp.25-52

3.7 Michell, A. G. M., "The Limits of Economy of Material in Frame Structures", *Philosophical Magazine*, Series 6, Vol.8, 1904, pp.589 et seq.

3.8 Hegemier, G. A. and Prager, W., "On Michell Trusses", *International J. of Mech. Sci.*, Vol.11, No.2, February 1969, pp.209-215

3.9 Shield, R. T., "Optimum Design Methods for Multiple Loading, *Zeitschrift für angewandte Mechanik und Physik*, Vol.14, January 1963, pp.38-45

3.10 Maxwell, C., Scientific Papers, Vol.2, 1869, pp.175 et seq.

3.11 Cox, H. L., *The Design of Structures of Least Weight*, 1st ed., Pergamon, Oxford, 1965

3.12 Chan, H. S. Y., "Optimum Structural Design and Linear Programming", College of Aeronautics Report No.175, Cranfield, England, 1964

3.13 Chan, A. S. L., "The Design of Michell Optimum Structures", British Aeronautical Research Council, R. & M. No.3303, 1960

3.14 Sved, G. and Ginos, Z., "Structural Optimization under Multiple Loading", International J. of Mech. Sci., Vol.10, No.10, October 1968, pp.803-805

Chapter 4

LITERATURE REVIEW AND ASSESSMENT OF THE PRESENT POSITION

by

L. A. Schmit

4.1 Introduction

There are several valuable reviews and annotated bibliographies already available in the literature. A rather comprehensive bibliography and assessment of optimum structural design concepts for aerospace vehicles through 1966 will be found in [4.1] and [4.2]. The literature review contained in [4.3] appeared in 1963 and the majority of the references cited deal with single load condition situations and assume a plastic collapse design philosophy; many references to the Russian and Polish literature are included. A comprehensive review of more recent developments in optimal structural design is given in [4.4]. This review makes clear the distinction between 'single purpose' and 'multipurpose structure' and it points out that currently research is proceeding on two fronts: (1) application of the numerical methods of mathematical programming to specific highly realistic problems and (2) analytical treatment of a variety of optimal design problems for structural elements and simple structures. The review presented in [4.5] deals specifically with the application of non-linear mathematical programming in structural design optimization through 1966.

The literature review to be presented in this Chapter will focus on applications of mathematical programming to structural design optimization and it will be limited to finite problems. In Section 4.2, an effort is made to trace the development of mathematical programming applications in structural design, using the philosophical framework set forth in Chapter 2 to help keep the review organized. Since the papers selected for discussion in Section 4.2 are limited in number, a more comprehensive list of references is given in Appendix A. In Section 4.3 under the heading of future trends, brief reviews of (1) structural optimization in the dynamic response regime; and (2) reliability based structural optimization are offered. In Chapter 10, reliability based structural optimization is discussed in more detail. The dynamic response regime and particularly the subject of structural optimization considering aeroelastic constraints is examined in greater depth in Chapter 11. Finally in Chapter 12, overall configuration considerations and optimization methods in preliminary design are considered.

4.2 Selective Review

It is to be understood that the literature survey given in this section is not intended to be exhaustive. Rather, it is a careful but probably somewhat subjective selection of a collection of papers that are thought to have strongly influenced the development of mathematical programming applications in structural design optimization during the last decade. Several of the references discussed are summarized in Tables 4.1 and 4.2 using the framework set forth in Chapter 2.

In [4.6], published in 1955, Klein pointed out that an important set of minimum weight structural design problems could be viewed as non-linear mathematical programming problems. The importance of inequality constraints in properly stating structural design optimization problems was clearly recognized. The influence of this paper was probably limited by the fact that the problem was treated in classical form using Lagrange multipliers and slack variables (see Section 2.6.4). The large number of unknowns and the need for finding all the solutions of the governing set of non-linear simultaneous equations were discouraging when larger problems were contemplated.

In [4.7], published in 1958, Pearson working within the plastic design philosophy treats the minimum weight design problem considering a multiplicity of overload conditions. Displacement constraints under service load conditions are ignored and compatibility conditions can be neglected under overload conditions since the plastic collapse design philosophy is adopted. The problem is treated as a simultaneous analysis-design optimization problem. Dealing primarily with planar trusses and frames, each redundant in each load condition is considered an independent variable. The equilibrium equations are used to determine all other member forces given a set of values for the redundant forces. The member section properties are computed by requiring that the yield stress is not exceeded in any member in any load condition. The key idea is using the redundants as the design-variables. The essentials of the approach can be summarized for the case of a general truss structure as follows:

Let A_i denote the cross-sectional area of the ith member,
F_{ij} the force in the ith member under the jth load condition,
R_{kj} the value of the kth redundant force under the jth load condition.

Given the yield stresses σ_i^+ and σ_i^-, the geometric configuration and the load conditions, find the R_{kj} such that

$$\sigma_i^- A_i \leq F_{ij} \leq \sigma_i^+ A_i \tag{4-1}$$

$$W(R_{kj}) = \sum_{i=1}^{I} \rho_i L_i A_i \rightarrow \text{Min} \tag{4-2}$$

where

$$A_i = \operatorname*{Max}_{j=1}^{J} \frac{|F_{ij}|}{C_i} \tag{4-3}$$

$$F_{ij}(R_{kj}) = \sum_{k=1}^{K} \alpha_{ik} R_{kj} + \beta_{ij} \tag{4-4}$$

and

$$C_i = \begin{cases} \sigma_i^+ & ; \text{ if } F_{ij} \geq 0 \\ |\sigma_i^-| & ; \text{ if } F_{ij} < 0 \end{cases} \quad . \tag{4-5}$$

A method of random steps is employed to seek the unconstrained minimum of $W(R_{kj})$ using only function evaluations [no gradients of $W(R_{kj})$ are calculated]. The fascinating aspect of this approach is that it simultaneously seeks an optimum design and the critical collapse mechanisms for each load condition. It should be noted that the problem dealt with in [4.7] can be alternatively cast as a linear programming problem in an extended space spanning the A_i's and the R_{kj}'s.

In [4.8] published in 1959, Livesley working within the plastic design philosophy studied the minimum weight design of planar frames and emphasized the importance of considering multiple loading conditions, postulating that a structure should be designed so that its behavior will be satisfactory for any condition within a prescribed loading envelope. Let every load condition within such an envelope be represented by a vector $\vec{P}$ that is a linear combination of several component loading systems $\vec{P}_i$, i.e.

$$\vec{P} = \sum_i \alpha_i \vec{P}_i \tag{4-6}$$

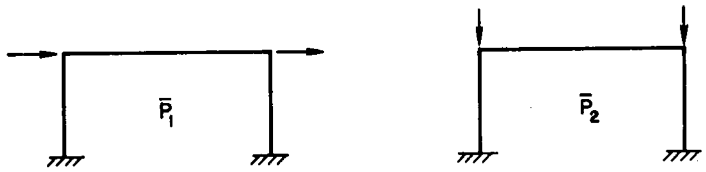
where the loading envelope is specified by defining a region R in $\vec{\alpha}$ space. A typical component of $\vec{\alpha}$ is denoted by α_i and each point $\vec{\alpha}$ in the region R defines a possible load condition. The envelope idea is illustrated by a simple example in Fig.4.1. Using a finite number of distinct loading conditions, an approximation of the loading envelope can be obtained by considering a set of points on the boundary of the region R. For example, one may elect to consider a set of J distinct load conditions defined by distinct points in the $\vec{\alpha}$ loading space, that is

$$\vec{P}_j = \sum_i \alpha_{ij} \vec{P}_i \quad ; \quad j = 1,2,\ldots J \quad . \tag{4-7}$$

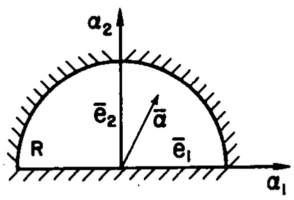
The notion of approximating a loading envelope with distinct load conditions is illustrated in Fig.4.2.

Ref. [4.9], published in 1960, showed that working within the elastic design philosophy the minimum weight design of elastic statically indeterminate structures could be cast as a non-linear programming problem in design variable space. The formulation set forth there considered a multiplicity of distinct load conditions and a variety of inequalities, including stress, displacement and side constraints. It was pointed out that the minimum weight design for a statically indeterminate structure is not necessarily one in which each member is fully stressed in at least one load condition. Since the design optimization problem formulated had the form of a non-linear programming problem, it followed that the optimum design did not necessarily lie at a vertex in the design space. The algorithm used to generate solutions for several simple three bar truss examples was a rather primitive version of a feasible direction method, that was called the method of alternate steps.

Ref. [4.10], published in 1963, reported an automated minimum weight optimum design capability for rectangular simply supported waffle plates (see Fig.4.3 in which the 7 design variables are identified) subject to a multiplicity of load conditions each of which was specified by giving the inplane force resultants N_x, N_y and N_{xy}. The failure mode concept was broadened and elastic instability as well as combined stress yield constraints were included in addition to uniaxial yield stress limits and side constraints. The influence of the total depth (H) available and the material selected, on the optimum design concept was illustrated by the numerical examples reported in that paper. As the total depth available was increased the optimum design shifted from a thick sheet, to a thin sheet with heavy stiffeners, to a thin sheet with light stiffeners and finally, if enough depth was available, the full depth was not used, suggesting the need for flanged stiffeners. The results reported exhibited relative minima in the design space and it was possible to associate the various major pockets with distinct subconcepts embedded within the statement of the mathematical programming problem. It was also found that the minimum weight design was often not unique. In particular, many designs all having the same minimum weight with different values of b_x, t_{w_y}, b_y, t_{w_x} but invariant ratios b_x/t_{w_y} and b_y/t_{w_x} were found. It was also noted that the payoff for permitting unsymmetric designs tends to decrease when there are many load conditions.



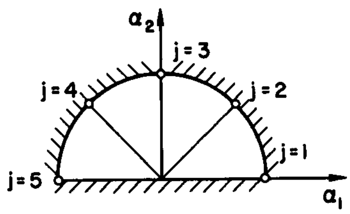
$$\bar{P} = \alpha_1 \bar{P}_1 + \alpha_2 \bar{P}_2$$
$$\bar{\alpha} = \alpha_1 \bar{e}_1 + \alpha_2 \bar{e}_2$$



REGION R DEFINED BY

$$\alpha_1^2 + \alpha_2^2 \le 1$$
$$\alpha_2 \ge 0$$

Fig.4.1 Loading Envelope Concept



$$\bar{P}_j = \sum_i \alpha_{ij} \bar{P}_i \qquad j = 1, 2, \ldots, 5$$

TABLE OF α_{ij}

i \ j	1	2	3	4	5
1	1	$+\frac{\sqrt{2}}{2}$	0	$-\frac{\sqrt{2}}{2}$	-1
2	0	$+\frac{\sqrt{2}}{2}$	1	$+\frac{\sqrt{2}}{2}$	0

Fig.4.2 Approximation of Load Envelope

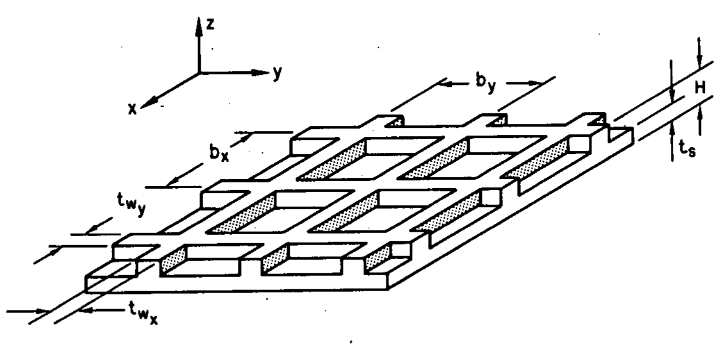


Fig.4.3 Waffle Plate

In [4.11], published in 1964, Moses introduced the idea of treating the structural design optimization problem as a sequence of linear programs. The integrated analysis-design optimization problem was stated in an extended space where the vector of unknowns $\vec{X}$ represents a concatenation of design and analysis variables. The inequality constraints were drastically simplified and non-linearity was confined to the analysis equations. A simple planar truss and a planar frame example were used to illustrate the method employed. The principal disadvantage of this formulation is that the dimensionality of the vector $\vec{X}$ grows rapidly, particularly for problems involving a large number of analysis variables and load conditions. It should be noted in passing that the integrated analysis-design optimization approach has also been explored using a penalty function formulation to transform the problem into a sequence of unconstrained minimizations [4.12], [4.13].

In [4.14], published in 1966, Reinschmidt, Cornell and Brotchie applied the sequence of linear programs formulation to the structural design optimization problem stated as an inequality constrained minimization problem in design variable space (see Section 2.6.1). A substantial number of planar truss and frame examples were studied and the need for convergence aids was revealed. Several techniques for coping with difficulties encountered in applying the SLP formulation were suggested in [4.14] and are discussed in Chapter 5. It should be noted that Pope [4.15] and Romstad and Wang [4.16] have also made contributions recently relevent to the minimum weight design of structures having prescribed geometric configuration using the SLP approach.

In [4.17], published in 1966 by Brown and Ang, the inequality constrained minimum weight structural design problem was dealt with directly in design variable space employing a modified gradient projection method (see Chapter 7). The capability reported treats planar trusses and frames and includes stress and displacement limits based on the American Institute of Steel Construction (AISC) Code. Multiple service load conditions are considered. Area and moment of inertia design variables are treated as continuous design variables and then a special program is used to transform the continuous solution into an optimum available section solution. The main computer program [4.18] is modular and hence applicable to other problems where user generated auxiliary programs compute the objective function $M(\vec{D})$, the constraint functions $h_j(\vec{D})$ and the gradients of the critical constraint functions $\nabla h_j(\vec{D})$; $j \in J_c$.

Dorn, Gomory and Greenberg [4.19], Hemp [4.20] and Fleron [4.21], all published studies in 1964, on the minimum weight design of planar trusses including both member location and sizing within the plastic design philosophy. Variation of topology was achieved by optimizing over a large preselected set of admissible members. The formulations of [4.19] and [4.20] lead to large linear programs. It should be noted that these studies were limited to structures that were statically determinate externally and subject to a single load condition. Minimum weight planar truss configurations were found to be statically determinate under a single load condition. It was shown by Dorn, Gomory and Greenberg through an interpretation of the dual LP problem that the minimization of weight is equivalent to the maximization of work done by the external loads on the joint displacements. Problems of this type were also discussed in the preceding Chapter. It should be noted that Felton and Dobbs [4.22] have recently examined the problem of truss member location and sizing considering multiple load conditions. An elastic design philosophy is adopted and a direct stiffness method of analysis is employed; both stress and member buckling constraints are considered.

In [4.23], published in 1966, Goble and DeSantis reported on an optimum design capability for continuous composite welded girders using mixed steels. The objective function to be minimized is a cost function including both material and fabrication costs. The design variables include cross sectional dimensions of discrete segments along the girder and steel type based on yield strength, as well as the location and number of splice points. The formulation considers moving loads and the constraints are based on the American Association of State Highway Officials (AASHO) Code. Optimum designs are sought employing heuristic decomposition in conjunction with a dynamic programming technique. This work is viewed as a pioneering effort in that it tackles cost as an objective function, discrete variables and moving load conditions. Cost has also been used successfully as an objective function by Moe and his coworkers [4.24], [4.25] in the context of ship structures.

The minimum weight design of stiffened cylindrical shells represents a recurring problem of fundamental importance in aerospace applications. The application of mathematical programming methods to this problem was first studied by Kicher [4.26]. A capability for the automated minimum weight design of stiffened cylindrical shells representative of the state-of-the-art (circa 1968) was reported in [4.27]. The problem is formulated using the Fiacco-McCormick interior penalty function (see Section 2.6.2 and Chapter 6) and numerical results are obtained by executing a sequence of unconstrained minimizations using the variable metric algorithm described in Chapter 6. The constraint repulsion characteristic of this formulation made it possible to employ approximate buckling analyses during major portions of the optimization. This work is discussed more fully in Chapter 9. The SUMT formulation has also been applied to the minimum weight design of stiffened fiber composite cylinders by Chao [4.28]. In this study fiber volume fraction and ply orientations are added to the collection of design variables.

In [4.29], published in 1968, Thornton and Schmit reported on an application of mathematical programming to the automated minimum weight design of a thermo-structural panel. Both time and distance through the thickness of the various layers were treated parametrically. This work which is described in Chapter 9 is thought to have been the first structural design application of the integrated penalty function formulation outlined by Eq. (2-10) and (2-11) in Section 2.6.2.

Gellatly reported in [4.30] on the development of a large scale automated minimum weight optimum design capability based on a displacement method finite element analysis and a feasible directions search procedure. This contribution is discussed in Chapter 8.

Melosh and Luik pointed out in [4.31] that the structural analysis problem associated with design optimization has the special characteristic of requiring the analysis of a large number of structures of similar form. Attention is focused on methods for the efficient analysis of a family of similar structures (multiple configuration analysis) used in conjunction with a univariate allocation scheme.

It is shown that the analysis scheme employed provides an efficient method for obtaining excellent approximations to the stress and displacement behavior as the design is modified. The method is applied to the minimum weight design of indeterminate space trusses considering stress constraints under multiple load conditions as described further in Chapter 8. The design variables are cross sectional areas and selection from an available set of discrete values introduces no special difficulties. This capability points up the importance of considering the relationships between structural analysis methods and design optimization techniques.

Karnes and Tocher [4.32] reported on a large scale automated minimum weight structural design capability, for stressed skin structure using a feasible direction method. Their work is described in Chapter 8.

Having used the framework presented in Chapter 2 to construct the summary review contained in Table 4.1, it may be observed that advances in the application of mathematical programming techniques to structural design optimization have usually exhibited one or more of the following characteristics:

(1) broadening of the design philosophy by considering a wider range of load conditions and failure modes,

(2) extending the approach to more appropriate and often more complex objective functions,

(3) consideration of a widening class of design variables from both a mathematical and a physical viewpoint,

(4) application of more sophisticated mathematical programming techniques including formulative and algorithmic innovations often based on engineering insight and physical understanding of the structural system,

(5) applications to large systems or to special problems with unusually complex loading environments and failure mode analyses.

4.3 Future Trends

The application of mathematical programming techniques to structural design is still a relatively new and growing area of interest and activity. In this Section, some current trends are identified and a few speculations concerning future research directions are offered. In Section 4.3.1, a brief review of some applications of mathematical programming to structural systems subject to dynamic response constraints is given. This subject and in particular the optimum design of structures subject to aeroelastic behavior constraints is treated further in Chapter 11. In Section 4.3.2, a brief survey of applications of mathematical programming to probability based structural design optimization problems is given. This topic is discussed in greater detail in Chapter 10. A few miscellaneous speculations about future trends, including the anticipated importance of various levels of approximate analysis, are discussed in Section 4.3.3.

4.3.1 Dynamic Response Regime

An area of investigation that has recently started to receive considerable attention is structural optimization in the dynamic response regime. The need for considering dynamic response in structural optimization is particularly pressing in lightweight flexible structures such as those that find application in aeronautical engineering. It is to be emphasized, however, that consideration of failure modes that require dynamic analysis should be in addition to appropriate static stress, displacement, and buckling limitations. In the recent literature, several structural optimization investigations have been reported that deal with one particularly troublesome behavior constraint. For example, in [4.33] and [4.34] attention has been focused on the flutter constraint while in [4.35] and [4.36], effort was centered on the natural frequency requirement. A highly idealized double wedge wing example that considered a plausible mix of constraints was reported in [4.37]; these included limitations on flutter, static stress, displacement, and angle of attack. Fox and Kapoor [4.38] have reported a capability for minimum weight optimum design of planar truss-frame structures with distributed and concentrated mass. Inequality constraints are placed on the maximum dynamic displacements and stresses, and the natural frequencies of the structure are excluded from certain bands. The limited class of structures notwithstanding (tubular members, planar truss-frames), this work represents one of the most comprehensive structural optimization investigations carried out to date in the dynamic response regime.

4.3.2 Probability Based Optimization

A steady improvement in our tools for achieving optimum designs may have a substantial influence upon design philosophy. In particular, our ability to generate designs that press right up against the limits of current specifications may lead to structures with a lower probability of survival than those usually designed against the same specifications using conventional design procedures. Thus, as optimum designs are achieved more frequently, it may become necessary to re-examine existing structural design specifications. Recognition of the philosophical attractiveness of seeking to design directly against a limited probability of failure can be expected to grow, in spite of the formidable difficulties inherent in implementing the probability based approach.

During the last decade, the foundations of structural design within a reliability philosophy have been set forth. The design problems studied to date are primarily illustrative and they indicate some of the problems that can be expected in both analysis of failure probabilities and design based on an allowable probability of failure. By and large, these studies have assumed that the environment can be replaced by a discrete set of load conditions; however, the loading magnitude and the strengths of the structural elements have been treated as random variables with a specified statistical description. Using mathematical programming methods, it has been possible to proportion member sizes of simple trusses and frames for minimum weight subject to a constraint on the overall probability of failure. One of the first papers to report on structural optimization with reliability constraints was presented by Hilton and Feigen [4.39]. Considering a single load condition, they used a Lagrange multiplier formulation to minimize weight subject to a probability of failure constraint, based on the assumption that the contributions of individual member failure probabilities to the overall probability of failure are independent. Significant weight savings compared with that obtained using a design rule based on an equal failure probability in each member resulted because lower failure probabilities were allocated to lighter members than the heavier members. Kalaba [4.40] showed that a dynamic programming formulation would give the optimum member proportions more efficiently than the Lagrange multipler technique. A necessary condition for the dynamic programming method to be applicable is that the contributions of the member failure probabilities to the overall probability of failure are independent. Switsky [4.41], followed Hilton and Feigen's Lagrange multiplier formulation and showed that several additional but reasonable assumptions lead to a simple scheme for proportioning members so as to achieve minimum weight and specific overall failure probability. In particular, Switsky showed that at the optimum, the weight of member i divided by the total weight equals the probability of failure of the ith member divided by the overall allowable probability of failure.

Moses and Kinser [4.42] report the minimum weight optimum design of multi-element statically indeterminate structures subject to multiple load conditions and an allowable overall probability of failure. By considering system interaction in the failure probability analysis, it was shown that significant weight reductions could be achieved particularly for systems with large numbers of members and failure modes. The probability of failure analysis computes the statistical correlation between failure modes and an ordering method was developed to find probabilities of failure of a mode conditional upon survival in the other modes. The probability of failure analysis presented is applicable to any elastically designed structure and it can treat any frequency distribution for each loading and element strength. Minimum weight results exhibit the characteristic that heavy members appear to have lower safety factors than light members when the structure is viewed deterministically. Recently, Shinozuka and Yang [4.43] extended the model of Kinser and Moses to an aerospace application in which proof-loading could be used. Minimum cost, including costs of members, of failure and of proof-loading became the objective function. Moses and Stevenson [4.44] report the reliability based minimum weight design of planar frames based on plastic collapse analysis. The method presented is, however, applicable to any redundant structure for which the collapse mode equations can be written as a combination of load and strength random variables. The feasible direction method of Zoutendijk (see Chapter 7) was introduced as an efficient method for reliability based optimization in which weight was the objective function and overall probability of failure the only behavior constraint. In frames it was found that traditional safety factors were a poor guide in indicating failure probabilities, particularly near a minimum weight design. It should be noted that any frequency distribution for independent load and strength variables can be handled by the method employed.

Chapter 10 contains a rather comprehensive review of approaches to structural reliability and optimization. It would appear that reliability based optimum design facilitates solution of the mathematical optimization problem by replacing the numerous behavior limitations of deterministic design by a single constraint on overall probability of failure. However, the conservation of difficulty principle applies since the mathematical and computational complexities have been transferred from the design optimization aspect to the analysis of the probability of failure.

4.3.3 Projections and Speculations

In this Section, some unsolved problems are identified and the importance of considering various levels of approximation in structural analysis is discussed.

4.3.3.1 Relative Minima

The existence of relative minima in many structural design optimization problems represents a basic difficulty. There is evidence, see for example [4.10] and [4.27], which suggests that relative minima are often associated with subconcepts present within the problem statement. The selection of initial trial designs, side constraints, design variable linking options, and the option to preassign any subset of the design variables can all be used to guide automated optimum design capabilities into various anticipated subconcept regions. In this connection, the complementary relationship between automated structural design and computer aided design employing man machine interactions should be emphasized. The relative minima problem must be recognized as one of the longstanding fundamental problems of design optimization and the view that it is in some sense a mathematical manifestation of the design creativity problem merits continuing re-examination.

4.3.3.2 Integer Variables

The problems associated with integer and strictly discrete variables are important and difficult. Techniques for dealing with mathematical programming problems with integer or mixed integer and continuous variables should be studied within the context of structural design applications. The idea of using 0-1 integer variables to declare the absence or presence of members in a structural system should be studied further. Structural optimization of rectangular multistory steel frames with respect to 0-1 topological variables and geometric layout has been studied by Soosaar and Cornell [4.45]. Toakley [4.46] has investigated the application of discrete programming techniques to the optimum design of planar frames and trusses using available sections. Porter Goff [4.47] has reported on the use of dynamic programming to obtain minimum weight layouts for cantilever trusses.

4.3.3.3 Parametric Constraints

The common occurrence of parametric inequality constraints (see Eq. (2-10) and (2-11)) in structural design problems suggests that further attention should be given to finding efficient schemes for dealing with such constraints. Parametric constraints can arise in a variety of ways. For example, the transverse displacement of a plate $w(x, y, t)$ may be limited over some time period of interest and over a specified two-dimensional region. Moving load conditions and loading envelopes represent other sources that can generate parametric constraints.

4.3.3.4 Decomposition

The study of formalized schemes for the decomposition of structural design-analysis problems into manageable subproblems which can be linked together and treated iteratively, warrants attention. The conventional separation of structural analysis and design procedures may be viewed as a traditionally accepted decomposition scheme. Note that substructuring concepts may be viewed as a form of decomposition in structural analysis. The separate consideration in aeronautical engineering of structure, weight and balance, aerodynamics, power plant, etc., while iterating through the overall systems design problem, may be thought of as an intuitive decomposition scheme. Formalizing the analysis and design decomposition of large structural systems poses a formidable challenge.

4.3.3.5 Approximate Methods of Analysis

The use of various levels of approximation as well as iterative solution methods are time honored practices in structural analysis and design. It is thus only reasonable to expect that these ideas have a place in the application of mathematical programming methods in this field. The multiple configuration analysis employed by Melosh and Luik [4.31], is an example of an iterative method in which approximations of the structural behavior are used to guide the optimization procedure. Iterative methods of this type together with those based on energy search methods [4.48] make it possible to guide a design optimization procedure using analysis information that is subject to gradual refinement as the design evolves. It should be noted also in this connection that Fox and Kapoor [4.49], have reported on an iterative method for finding eigenvalues and eigenvectors based upon minimization of the Rayleigh quotient. This method appears to be particularly well suited to dealing with the problem of normal mode analysis that is central to structural optimization in the dynamic response regime.

In [4.27], approximate shell buckling analyses were used during major portions of the structural synthesis procedure. In this instance, the shell buckling analyses were approximate in the sense that only a small number of possible buckling mode shapes were examined. It is emphasized that the constraint repulsion characteristic of the interior penalty function formulations (such as the Fiacco-McCormick method, see Section 2.6.2 and Chapter 6) often make it possible to use approximate analyses during major portions of the optimization process while still generating a sequence of steadily improving designs each of which is acceptable (even with respect to more refined analyses).

Exploration of the potential benefits to be gained from using iterative methods of analysis and various levels of analysis approximation in structural synthesis has just begun. Numerous opportunities exist for exploiting the idea of using approximate analyses during major portions of a structural optimization procedure. For example, consider the problem of limiting the maximum transverse displacement of a plate when the location at which the maximum occurs is not known. A coarse mesh of locations could be used for the approximate analysis while a fine mesh could be used to locate the maximum deflection more precisely at the end of each unconstrained minimization stage.

Useful approximations of structural behavior can often be obtained using Taylor series expansions of the analysis variables $(\vec{Y})$ as functions of the design variables $(\vec{D})$. Assume that a static linear structural analysis of the form

$$A\,\vec{Y} = \vec{B} \tag{4-8}$$

governs the behavior of a structural system under investigation. For example, in the case of a linear static displacement method of structural analysis, A would become the system stiffness matrix (K), $\vec{Y}$ would become the vector of independent generalized displacements $(\vec{u})$ and $\vec{B}$ would become the load vector for a particular load condition $(\vec{P})$. Given the results of an analysis for a design $\vec{D}_q$, i.e. $\vec{Y}(\vec{D}_q)$ and a first order sensitivity analysis*,

$$\frac{\partial \vec{Y}}{\partial D_i}(\vec{D}_q) \quad ; \quad i = 1,2,\ldots I$$

for a design $\vec{D}_q$, a first order Taylor series expansion for each analysis variable Y_k can be written as follows:

$$Y_k(\vec{D}) = Y_k(\vec{D}_q) + (\vec{D} - \vec{D}_q)^T \,\nabla Y_k(\vec{D}_q) + \ldots \tag{4-9}$$

where it is understood that the elements of the vector $\nabla Y_k(\vec{D}_q)$ are

$$\frac{\partial Y_k}{\partial D_i}(\vec{D}_q) \quad ,$$

*This refers to the sensitivity of the analysis variables to changes in the design variables as distinguished from the sensitivity of the optimum design to changes in the limitations imposed by the inequality constraints.

i.e. the partial derivatives of the kth analysis variable with respect to the ith design variable evaluated at $\vec{D}_q$. If a second order sensitivity analysis is available, then a second order Taylor series approximation can be formed by adding the following term to the right hand side of Eq. (3-9)

$$+ \tfrac{1}{2}(\vec{D} - \vec{D}_q)^T \left[\frac{\partial^2 Y_k}{\partial D_i \, \partial D_j} (\vec{D}_q) \right] (\vec{D} - \vec{D}_q) \quad . \tag{4-10}$$

It is interesting to note that if A depends on the D_i linearly and B is independent of the D_i, then it can be shown that

$$\frac{\partial^2 Y_k}{\partial D_i \, \partial D_j} = - K^{-1} \left[\frac{\partial K}{\partial D_i} \frac{\partial \vec{Y}}{\partial D_j} + \frac{\partial K}{\partial D_j} \frac{\partial \vec{Y}}{\partial D_i} \right] \quad . \tag{4-11}$$

Thus it is seen that first or second order Taylor series expansions can be used to generate approximations of the analysis variables Y_k that are useful over some region of the design space in the neighborhood of $\vec{D}_q$.

Another powerful collection of approximate analysis methods is based upon the idea of using a limited basis to represent the solution vector of a set of simultaneous equations or an eigenproblem. It has often been observed that the number of degrees of freedom required to adequately represent the behavior of a structure is frequently far less than that dictated by its geometry and the idealization techniques available. Thus, in dynamic analysis, it is common practice to express the displacement behavior in terms of a reduced set of generalized coordinates and normal modes. It is interesting to note that Turner [4.33] works with a fixed set of normal modes to seek a first approximation to the optimum design. When the first stage of the optimization is completed, a new set of normal modes (for the first approximation optimum design) is calculated and used to obtain a second approximation of the optimum design.

The idea of expressing the approximate solution of the analysis as a linear combination of a few vectors containing information about the behavior of the structural system can be used in a variety of ways. For example*, the analysis variables $\vec{Y}$ for the design $\vec{D}^{q+1}$ can be approximated by the linear combination of

(a) the analysis variables for the initial trial design $\vec{Y}(D^{(1)})$

(b) the analysis variables for the current or qth trial design $\vec{Y}(\vec{D}^{(q)})$

and (c) the directional derivative of the analysis vector along the design modification vector $\vec{S}(q)$

$$\vec{Y}(\vec{D}^{q+1}) = \beta_1 \, \vec{Y}(\vec{D}^{(1)}) + \beta_2 \, \vec{Y}(\vec{D}^{(q)}) + \beta_3 \, \vec{S}^q \cdot \nabla\vec{Y}(\vec{D}^q) \tag{4-12}$$

where the β's are undetermined coefficients. Another variation of the limited basis idea that has been explored by Fox and Muira [4.50], is to approximate the analysis vector as a linear combination of the results from r previously analyzed designs, that is let

$$\vec{Y}(\vec{D}) = \sum_{k=1}^{r} \beta_k \, \vec{Y}(\vec{D}^{(k)}) \quad . \tag{4-13}$$

Substituting either Eq. (4-12) or Eq. (4-13) into the appropriate energy statement, the stationary condition will yield a set of simultaneous equations to be solved for the β's.

4.3.4 Concluding Remarks

Current trends in the application of mathematical programming methods to structural design optimization seem to be characterized by: (a) efforts to generate large scale structural capabilities involving drastic idealization and consideration of a limited class of failure modes (see Chapter 8); (b) efforts to generate structural optimization capabilities for relatively small special problems considering complex failure mode analyses involving less idealization (see Chapter 9) and (c) applications in preliminary design of vehicle configuration (see Chapter 12). In dealing adequately with a small subsystem type problem, the engineer runs the risk of dealing adequately with the wrong problem. On the other hand, in seeking to deal with the large system, it is inevitable that idealizations and simplifications will be found necessary and, therefore, the engineer runs the risk of treating an inadequate representation of the right problem.

---

*This suggestion can be viewed as a generalization of the approach taken by Melosh and Luik [4.31].

Table 4.1

SUMMARY OF SELECTED REFERENCES (Deterministic)

	Ref. [4.6] Klein 1955	Ref. [4.7] Pearson 1958	Ref. [4.8] Livesley 1959	Ref. [4.9] Schmit 1960	Ref. [4.10] Schmit, Kicher Morrow 1963	Ref. [4.11] Moses 1964
Kinds of Failure Modes	σ, u	σ_{yld} Plastic Collapse	σ_{yld} Plastic Collapse	σ, u	Plate Buckling Combined σ	σ
Kind of Load Conditions	Service Single	Overload Multiple	Overload Multiple	Service Multiple	Service Multiple	Service Multiple
Kind of Design Variables	Continuous Sizing	Continuous Sizing	Continuous Sizing	Continuous Sizing	Continuous Sizing Config.	Continuous Sizing
Objective Function	Weight Non-linear	Weight Linear	Weight Linear	Weight Linear	Weight Non-linear	Weight Linear
Formulation and Algorithm	Classical ---	---- Random Steps	Linear Program	NLP Alternate Step	NLP Alternate Step	SLP (extended space) Simplex
Type of Structure	Beam	Planar Trusses + Frames	Planar Frame	Simple Planar Truss	Waffle Plate	Planar Truss and Frame

	Ref. [4.14] Reinschmidt Cornell, Brotchie 1966	Ref. [4.17] Brown and Ang 1966	Ref. [4.19] Dorn, Gomory Greenberg 1964	Ref. [4.23] Goble and DeSantis 1966	Ref. [4.27] Morrow, Schmit 1968	Ref. [4.29] Thornton and Schmit 1968
Kinds of Failure Modes	σ	σ, u AISC	σ_{yld} Plastic Collapse	σ AASHO	Shell Buckling Combined σ	Temp., ε Combined σ
Kind of Load Conditions	Service Multiple	Service Multiple	Overload Single	Service Moving	Service N, p, T Multiple	Service Parametric (Re-entry)
Kind of Design Variables	Continuous Sizing	Continuous Sizing	Continuous Sizing and Location	Discrete Sizing Config. Material	Continuous Sizing, Config.	Continuous Sizing Config.
Objective Function	Weight Linear	Weight Non-linear	Weight Linear	Cost Non-linear	Weight Non-linear	Weight or Depth Non-linear or Linear
Formulation and Algorithm	SLP Simplex	NLP Grad. Proj.	Linear Program	Heuristic Decomposition Dynam. Prog.	SUMT Fletcher-Powell	SUMT Fletcher-Powell
Type of Structure	Planar Trusses Frames	Planar Trusses Frames	Planar Trusses	Continuous Welded Girders	Integrally Stiffened Cylindrical Shell	Thermo-Structural Panel

Table 4.1

SUMMARY OF SELECTED REFERENCES (Deterministic) (Contd)

	Ref. [4.30] Gellatly 1966	Ref. [4.31] Melosh and Luik 1967	Ref. [4.32] Tocher and Karnes 1968	Ref. [4.38] Fox and Kapoor 1969
Kinds of Failure Modes	σ, u	σ	σ, u	σ, u Dynamic
Kind of Load Conditions	Service Multiple	Service Multiple	Service Multiple	Service Single
Kind of Design Variables	Continuous Sizing	Discrete Sizing	Continuous Sizing	Continuous Sizing
Objective Function	Weight Linear	Weight Linear	Weight Linear	Weight Linear
Formulation and Algorithm	NLP Alternate Step	NLP Univariate Search	NLP Feasible Direction Zoutendijk	NLP Feasible Direction Zoutendijk
Type of Structure	Bars, Shear Panels, Membrane Plates	Planar and Space Trusses	Bars and Triangular Membranes	Tubular Planar Truss-Frames

Table 4.2

SUMMARY OF SELECTED REFERENCES (Probability Based)

	Ref. [4.39] Hilton and Feigen 1960	Ref. [4.42] Moses and Kinser 1967	Ref. [4.44] Moses and Stevenson 1968	Ref. [4.43] Shinozuka and Yang 1969
Kinds of Failure Modes	σ	σ	σ_{yld} Plastic Collapse	σ
Kind of Load Conditions	Service Single	Service Multiple	Service Multiple	Service and Proof Loading Multiple
Kind of Design Variables	Continuous Sizing	Continuous Sizing	Continuous Sizing	Continuous Sizing
Objective Function	Weight Linear	Weight Linear	Weight Non-Linear	Cost Non-linear
Formulation and Algorithm	Classical ----	NLP Alternate Step	NLP Feasible Direction Zoutendijk	NLP Feasible Direction Zoutendijk
Type of Structure	2 Member Structure	Indeterminate Trusses	Planar Frames	Determinate Trusses

List of References

Ref.

4.1 Gerard, G., "Optimum Structural Design Concepts for Aerospace Vehicles: Bibliography and Assessment", USAF, AFFDL-TR-65-9, June 1965

4.2 Gerard, G., "Optimum Structural Design Concepts for Aerospace Vehicles: Bibliography and Assessment", USAF, AFFDL-TR-66-188, December 1966

4.3 Wasiutynski, Z. and Brandt, A., "The Present State of Knowledge in the Field of Optimum Design of Structures", *Applied Mechanics Reviews,* May 1963, pp.341-350

4.4 Sheu, C. Y. and Prager, W., "Recent Developments in Optimal Structural Design", *Applied Mechanics Reviews,* Vol.21, No.10, October 1968, pp.985-992

4.5 Kowalik, J., "Non-linear Programming Procedures and Design Optimization", Acta Polytechnica Scandinavica, Mathematics and Computer Machinery Series NR. 13, Trondheim, Norway, 1966

4.6 Klein, B., "Direct Use of Extremal Principles in Solving Certain Optimization Problems Involving Inequalities", *Operations Research,* Vol.3, 1955, pp.168-175

4.7 Pearson, C. E., "Structural Design by High Speed Computing Machines", Proceedings of the First Conference on Electronic Computation, ASCE, New York, 1958, pp.417-436

4.8 Livesley, R. K., "Optimum Design of Structural Frames for Alternative Systems of Loading", *Civil Engr. and Public Works Review,* Vol.54, No.636, June 1959, pp.737-740

4.9 Schmit, L. A., "Structural Design by Systematic Synthesis", Proc. of the Second National Conference on Electronic Computation, Structural Division, ASCE, Pittsburgh, Pa., September 1960, pp.105-132

4.10 Schmit, L. A., Kicher, T. P. and Morrow, W. M., "Structural Synthesis Capability for Integrally Stiffened Waffle Plates", *AIAA Journal,* Vol.1, No.12, December 1963, pp.2820-2836

4.11 Moses, F., "Optimum Structural Design using Linear Programming", *J. of the Structural Division, ASCE,* Vol.90, No.ST6, December 1964, pp.89-104

4.12 Schmit, L. A. and Fox, R. L., "An Integrated Approach to Structural Synthesis and Analysis", *AIAA Journal,* Vol.3, No.6, June 1965, pp.1104-1112

4.13 Fox, R. L. and Schmit, L. A., "Advances in the Integrated Approach to Structural Synthesis", *J. of Spacecraft and Rockets,* Vol.3, No.6, June 1966, pp.858-866

4.14 Reinschmidt, K. F., Cornell, A. C. and Brotchie, J. F., "Iterative Design and Structural Optimization", *J of the Structural Division, ASCE,* Vol.92, No.ST6, December 1966, pp.281-318

4.15 Pope, G. G., "The Design of Optimum Structures of Specified Basic Configuration", *International Journal of Mech. Sci.,* Vol.10, No.4, April 1968, pp.251-263

4.16 Romstad, K. M. and Wang, C. K., "Optimum Design of Framed Structures", *Journal of the Structural Division, ASCE,* Vol.94, No.ST12, December 1968, pp.2817-2845

4.17 Brown, D. M. and Ang, A. H. S., "Structural Optimization by Non-linear Programming", *J. of the Structural Division, ASCE,* Vol.92, No.ST6, December 1966, pp.319-340

4.18 Brown, D. M. and Ang, A. H. S., "A Non-linear Programming Approach to the Minimum Weight Elastic Design of Steel Structures", Structural Research Series No.298, Univ. of Illinois, Civil Engineering Studies, Urbana, Ill., 1965

4.19 Dorn, W. S., Gomory, R. E., and Greenberg, H. J., "Automatic Design of Optimal Structures", *Journal de Mechanique,* Vol.3, No.1, 1964, pp.25-52

4.20 Hemp, W. S., "Studies in the Theory of Michell Structures", *Proc. of the 11th Int. Cong. Appl. Mech.,* 1964, Springer, Berlin, 1966, pp.621-628

4.21 Fleron, P., "The Minimum Weight of Trusses", *Bygningingsstatiske Meddelelser,* Vol.35, No.3, 1966, pp.81-96

4.22 Dobbs, M. W. and Felton, L. P., "Optimization of Truss Geometry", *Journal of the Structural Division, ASCE,* Vol.95, No.ST10, October 1969, pp.2105-2118

4.23 Goble, G. G. and DeSantis, P. V., "Optimum Design of Mixed Steel Composite Girders, *J. of the Structural Div., ASCE,* Vol.92, No.ST6, December 1966, pp.25-43

4.24 Moe, J. and Lund, S., "Cost and Weight Minimization of Structures with Special Emphasis on Longitudinal Strength Members of Tankers", *Trans. Royal Inst. of Nav. Arch.,* Vol.110, No.1, 1968, pp.43-70

4.25 Moe, J., "Design of Ship Structures by Means of Non-linear Programming Techniques", *Proc. AGARD Symposium on Structural Optimization,* Istanbul, October 1969, AGARD-CP-36-70

List of References (Contd.)

Ref.

4.26 Kicher, T. P., "Structural Synthesis of Integrally Stiffened Cylinders", *J. of Spacecraft and Rockets,* Vol.5, No.1, January 1968, pp.62-67

4.27 Morrow II, W. M., and Schmit, L. A., "Structural Synthesis of a Stiffened Cylinder", NASA CR-1217 December 1968

4.28 Chao, D., "Minimum Weight Design of Stiffened Fiber Composite Cylinders", USAF, AFML-TR-69-251, September 1969

4.29 Thornton, W. A. and Schmit, L. A., "The Structural Synthesis of an Ablating Thermostructural Panel", NASA CR-1215, December 1968

4.30 Gellatly, R. A., "Development of Procedures for Large Scale Automated Minimum Weight Structural Design", USAF, AFFDL-TR-66-180, December 1966

4.31 Melosh, R. J. and Luik, R., "Approximate Multiple Configuration Analysis and Allocation for Least Weight Structural Design", USAF, AFFDL-TR-67-59, April 1967

4.32 Tocher, J. L. and Karnes, R. N., "Automatic Design of Optimum Hole Reinforcement," No. D6-23359, May 1968, The Boeing Company, Commercial Airplane Division, Renton, Washington

4.33 Turner, M. J., "Optimization of Structures to Satisfy Flutter Requirements", *AIAA Journal,* Vol.7, No.5, May 1969, pp.945-951

4.34 McIntosh, S. C., Weisshaar, T. A. and Ashley, H., "Progress in Aeroelastic Optimization Analytical vs. Numerical Approaches", Department of Aeronautics and Astronautics, Stanford University, SUDAAR 383, July 1969

4.35 Rubin, C. P., "Minimum Weight Design of Complex Structures Subject to a Frequency Constraint", *AIAA Journal,* Vol.8, No.5, May 1970, pp.923-927

4.36 Zarghamee, M. S., "Optimum Frequency of Structures", *AIAA Journal,* Vol.6, No.4, April 1968, pp.749-750

4.37 Schmit, L. A. and Thornton, W. A., "Synthesis of an Airfoil at Supersonic Mach Number", NASA CR-144, January 1965

4.38 Fox, R. L. and Kapoor, M. P., "Structural Optimization in the Dynamic Response Regime: A Computational Approach", AIAA Structural Dynamics and Aeroelasticity Specialist Conference, New Orleans, 1969, pp.15-22

4.39 Hilton, H. H. and Feigen, M., "Minimum Weight Analysis Based on Structural Reliability", *J. of the Aerospace Sciences,* Vol.27, No.9, September 1960, pp.641-652

4.40 Kalaba, R., "Design of Minimal-Weight Structures for Given Reliability and Cost", *J. of the Aerospace Sciences,* Vol.29, No.3, March 1962, pp.355-356

4.41 Switsky, H., "Minimum Weight Design with Structural Reliability", AIAA Fifth Annual Structures and Materials Conference, April 1964, pp.316-322

4.42 Moses, F. and Kinser, D. E., "Optimum Structural Design with Failure Probability Constraints", *AIAA Journal,* Vol.6, No.6, 1967, pp.1152-1158

4.43 Shinozuka, M. and Yang, J. N., "Optimum Structural Design based on Reliability and Proof-Load Test", *Annals of Assurance Sciences,* Proceedings of Reliability and Maintainability Conference, Vol.8, July 1969, pp.375-391

4.44 Moses, F. and Stevenson, J. D., "Reliability Based Structural Design", Case Western Reserve University, DSMSMD Report No.16, January 1968

4.45 Soosaar, K. and Cornell, A. C., "Optimization of Topology and Geometry of Structural Frames", a paper presented at the ASCE Joint Specialty Conference on Optimization and Non-linear Problems, Chicago, Illinois, April 18-20, 1968

4.46 Toakley, A. R., "Optimum Design Using Available Section", *J. of the Structural Division, ASCE,* Vol.94, No.ST5, May 1968, pp.1219-1241

4.47 Porter Goff, R. F. D., "Decision Theory and the Shape of Structures", *J. of the Royal Aeronautical Society,* Vol.70, 1966, pp.448-452

4.48 Schmit, L. A. et al., "Developments in Discrete Element Finite Deflection Structural Analysis by Function Minimization", USAF, AFFDL-TR-68-126, September 1968

4.49 Fox, R. L. and Kapoor, M. P., "A Minimization Method for the Solution of the Eigenproblem Arising in Structural Dynamics", Proceedings of the Second Conference on Matrix Methods in Structural Mechanics, WPAFB, October 1968, AFFDL-TR-68-150, December 1969, pp.271-306

4.50 Fox, R. L. and Miura, H., "An Approximate Analysis Technique for Structural Optimization", submitted to the AIAA Journal for publication, July 1970

SECTION II

ALGORITHMIC TOOLS

Chapter 5

SEQUENCE OF LINEAR PROGRAMS

by

G. G. Pope

5.1 Introduction

Linear programming problems are of importance in their own right in many commercial and technological fields and consequently their mathematical properties have been studied in depth and efficient computer programs have been developed for their solution. This available expertise can be utilised in two distinct ways in the solution of non-linear programming problems. Firstly the choice of an efficient direction in which to search for a lighter feasible solution, starting from a feasible solution in which one constraint at least is active, may be expressed as a problem in linear programming, following the procedure due to Zoutendijk which is described in Chapter 7. Secondly the non-linear programming problem may itself be replaced by a sequence of linear programming problems. The latter approach which has the attraction of simplicity but which also contains some pitfalls for the unwary, is discussed in this Chapter. First, however, a brief description is given of the more important properties of linear programming problems themselves.

5.2 Linear Programming

In order to demonstrate clearly the duality properties of linear programming problems, it is convenient in this section to depart slightly from the vectorial notation used in the preceding text and to employ instead the well-known convention in which repeated suffices are used to denote summations, i.e.

$$a_{ij}\, b_i = \sum_{i=1}^{i=I} a_{ij}\, b_i \quad .$$

The fundamental theory of linear programming is developed rigorously in the texts by Hadley [5.1] and by Dantzig [5.2]. A completely general problem of this class may be expressed in the following form. Find a vector d_i of I terms which satisfies the equations

$$f_{ij}\, d_i = a_j \quad ; \quad j = 1,2,\ldots J \quad , \tag{5-1}$$

and the inequalities

$$h_{ik}\, d_i \geqslant b_k \quad ; \quad k = 1,2,\ldots K \quad , \tag{5-2}$$

$$d_i \geqslant 0 \quad ; \quad i = 1,2,\ldots I \tag{5-3}$$

and which minimizes a merit function defined by

$$M = e_i\, d_i \quad . \tag{5-4}$$

Extra positive variables, known as slack variables, may always be added so that the inequalities (5-2) may be incorporated in Eq. (5-1); conversely the latter may be expressed as the inequalities

$$\left.\begin{aligned} f_{ij}\, d_i &\geqslant a_j \quad , \\ -f_{ij}\, d_i &\geqslant -a_j \quad . \end{aligned}\right\} \tag{5-5}$$

Thus either Eq. (5-1) or the inequalities (5-2) may be omitted from the formulation without loss in generality.

5.2.1 Terminology and Method of Solution

Consider now a typical linear programming problem which is so formulated that the inequality constraints (5-2) do not appear explicitly. A feasible solution is defined as any solution which satisfies both Eq. (5-1) and the necessary condition (5-3) that the variables are positive. A basic solution is defined as a solution consisting of J non-zero variables and $(I - J)$ zero variables. Degenerate solutions in which there are more than $(I - J)$ zero variables can be ignored in practical computations. It may readily be demonstrated that, if a feasible solution exists at all, there must necessarily be a basic feasible solution which minimizes the merit function, although there may sometimes be other feasible solutions which reduce this function to the same value.

Linear programming problems are usually solved by the Simplex method developed by Dantzig or by methods closely related to it. Applications of these methods start from a known basic feasible solution and progress successively to solutions of the same type closer to the optimum until the latter is reached. Provided degenerate solutions do not occur this procedure necessarily converges in a finite number of operations. In many applications a combination of non-zero unknowns which will yield a basic feasible solution is known initially and can be used as a starting point. When no such prior information is available the linear programming problem may be enlarged artificially to a problem with an obvious basic feasible solution, using the following technique which is due to Zoutendijk [5.3]. Eq. (5-1) are first arranged in such a way that the constants a_j are all positive. A different additional variable is then added to each of the equations that do not include a slack variable preceded by a positive sign, and the merit function is modified to become

$$M' = e_i d_i + P(x_1 + x_2 + x_3 \dots x_J) \tag{5-6}$$

where x_1 to x_J are the additional variables and the slack variables of this type, and P is a large positive constant. A basic feasible solution to this enlarged problem is obtained by selecting the variables x_1 to x_J to be non-zero. The optimum solutions to the original and enlarged problems will obviously be the same provided, of course, that the former has a basic feasible solution and that the constant P is sufficiently large.

5.2.2 Duality

Consider a typical linear programming problem expressed for convenience in the form:

minimize

where

and

$$\left.\begin{aligned} M &= e_i d_i \\ d_i &\geqslant 0 \\ h_{ik} d_i &\geqslant b_k \quad ; \quad k = 1,2,\dots K \ . \end{aligned}\right\} \tag{5-7}$$

This is closely related to another linear programming problem involving the same coefficients e_i, b_k and h_{ik} which may be expressed as follows:

maximize

where

and

$$\left.\begin{aligned} N &= b_k y_k \\ y_k &\geqslant 0 \\ h_{ik} y_k &\leqslant e_i \quad ; \quad i = 1,2,\dots I \ . \end{aligned}\right\} \tag{5-8}$$

Whichever of the above problems is of primary interest in a particular application is referred to as the primal problem and the related problem is referred to as the dual problem.

The following duality properties are useful in the present context:

(1) The optimum solution of one problem may be deduced directly from the optimum solution of the other, and the merit function in both problems has the same optimum value.

(2) Consider the optimum solution of both problems when every constraint in each problem involves a slack variable. When the slack variable in the kth constraint in one problem is non-zero, the kth variable in the other problem vanishes; conversely, if the kth variable is non-zero in one problem, the slack variable in the kth constraint in the other problem is zero.

It sometimes proves more economical from the computational viewpoint to solve the dual problem rather than the primal problem, especially when a basic feasible solution is known initially to the former but not to the latter (see, for example, Paragraph 5.3.4). More general conditions under which it is preferable to solve the dual problem vary to some extent with the algorithm used in the solution and depend on the number of equality constraints in the primal problem that do not involve slack variables and on the ratio of the number of constraints to the number of variables.

5.3 The Reduction of Non-Linear Programming Problems to a Sequence of Problems in Linear Programming

The properties of non-linear programming problems may most readily be described by considering first problems in which all the constraints are expressed as inequalities and in which only two variables are involved. Consider the following linear programming problem:

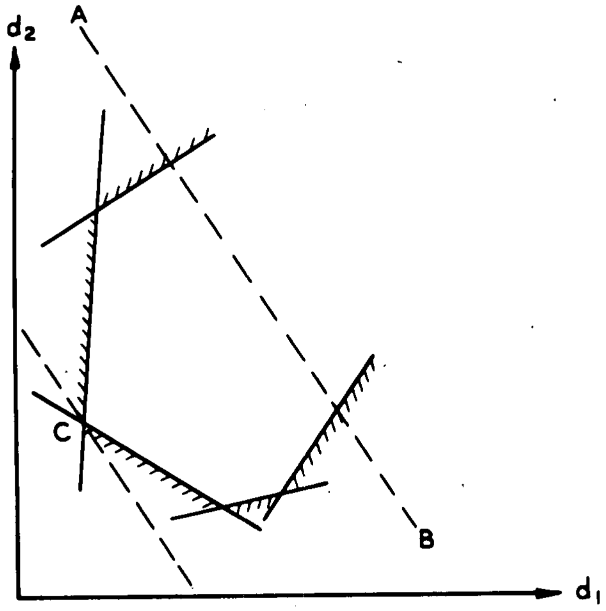


Fig.5.1 Linear Problem

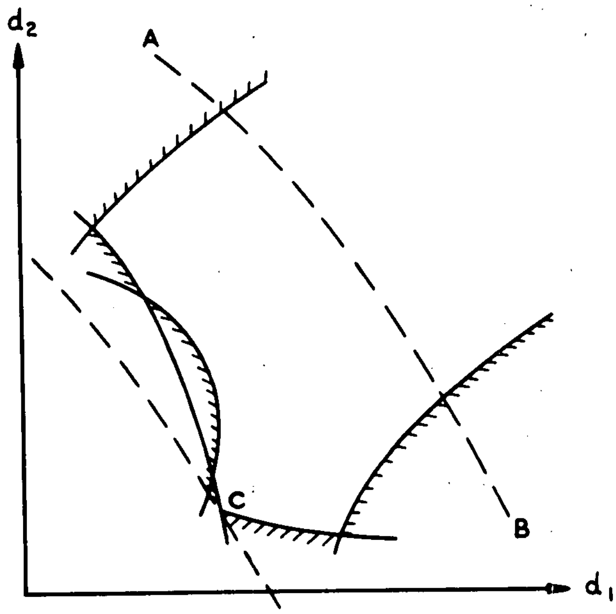


Fig.5.2 General Non-linear Problem

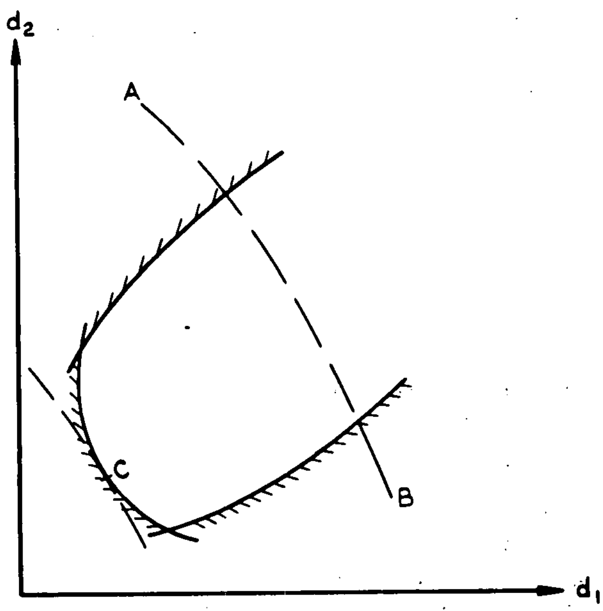


Fig.5.3 Convex Problem

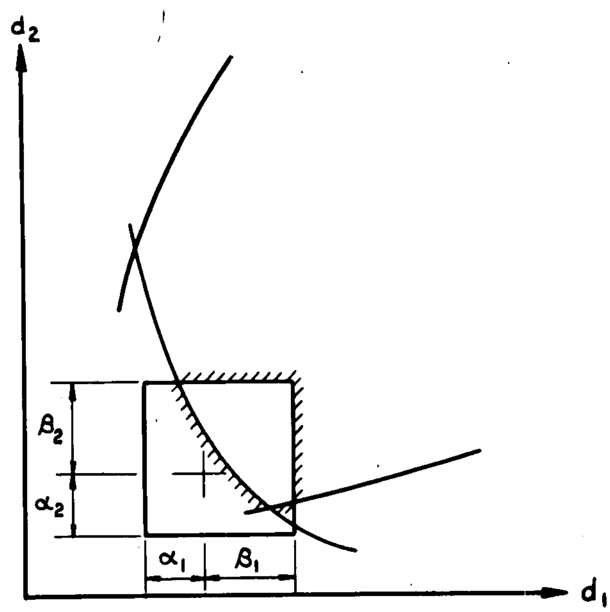


Fig.5.4 Additional Constraints in the Move Limit Method

minimize

$$M = e_1 d_1 + e_2 d_2$$

subject to the constraints

$$\left.\begin{aligned} & h_{1k} d_1 + h_{2k} d_2 \geqslant b_k \quad ; \quad k = 1,2,\ldots K \\ & d_1 \geqslant 0 \ , \ d_2 \geqslant 0 \ . \end{aligned}\right\} \qquad (5\text{-}9)$$

where

This problem is presented in graphical form in Fig.5.1. The constraints consist of a number of intersecting straight lines where the hatching indicates the edges of the region in which feasible solutions are obtained. The line AB indicates the locus of points along which the merit function M has a constant value and the corresponding loci for other values of M are lines parallel to AB; the problem of minimizing M reduces therefore to that of finding a line parallel to AB which passes through the extreme vertex C of the feasible region. Consider now the non-linear programming problem:

minimize

$$M(d_1, d_2)$$

subject to the constraints

$$\left.\begin{aligned} & h(d_1, d_2) \geqslant b_k \quad ; \quad k = 1,2,\ldots K \\ & d_1 \geqslant 0 \ , \ d_2 \geqslant 0 \ . \end{aligned}\right\} \qquad (5\text{-}10)$$

where again

In general, neither the boundaries of the feasible region nor the contours of equal values of the merit function are straight lines, and they may take such complex forms as are illustrated in Fig.5.2. It is immediately obvious therefore that the optimum solution need not necessarily be at an intersection of constraints, and also that local optima may occur in addition to the global optimum which is sought. This latter difficulty does not arise in problems where the constraints and merit function have the forms illustrated in Fig.5.3; such problems, which are usually referred to as convex problems, are difficult to identify when the number of unknowns is large. Consequently, since all deterministic solution techniques search in effect for local optima, it is strictly necessary to repeat solution procedures from several unrelated starting points before a calculated optimum can be treated with confidence as a global optimum.

If more general problems are now considered which are expressed purely in terms of inequality constraints and which involve N variables d_n, the (d_1, d_2) plane may be generalised into an N-dimensional space so that the constraint intersections on the edges of the feasible region in the plane become vertices on the boundaries of a corresponding region in the N-dimensional space. Using the notation of the preceding Chapters such problems may be expressed in the following form:

minimize

subject to the constraints

$$\left.\begin{aligned} & M(\vec{D}) \\ & h_k(\vec{D}) \geqslant 0 \quad ; \quad k = 1,2,\ldots K \end{aligned}\right\} \qquad (5\text{-}11)$$

where the column vector $\vec{D}$ corresponds to the variables $d_1 \ldots d_N$ but is not necessarily expressed in terms of components which are constrained to be positive.

5.3.1 The Simplest Approach

The following procedure, which has been employed in the structural design context by Moses [5.4] and by Karihaloo et al. [5.5] is the simplest possible for replacing a typical non-linear programming problem by a sequence of problems in linear programming:

(1) Linearise the constraints and the merit function in the neighbourhood of an arbitrary point $\vec{D}_o$ and solve the resulting linear programming problem which is given by

minimize

$$M(\vec{D}_o) + \nabla M(\vec{D}_o) \; . \; [\vec{D} - \vec{D}_o]$$

subject to the constraints (5-12)

$$h_k(\vec{D}_o) + \nabla h_k(\vec{D}_o) \; . \; [\vec{D} - \vec{D}_o] \;\geq\; 0 \quad ; \quad k = 1,2,\ldots K \quad .$$

(2) Repeat the process until the optimum solutions of successive linearised problems are virtually identical, redefining $\vec{D}_o$ each time as the optimum solution to the preceding problem.

This procedure will only converge if the optimum solution happens to be at a vertex of the feasible region in the N-dimensional space referred to above. If the curvature of the constraints or of the merit function is such that the optimum solution does not correspond to a vertex, the numerical results will oscillate indefinitely between adjacent vertices; such a situation is illustrated in Fig.5.3. This difficulty may be overcome in convex problems by the use of the procedures described in Sections 5.3.2 and 5.3.3; in more general applications the procedure described in Section 5.3.3 may be employed.

5.3.2 The Cutting Plane Method

The cutting plane method, which was developed independently by Cheney and Goldstein [5.6] and by Kelly [5.7], employs the useful property that linearised constraints in convex problems necessarily lie entirely outside the feasible region. Consequently an envelope of such constraints may be used to represent the critical non-linear constraints to any required degree of accuracy. A typical version of the method proceeds as follows when the objective function is linear:

(1) Linearise the constraints in the neighbourhood of an arbitrary point $\vec{D}_o$ and solve the resulting linear programming problem.

(2) Substitute the results of the linearised computation in the non-linear constraint equations and find which of the latter is most seriously violated.

(3) Linearise this constraint about the optimum solution $\vec{D}_p$ to the preceding linear programming problem and find the modified optimum solution when this additional linearised constraint is added.

(4) Repeat steps (2) and (3), adding an extra linearised constraint each time, until all non-linear constraints are satisfied to an acceptable standard of accuracy.

Cornell, Reinschmidt and Brotchie [5.8], [5.9] have studied the possibility of disgarding inactive constraints to reduce the size of the linear programming problems involved in the application of this method. They have found, however, that the computations required to identify the constraints that can rigorously be omitted are too lengthy in general to be of practical value. Simple semi-empirical rules suggested by these authors for the elimination of such constraints are unlikely to be suitable for general application. Difficulties of this kind are also discussed by Moses [5.10], [5.11].

The cutting plane method has two very undesirable features:

(1) When the optimum solution does not coincide with a vertex of the feasible region, the angle between the active linearised constraints is small; consequently round-off errors can sometimes debase numerical accuracy to an unacceptable extent.

(2) The method cannot be employed satisfactorily in problems which are not strictly convex since the linearised constraints may then exclude legitimate parts of the feasible region.

This second feature, in particular, makes the cutting plane method unacceptable in practical problems where convexity cannot be demonstrated.

5.3.3 The Move Limit Method

An alternative approach due to Griffiths and Stewart [5.12], which does not suffer the above deficiencies, makes use of artificial limits on the variation of the design variables in a typical linearised computation; it has been used successfully by several workers in the structural design field [5.8], [5.9], [5.13], [5.14], [5.16] and proceeds as follows:

(1) Linearise the constraints and the merit function in the neighbourhood of an arbitrary point $\vec{D}_o$ and impose additional constraints of the form

$$\vec{D}_o - \vec{\alpha} \;\leq\; \vec{D} \;\leq\; \vec{D}_o + \vec{\beta} \tag{5-13}$$

as illustrated in Fig.5.4, where $\vec{\alpha}$ and $\vec{\beta}$ are suitably chosen vectors of positive constants.

(2) Repeat the process, redefining $\vec{D}_o$ as the optimum solution to the preceding linear programming problem, until either no significant change occurs in the solution, or successive solutions start to oscillate between the vertices of the feasible region; in the latter event continue computations using suitably reduced values of $\vec{\alpha}$ and $\vec{\beta}$.

Discretion and experience must be employed in the choice of values for the components of the vectors $\vec{\alpha}$ and $\vec{\beta}$. For computational efficiency it is desirable to choose relatively large values initially so that the imposed limits do not impede rapid convergence to the immediate vicinity of the optimum solution. Insufficient evidence is yet available to indicate the best way to reduce these values when oscillation occurs; the author has, however, obtained satisfactory convergence by repeatedly halving the amplitude in structural applications in which equal values were employed for all the components of these vectors associated with the design variables. For computational efficiency it may, of course, be desirable to impose severe limits only on those variables which are immediately associated with oscillatory behaviour; this aspect has not yet been studied in depth.

The above method, which is known either as the 'move limit method' or as the 'method of approximation programming', involves a complete relinearisation of the non-linear problem before each linear sub-problem. Consequently, in structural problems where the constraints consist only of lower bounds on the design variables and upper bounds on the displacements and stresses, negligible additional effort is involved in factoring each linearised solution so that it is just a feasible solution of the relevant non-linear problem. Any increase in the value of the factored merit function after successive linearised computations may then be taken as an adequate indication that a reduction in the move limits is necessary.

It has been assumed in the foregoing discussion that each non-linear constraint has been represented by a single linear constraint in the individual sub-problems of the move limit method. Better approximations may, however, be incorporated by retaining appropriate non-linear terms in a Taylor's series expansion of the constraint about the starting point $\vec{D}_o$, and by representing this power series expansion approximately between the move limits by a series of tangent planes in the N-dimensional space referred to above. Such techniques are discussed by Cornell, Reinschmidt and Brotchie [5.8], [5.9] and by Moses [5.10], but little experience has yet been obtained in their use.

5.3.4 Use of the Dual Problem in the Structural Optimization Field

A useful property of any of the above methods when applied to structural problems is that the coefficients of the design variables in the objective function of the primal problem are nearly always all positive. Thus a basic feasible solution to the dual of this problem may be obtained directly by choosing the slack variables to be the non-zero variables [5.14]. The Simplex method may then be used to find the optimum solution of the dual problem and consequently of the primal problem as well. There is, in theory, a possibility that no feasible solution exists to the primal problem; the objective function of the dual problem can then take an indefinitely large value. Difficulties of this kind cannot, of course, occur if the linearisation process starts from a feasible solution of the non-linear problem; they are only likely in practice when upper limits are imposed on the permissible values of the design parameters in the non-linear problem, or when lower limits are placed on the absolute values of the displacements. Under these circumstances a detailed investigation may be required to show whether the difficulty is due to linearisation about an inappropriate point or whether the non-linear problem itself has no feasible solution.

5.3.5 Discrete Variables

Variables that can only take discrete values introduce major complications whatever solution procedure is employed. Such variables may in theory be incorporated in procedures based on linear programming with the aid of the integer-programming techniques developed by Gomory and by Beale, see Dantzig [5.2]. Convergence difficulties were, however, experienced by Toakley [5.15] when he employed Gomory's algorithm in the structural optimization field.

Acknowledgement - This Chapter is British Crown Copyright, reproduced with the permission of the Controller, Her Majesty's Stationery Office.

List of References

Ref.

5.1 Hadley, G., *Linear Programming,* 1st ed., Addison-Wesley, Reading, Mass., 1962

5.2 Dantzig, G. B., *Linear Programming and Extensions,* 1st ed., Princeton University Press, Princeton, New Jersey, 1963

5.3 Zoutendijk, G., 'Nonlinear Programming : A Numerical Survey,' *SIAM Journal on Control,* Vol.4, No.1, February 1966, pp.194-210

5.4 Moses, F., 'Optimum Structural Design using Linear Programming,' *J. of the Structural Division, ASCE,* Vol.90, No.ST6, December 1964, pp.89-104

5.5 Karihaloo, B. L., Pathare, P. R. and Ramash, C. K., 'The Optimum Design of Space Structures by Linear Programming using the Stiffness Matrix Method of Analysis,' *Space Structures* (edited by R. M. Davies), pp.278-290, Blackwell Scientific Publications, Oxford, England, 1967

5.6 Cheney, E. W., and Goldstein, A. A., 'Newton's Method for Convex Programming and Tchebycheff Approximation,' *Numerische Mathematik,* Vol.1, 1959, pp.253-268

5.7 Kelley, J. E., 'The Cutting Plane Method for solving Convex Programs,' J. of SIAM, 1960, pp.703-712

5.8 Cornell, C. A., Reinschmidt, K. F., and Brotchie, J. F., 'Structural Optimization', Research Report R65-26, Part 2, Dept. of Civil Engineering, MIT, Cambridge, Mass., September 1965

5.9 Reinschmidt, K. F., Cornell, C. A., and Brotchie, J. F., 'Iterative Design and Structural Optimization,' *J. of the Structural Division, ASCE,* Vol.92, No.ST6, December 1966, pp.281-318

5.10 Moses, F., 'Some Notes and Ideas on Mathematical Programming Methods for Structural Optimization,' Meddelelse SKB 11/M8, Norges Tekniske Høgskole, Trondheim, Norway, January 1967

5.11 Moses, F. and Onada, S., 'Minimum weight design of structures with application to elastic grillages,' Int. J. for Numerical Methods in Engineering, Vol.1, 1969, pp.311-331

5.12 Griffith, R. E. and Stewart, R. A., 'A Non-linear Programming Technique for the Optimization of Continuous Processing Systems,' *Management Science,* Vol.7, 1961, pp.379-392

5.13 Pope, G. G., 'The Design of Optimum Structures of Specified Basic Configuration,' *International J. of Mech. Sci.,* Vol.10, No.4, April 1968, pp.251-263

5.14 Romstad, K. M. and Wang, C. K., 'Optimum Design of Framed Structures,' *J. of the Structural Division, ASCE,* Vol.94, No.ST12, December 1968, pp.2817-2845

5.15 Toakley, A. R., 'Optimum Design using Available Section, *J. of the Structural Division, ASCE,* Vol.94, No.ST5, May 1968, pp.1219-1241

5.16 Pope, G. G., 'The Application of Linear Programming Techniques in the Design of Optimum Structures', *Proc. AGARD Symposium on Structural Optimization,* Istanbul, October 1969, AGARD-CP-36-70

Chapter 6

UNCONSTRAINED MINIMIZATION APPROACHES TO CONSTRAINED PROBLEMS

by

R. L. Fox

6.1 Introduction

There are many approaches to the constrained minimization problem. Methods which have developed a great deal of currency are the unconstrained minimization formulations of the constrained problem. The basic idea of these methods is to convert the constrained problem, with its objective function and equality and inequality constraints, into a problem in which some new function is minimized without regard for constraints. The solution to the original constrained minimization problem is then developed through a sequence of unconstrained minimizations.

There are several reasons for the appeal of the unconstrained minimization formulations and it is useful to examine some of these briefly even before looking at the structure of the methods themselves. One is that algorithms for the unconstrained minimization of rather arbitrary functions are well studied and generally are quite reliable. These methods are establishing a solid place for themselves in the numerical analysis spectrum and they have a considerable and sophisticated literature. A second reason for the appeal of the unconstrained formulation of the constrained problems is that the sequential nature of the methods allows, in some cases, a gradual or sequential approach to criticality of the constraints. In addition, the sequential process permits a graded approximation to be used in the analysis of the system. This latter allows coarse approximations to be used during early stages of the optimization procedure and finer or more detailed analysis approximations to be used during the later stages. A final reason for the appeal of the methods is that for some types of problems, the formulation and implementation using available computer programs is quite straightforward. This characteristic permits the generation of capabilities for solving the constrained optimization problem with a minimum of programming time. This is in contrast with the direct methods, discussed elsewhere in this volume, which may require extensive computer programming for their implementation.

A brief introduction to the basic structure of unconstrained formulations should help to provide an orientation for what follows. First of all, to restate the basic optimization problem we first examine the problem with inequality constraints only, of the form:

$$\text{Find } \vec{D} \text{ such that } M(\vec{D}) \rightarrow \text{minimum and}$$

$$h_j(\vec{D}) \leqslant 0 \quad ; \quad j = 1,2,\ldots J \quad . \tag{6-1}$$

This problem is converted to an unconstrained minimization problem by constructing a function of the form;

$$\phi(\vec{D},r) = M(\vec{D}) + P\,[h_1(\vec{D})\ ,\ h_2(\vec{D})\ ,\ \ldots\ h_J(\vec{D}),r] \tag{6-2}$$

where P is some function, which will be discussed later, of the constraints and of a parameter r such that violations of the constraints produce a penalty to be appended to the objective function in such a way that unconstrained minimization of ϕ tends, in a variety of ways for different methods, to the solution of the constrained minimization problem given by Eq. (6-1). There are a variety of P functions and strategies for applying the method and the most applicable of these will be discussed in later sections of this Chapter. Optimization problems involving both equality and inequality constraints may also be expressed in an unconstrained form. This is done through functions similar to (6-2), but including the effects of the equality constraints.

In any event, the ultimate goal of the formulation is to convert the original problem into an unconstrained problem in which the function ϕ can be minimized without regard for constraints and for which the minimum tends, in some sequential way, to the solution of the original problem. Therefore, the second aspect of these approaches is the utilization of an unconstrained minimization algorithm. Techniques for unconstrained minimization usually take the form of an iteration:

$$\vec{D}_{q+1} = \vec{D}_q + \alpha\,\vec{S}_q \tag{6-3}$$

where in α is a 'step-length' in some direction given by $\vec{S}_q$. This iteration is applied to the ϕ-function until a point is reached which is determined to be its minimum. Most of these methods owe their particular characteristics to the rationale used to determine α and $\vec{S}_q$. The effectiveness of the unconstrained minimization algorithm is crucial to the operation of the overall method and therefore a detailed discussion of some of these procedures will be taken up before their application to the constrained problem is considered.

6.2 Unconstrained Minimization Methods

6.2.1 Some Early Methods

In this section, we will examine the methods of solving the problem:

$$\text{Find } \vec{X} \text{ such that } F(\vec{X}) \rightarrow \text{Min} \tag{6-4}$$

where there are *no restrictions on the choice of* $\vec{X}$. The earliest and most primitive approach to the problem is that which goes under the various names gridding, exhaustive enumeration or exhaustive search. This approach is simply to select for each of the variables a range and an increment or spacing within this range and then to examine all possible combinations of the variables selecting that combination which produces the least value of the F-function. Simple arithmetic will reveal that in order to obtain any reasonable accuracy for even a modest number of variables, an enormous computational effort is involved in obtaining a solution. An alternative which is only slightly more effective is the random search.

The random search is nearly as simple as the grid search, but it has the advantage that on each successive sample, every point in the space is equally likely to be tested. It consists of generating a set of $\vec{X}$'s each component of which is a random number in some preselected range. Most computer libraries have random number generators and usually this can be done quite conveniently.

Comparison between the two methods (grid and random) is probably fruitless inasmuch as the results will depend heavily upon the function being searched and also because the methods would be used only when efficiency is really no object.

A random-based method which is somewhat more sophisticated is the random walk. The version which we will discuss is based upon the idea of a sequence of improved approximations to the minimum, each of which is derived from the preceding approximation. The sequence is determined from the prescription:

$$\vec{X}_{q+1} = \vec{X}_q + \rho\,\hat{e}_r \tag{6-5}$$

where $\vec{X}_q$ is the 'old' approximation to the minimum and $\vec{X}_{q+1}$ is the 'new' approximation, ρ is a scalar step length and $\hat{e}_r$ is a unit random vector. The algorithm is based upon the following steps:

(i) Choose a starting point $\vec{X}_o$ and a step length ρ which is large in relation to the final accuracy desired.

(ii) Generate $\hat{e}_r$.

(iii) Calculate $\tilde{F} \equiv F(\vec{X}_q + \rho\,\hat{e}_r)$.

(iv) If the result of (iii) is less than $F(\vec{X}_q)$, then set $\vec{X}_{q+1} = \vec{X}_q + \rho\,\hat{e}_r$ and repeat (iii) and (iv); otherwise, just repeat (ii), (iii) and (iv).

(v) If a sufficient number of trials produces no acceptable $\vec{X}_{q+1}$, reduce ρ and continue (ii), (iii), (iv).

(vi) When ρ has been reduced to within the accuracy desired, terminate.

This method, while slightly more efficient than the grid or pure random search, is still quite inefficient except on very small problems and is recommended only in cases where programming ease is the principal objective.

Further methods which should be mentioned are the gradient or steepest descent methods of unconstrained minimization. These methods are based upon the well-known property that the negative of the gradient direction is the direction in which the function decreases at the greatest possible rate. These methods all utilize the iteration of Eq. (6-3), with $\vec{S}_q$ equal to $-\nabla F$ evaluated at X_q. The different steepest descent methods are based upon different strategies and techniques for choosing α. The basic drawback of the gradient methods is that for functions with any degree of ill-conditioning, the iteration usually settles into a steady N-dimensional zig-zag and convergence becomes very slow. It should be noted that the ϕ—functions used in the methods discussed subsequently in this Chapter tend by their nature to be ill-conditioned.

6.2.2 One-Dimensional Minimization

One form of steepest descent method, while not notably effective as an overall method is based upon a strategy for picking α in Eq. (6-3) which has *important implications for other more practical* methods. The idea is to choose the α which minimizes F in the direction $\vec{S}_q = -\nabla F(\vec{X}_q)$. An obvious advantage of this approach is that each step will produce the greatest possible reduction in F and hence one might expect the process to converge faster than if the minimum were not sought. Another, more important, advantage, which will be discussed subsequently, is that by taking the minimizing step at each iteration of Eq. (6-3), certain very valuable properties will pertain.

Consider <u>any</u> vector $\vec{S}_q$ and the move prescription:

$$\vec{X} = \vec{X}_q + \alpha \vec{S}_q \tag{6-6}$$

where, if α is thought of as a variable, then the locus of $\vec{X}$ for a range of values of α is a straight line. Substituting this formally into $F(\vec{X})$ we obtain

$$F(\vec{X}) = F(\vec{X}_q + \alpha \vec{S}_q) = F(\alpha) \tag{6-7}$$

since F can be considered a function of α alone, ($\vec{X}_q$ and $\vec{S}_q$ are considered fixed). Here the value of α which minimizes $F(\alpha)$ is sought. Note that this value, denoted α^*, does not produce the global minimum of F unless, of course, the line $\vec{X} = \vec{X}_q + \alpha \vec{S}_q$ contains the global minimum point.

With this concept, the problem of minimizing $F(\vec{X})$ can be reduced to a succession of one-dimensional minimization problems regardless of the dimensionality of $\vec{X}$. In practice, α^* can rarely be obtained explicitly and generally we must resort to a numerical means for finding α^*.

Consider approximating the function $F(\alpha)$ by a function $h(\alpha)$ which has an easily determined minimum point. The simplest one variable function possessing a minimum is the quadratic

$$h(\alpha) = a + b\alpha + c\alpha^2 \tag{6-8}$$

the minimum of which occurs where

$$\frac{dh}{d\alpha} = b + 2c\alpha = 0 \tag{6-9}$$

or

$$\alpha^* = -\frac{b}{2c} . \tag{6-10}$$

The constants b and c for the approximating quadratic (a is not needed) can be determined by sampling the function at three different α values, $\alpha_1, \alpha_2, \alpha_3$ and solving the equations

$$\left.\begin{aligned} F_1 &= a + b\alpha_1 + c\alpha_1^2 \\ F_2 &= a + b\alpha_2 + c\alpha_2^2 \\ F_3 &= a + b\alpha_3 + c\alpha_3^2 \end{aligned}\right\} \tag{6-11}$$

where F_1 denotes the value $F(\alpha_1)$, etc. A choice of α_1, α_2 and α_3 for which Eq. (6-11) are particularly easy to solve and which can save one function evaluation is $0, t, 2t$ where t is a preselected trial step. Note that if F at $\alpha = 0$ is presumed known from the previous iteration only two new function evaluations are required. With this choice, Eq. (6-11) become

$$\left.\begin{aligned} F_1 &= a \\ F_2 &= a + bt + ct^2 \\ F_3 &= a + 2bt + 4ct^2 \end{aligned}\right\} \tag{6-12}$$

from which

$$a = F_1$$

$$b = \frac{4F_2 - 3F_1 - F_3}{2t} \tag{6-13}$$

$$c = \frac{F_3 + F_1 - 2F_2}{2t}$$

and

$$\alpha^* \simeq \frac{4F_2 - 3F_1 - F_3}{4F_2 - 2F_3 - 2F_1}\, t \quad . \tag{6-14}$$

For α^* to correspond to a minimum and not a maximum of $h(\alpha)$, α^* must satisfy

$$\left.\frac{d^2h}{d\alpha^2}\right|_{\alpha=\alpha^*} > 0 \quad . \tag{6-15}$$

For the case where h is quadratic, this requires $c > 0$ or

$$F_3 + F_1 > 2F_2 \quad . \tag{6-16}$$

A scheme for insuring that the condition Eq. (6-16) is satisfied and further that the minimum lies in the interval $0 < \alpha < 2t$ is as follows:

(i) Choose an initial value for $t = t_o$ based upon previous iterations or other information regarding a reasonable value for the step length. Ideally, t_o would be of the order of α^*.

(ii) Compute $F(t)$.

(iii) If $F(t) > F(0) \equiv F_1$ then set $F_3 = F(t)$ and cut t in half and repeat (ii); otherwise, set $F_2 = F(t)$, double t, and repeat (ii).

(iv) When a value of t has been obtained such that $F_2 < F_1$ and $F_3 > F_2$, compute α^* according to Eq. (6-14).

It should be noted that even a function possessing a single minimum in the space of $\vec{X}$ may have multiple minima along a line. If a test is made to insure that $F(\alpha^*) < F_1$, the process will not diverge or cycle; however, it is a good rule to try to select t so that only the nearest minimum to $\vec{X}_q$ is included in the interval $0 < \alpha < 2t$ if possible. This precaution is wise because some of the methods to be discussed later depend for their efficiency upon a smooth progression along the contours of the function.

A variety of tests are possible to ascertain if the approximation to the minimum (call it $\tilde{\alpha}^*$) is a sufficiently good approximation to the exact α^*. A sort of ultimate criterion is

$$F(\tilde{\alpha}^*) < \begin{cases} F(\tilde{\alpha}^* + \varepsilon) \equiv F^+ \\ \quad\text{and} \\ F(\tilde{\alpha}^* - \varepsilon) \equiv F^- \end{cases} \tag{6-17}$$

where ε is the minimum significant change of the variable in the direction under consideration. Computationally, this criterion has two main disadvantages: first, it requires two extra function evaluations and secondly, it is not really as certain as it seems since the values F^+ and F^- may be contaminated by roundoff noise rendering the results of the test inconclusive. An alternative is to compute an approximation to $dF/d\alpha$ at $\tilde{\alpha}^*$ as

$$\tilde{F}' \equiv \frac{F(\tilde{\alpha}^* + \Delta) - F(\tilde{\alpha}^* - \Delta)}{2\Delta} \tag{6-18}$$

where Δ is a numerically significant, but still small, change in α, and compare this with zero. The range of the absolute value of the derivative of h in the interval 0 to $2t$ can be used as a basis of comparison; in other words, the maximum value of $|dh/d\alpha|$ is either b (at $\alpha = 0$) or $b + 4ct$ (at $\alpha = 2t$) and these can be used to determine if $\tilde{F}'$ is sufficiently small. For example we might require

$$\tilde{F}' < \frac{2|b| + 4|c|t}{200} \tag{6-19}$$

which is 1/100th of the average of $|h'(0)|$ and $|h'(2t)|$. This sort of criterion still requires two additional function evaluations and it is not foolproof.

An alternative criterion which is practically 'free' from the computational point of view is the following: compare $F(\tilde{\alpha}^*)$ and $h(\tilde{\alpha}^*)$ and consider $h(\tilde{\alpha}^*)$ a sufficiently good approximation if they differ by a small amount. It can be shown that

$$h(\tilde{\alpha}^*) = F_1 \frac{(4F_2 - 3F_1 - F_3)^2}{8(F_1 - 2F_2 + F_3)} = a - \frac{b^2}{4c} \quad . \tag{6-20}$$

For example, we might require

$$\frac{|h(\tilde{\alpha}^*) - F(\tilde{\alpha}^*)|}{|h(\tilde{\alpha}^*)|} < \varepsilon \tag{6-21}$$

where ε is a small fraction, say 0.01.

If the criterion chosen for the accuracy of the minimum is not satisfied, the original algorithm can be reapplied at $\tilde{\alpha}^*$ or t, whichever is a better approximation, or a general quadratic fit can be made using the 'best' 3 of the points 0, t, $2t$ and $\tilde{\alpha}^*$.

It is easy to concoct numerous function interpolation schemes based on higher order polynomials using more sample points or finite approximations to derivatives. Such algorithms may have advantages in certain problems, but in giving the rein to one's imagination, care should be exercised to avoid excessive function calculation and algorithmic complication. If refinement of the minimum is necessary in ill-behaved problems, it is generally better to apply the same simple algorithm repeatedly in successive approximations than to attempt to construct an air-tight technique to secure the minimum in one trial.

In some cases, a higher order interpolation for the one-dimensional minimization is appropriate. In particular, if the function has continuous first partial derivatives, a 2-point cubic fit can be used economically. If the gradient of the function being minimized is easily obtained, it is reasonable to consider a minimization algorithm based upon derivatives of the function. Note that the derivative $dF/d\alpha$ is

$$\frac{dF}{d\alpha} = \sum_{i=1}^{N} \frac{\partial F}{\partial x_i} \frac{\partial x_i}{\partial \alpha} \quad . \tag{6-22}$$

In a move of the form of Eq. (6-6) $\partial x_i/\partial\alpha = s_i^{(q)}$. Therefore

$$\frac{dF}{d\alpha} = \sum_{i=1}^{N} \frac{\partial F}{\partial x_i} s_i^{(q)} \equiv \nabla F^T \vec{S}_q \tag{6-23}$$

where ∇F is evaluated at the point along $\vec{S}_q$ where the slope is to be determined. As with the previous method this method hinges on approximating $F(\vec{X} + \alpha\vec{S}) \equiv F(\alpha)$ by a function $h(\alpha)$. However, in this case rather than a quadratic, h is taken to be the cubic

$$h(\alpha) = a + b\alpha + c\alpha^2 + d\alpha^3 \quad . \tag{6-24}$$

Values of $F(A)$, $(dF/d\alpha)_A$, $F(B)$ and $(dF/d\alpha)_B$ are available and thus the parameters of $h(\alpha)$ can be determined from the solution of

$$\left.\begin{aligned} a + bA + cA^2 + dA^3 &= F_A \equiv F(A) \\ a + bB + cB^2 + dB^3 &= F_B \equiv F(B) \\ b + 2cA + 3dA^2 &= F'_B \equiv (dF/d\alpha)_A \\ b + 2cB + 3dB^2 &= F'_B \equiv (dF/d\alpha)_B \end{aligned}\right\} \tag{6-25}$$

and the minimum would be one of the two points where

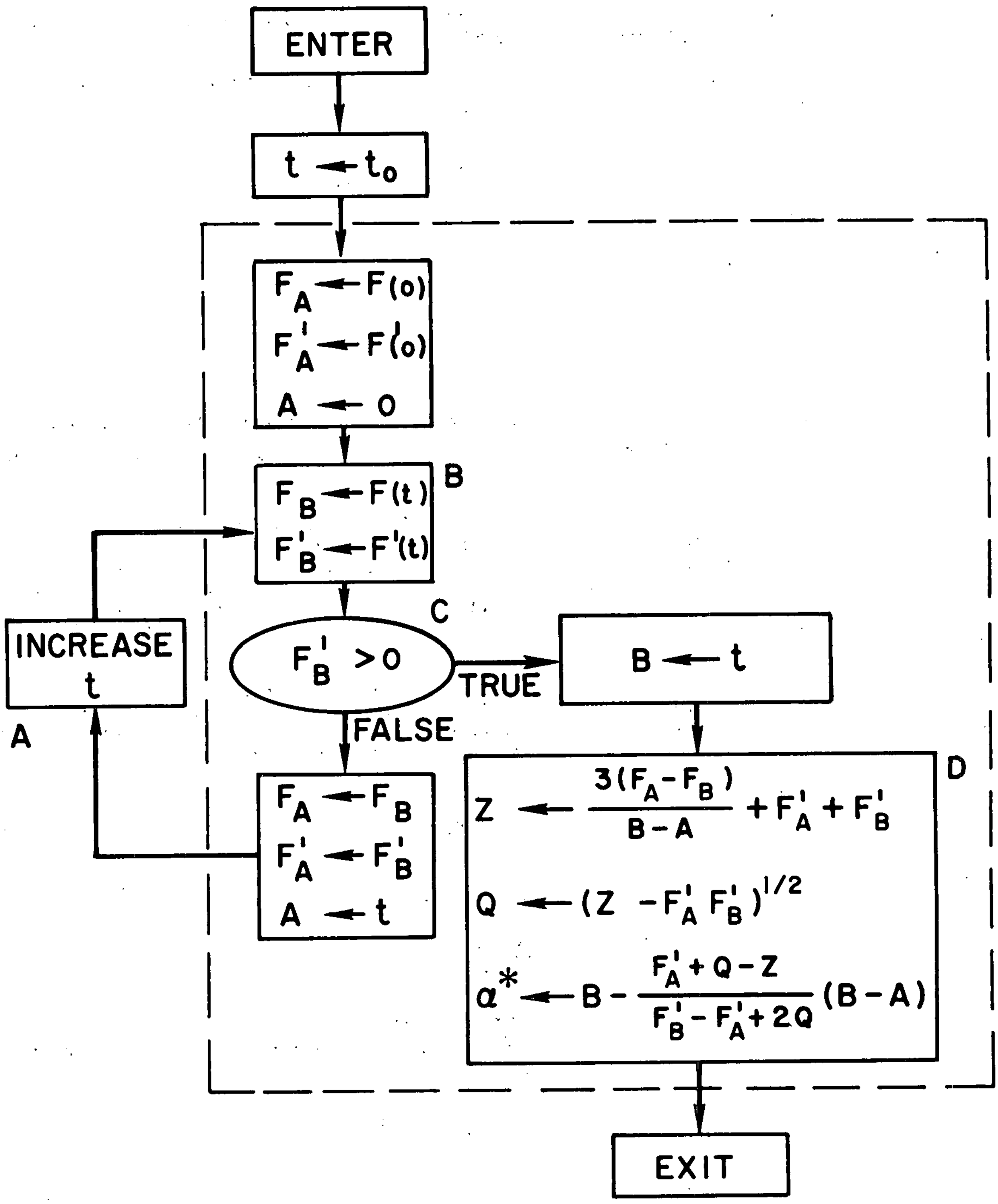


Fig.6.1 Flow Diagram for Cubic Interpolation

$$\frac{dh}{d\alpha} \equiv h' = 0 \tag{6-26}$$

or where

$$b + 2c\alpha + 3d\alpha^2 = 0 \quad . \tag{6-27}$$

Defining the quantities

$$Z = \frac{3(F_A - F_B)}{B - A} + F'_A + F'_B \tag{6-28}$$

and

$$Q = [Z^2 - F'_A F'_B]^{\frac{1}{2}} \tag{6-29}$$

the solution of Eq. (6-27) can be expressed as [6.7]

$$\alpha^* = B - \frac{F'_B + Q - Z}{F'_B - F'_A}(B - A) \quad . \tag{6-30}$$

The conditions $F'_A < 0$, $F'_B > 0$ insure that the estimated minimum point, α^*, will lie between A and B.

The flow diagram shown in Fig.6.1 is the logic for a basic algorithm using cubic interpolation. The two items left undetermined in this flow diagram, t_o and the contents of block A, are somewhat related and will be discussed together. The choice of t_o is crucial to efficiency since each traverse of the loop containing block A adds significantly to the labor involved in making the step. Indeed, in most problems the major effort of making an iteration is that expended in block B and ideally it would be done only once. The conflict is this: if t_o is chosen comfortably large so that F'_B is certain to be positive in the first pass through test C, the interpolation may take place over so large an interval as to produce a poor approximation. On the other hand, if t_o is too small, numerous increases in t may be necessary before test C is satisfied.

A number of techniques have been used to attempt to establish a proper range for t_o. Perhaps the most widely used a priori method is to assume initially that $F(\alpha)$ can be approximated by a quadratic and use $F(0)$, $F'(0)$ and a guess at the minimum value of the function, $\tilde{F}$, along $\vec{S}$, as the data for interpolation. Of course, this still leaves $\tilde{F}$ to be estimated. A low estimate of the minimum of $F(\alpha)$ may often be obtained easily and the use of this will generally result in overestimating t_o. Another approach is to estimate the expected reduction in F based upon preceding iterations.

The possibilities for estimating t_o are endless and what is efficient in one problem may be inappropriate in another. A careful eye should be kept on this aspect of the minimization routine, however, since this is usually where the time consuming computation is generated.

Once an estimate of α^* has been obtained the $F_* \equiv F(\vec{X} + \alpha^* \vec{S})$ can be computed. If F_* is less than both F_A and F_B, then at least $\vec{X}_*$ is a candidate for a minimum point. If this is indeed the case, the goodness of fit can be checked by calculating c, a measure of the orthogonality between the direction $\vec{S}$ and the gradient at α^*, which is given by

$$c = \frac{\vec{S}^T \vec{G}_*}{|\vec{S}|\ |\vec{G}_*|} \tag{6-31}$$

where $\vec{G}_* = \nabla F(\vec{X} + \alpha^* \vec{S})$.

The test $|c| < \varepsilon$ may be used as the final criterion for acceptance of α^*. Values for ε of from 10^{-2} down to 10^{-m}, where m is the number of working digits in the computer, have been used; however, these lower values can be very difficult to satisfy especially if there are many variables in the problem. The stringency of this orthogonality requirement should bear a relationship to the overall method in which the minimizing step routine is embedded and even at this level it cannot be stated with certainty what the best strategy is.

If the test for a minimum fails, then block D of Fig.6.1 may be re-entered and a new interpolation attempted. Before entering, it is merely necessary to test the sign of $\vec{S}^T \vec{G}_*$, if it is positive then set

$$B \leftarrow \alpha^*$$

$$F_B \leftarrow F_* \qquad (6\text{-}32)$$

$$F'_B \leftarrow \vec{S}^T \vec{G}_*$$

otherwise set

$$A \leftarrow \alpha^*$$

$$F_A \leftarrow F_* \qquad (6\text{-}33)$$

$$F'_A \leftarrow \vec{S}^T \vec{G}_* \ .$$

Since the formula for α^* is arranged so that $A \leqslant \alpha^* \leqslant B$, each refit will narrow the gap $B-A$, the size of which can also be tested as a precaution against pathological functions or overly zealous criteria for the minimum, and in principle the minimum can be located to within the desired accuracy by successive refits.

There are several types of schemes for one-dimensional minimization which have been omitted in this discussion and the reader is referred to the literature. Most of these are highly organized hunt and peck schemes with elegant logic behind them; however, their usefulness is generally limited to problems where the interpolation methods fail, for example in some of the discontinuous derivative cases. One method in particular deserves mention; the Fibonacci search which is based upon the fascinating Fibonacci numbers. This is a sampling method which traps the minimum in successively smaller intervals. For a lucid explanation of this and some related techniques, see Wilde and Beightler [6.1].

6.2.3 Quadratically Convergent Methods

Because most of the functions we will be minimizing have a convergent Taylor series at and near the minimum, it is useful to consider a quadratic approximation to the function. A Taylor series about any point $\vec{X}_o$ is of the form,

$$F(\vec{X}) \simeq F(\vec{X}_o) + (\vec{X} - \vec{X}_o)^T F(\vec{X}_o) + \tfrac{1}{2}(\vec{X} - \vec{X}_o)^T J(\vec{X} - \vec{X}_o) + \ldots \qquad (6\text{-}34)$$

where J is the matrix

$$J = \left[\frac{\partial^2 F}{\partial x_i \partial x_j}\right] \qquad (6\text{-}35)$$

and hence in the vicinity of the minimum we may think of F as approximated by

$$F \simeq \vec{X}^T A\vec{X} + \vec{X}^T\vec{B} + c \equiv Q \qquad (6\text{-}36)$$

for some matrix A, vector $\vec{B}$ and scalar c. A minimization method is said to converge quadratically if it will minimize a general quadratic in a finite and predetermined number of steps.

It is found that in practice a surprising number of functions are well approximated by a quadratic even at points moderately distant from $\vec{X}_m$ (the minimum point) and hence quadratically convergent methods are usually far more efficient for general applications than those lacking this property.

Most quadratic methods are based, in one way or another, on the concept of conjugate directions. In the context of the minimization of a quadratic function a set of N directions $\vec{S}_q$ is said to be conjugate or more accurately A-conjugate if

$$\vec{S}_i^T A \vec{S}_j = 0 \quad , \quad \text{for all } i \neq j \qquad (6\text{-}37)$$

where A is an $N \times N$ symmetric matrix.

A set of such directions possesses an extremely powerful property:

If a quadratic function Q is minimized sequentially, once along each direction of a set of N linearly independent, A-conjugate directions, the global minimum of Q will be located at or before the Nth step regardless of the starting point.

Note that the order in which the directions are used is immaterial to this property.

There is an interesting geometrical interpretation of this property. Starting from the point $\vec{X}_1$, if we minimize Q along $\vec{S}_1$, and then from the resulting point $\vec{X}_2$ minimize along $\vec{S}_2$ (which is A-conjugate to $\vec{S}_1$) then the resulting point is the minimum of Q in the plane containing $\vec{S}_1$ and $\vec{S}_2$ and passing through $\vec{X}_1$. In other words, it is the minimum in the plane

$$\vec{X} = \alpha_1\vec{S}_1 + \alpha_2\vec{S}_2 + \vec{X}_1 \tag{6-38}$$

where α_1 and α_2 are variables.

This result generalizes to the jth cycle in that the sequential minimization along the conjugate vectors $\vec{S}_i$, $i = 1,2,\ldots j$ produces the minimum point of Q in the subspace spanned by the vectors $\vec{S}_1,\ldots,\vec{S}_j$. Thus, at or before the Nth step, the global minimum point of Q will be reached.

It should be noted that these results require that each step must terminate at a minimum in the given direction. This point is emphasized because it is precisely the numerical difficulty of computing exactly the minimizing steps at each iteration which causes most of the practical problems with these methods.

The conjugacy relations do not define a unique set of directions, but any set of N independent, mutually A-conjugate directions will suffice. The various ways for generating such directions without knowing A form the basis for different methods which are quadratically convergent.

6.2.4 Powell's Method

A quadratically convergent method [6.2] which does not require the evaluation of the gradient of the function or any other derivatives will now be discussed. Consider a set of directions $\vec{S}_q$, $q = 1,2,\ldots N$ which are initially set equal to the coordinate vectors. That is, if we denote the ith component of $\vec{S}_q$ by s_{iq} then

$$s_{iq} = \delta_{iq} \quad ; \quad i,q = 1,2,\ldots N$$

where δ_{iq} is the usual Kronecker delta.

The method may be concisely outlined as follows:

(i) $\vec{Y} \leftarrow \vec{X}$ arbitrary ,

(ii) $\vec{X} \leftarrow \vec{X} + \alpha_i^* \vec{S}_i$; $i = 1,2,\ldots N$,

(iii) $\vec{S}_{N+1} \leftarrow \vec{X} - \vec{Y}$,

(iv) $\vec{X} \leftarrow \vec{Y} \leftarrow \vec{X} + \alpha_{N+1}^* \vec{S}_{N+1}$,

(v) $\vec{S}_i \leftarrow \vec{S}_{i+1}$, $i = 1,2,\ldots N$,

(vi) return to (ii) .

Thus, the method involves minimizing first once in each of the coordinate directions (actually any set of independent directions will do) and then in the direction defined by a vector from the starting point of the cycle to the ending point of the cycle. This so-called, 'pattern move' is in the direction of the trend of the collective minimizations in the coordinate directions. After this minimization is carried out, $\vec{S}_1$ is dropped and replaced by $\vec{S}_2$, $\vec{S}_2$ is replaced by $\vec{S}_3$ and so on until $\vec{S}_N$ is replaced by the pattern direction. The process is then repeated with the new set of directions.

Theoretically, more is required to make the method truly efficient on general functions, but the idea is contained in the above. The flow diagram shown in Fig.6.2 is a codification of the simplest version of the method. Note that a pattern direction is constructed (block A), then used for a minimization step (blocks B and C) and then it is stored in $\vec{S}_N$ (block D) as all of the directions are up-numbered and $\vec{S}_1$ discarded. The direction $\vec{S}_N$ will then be used for a step to a minimum just prior to the construction of the next pattern direction. As a consequence of this for the second cycle both $\vec{X}$ and $\vec{Y}$ in block A are points which are minima along $\vec{S}_N$, the last pattern direction. This sequence will impart special properties to $\vec{S}_{N+1} = \vec{X} - \vec{Y}$ which are the source of the rapid convergence of the method.

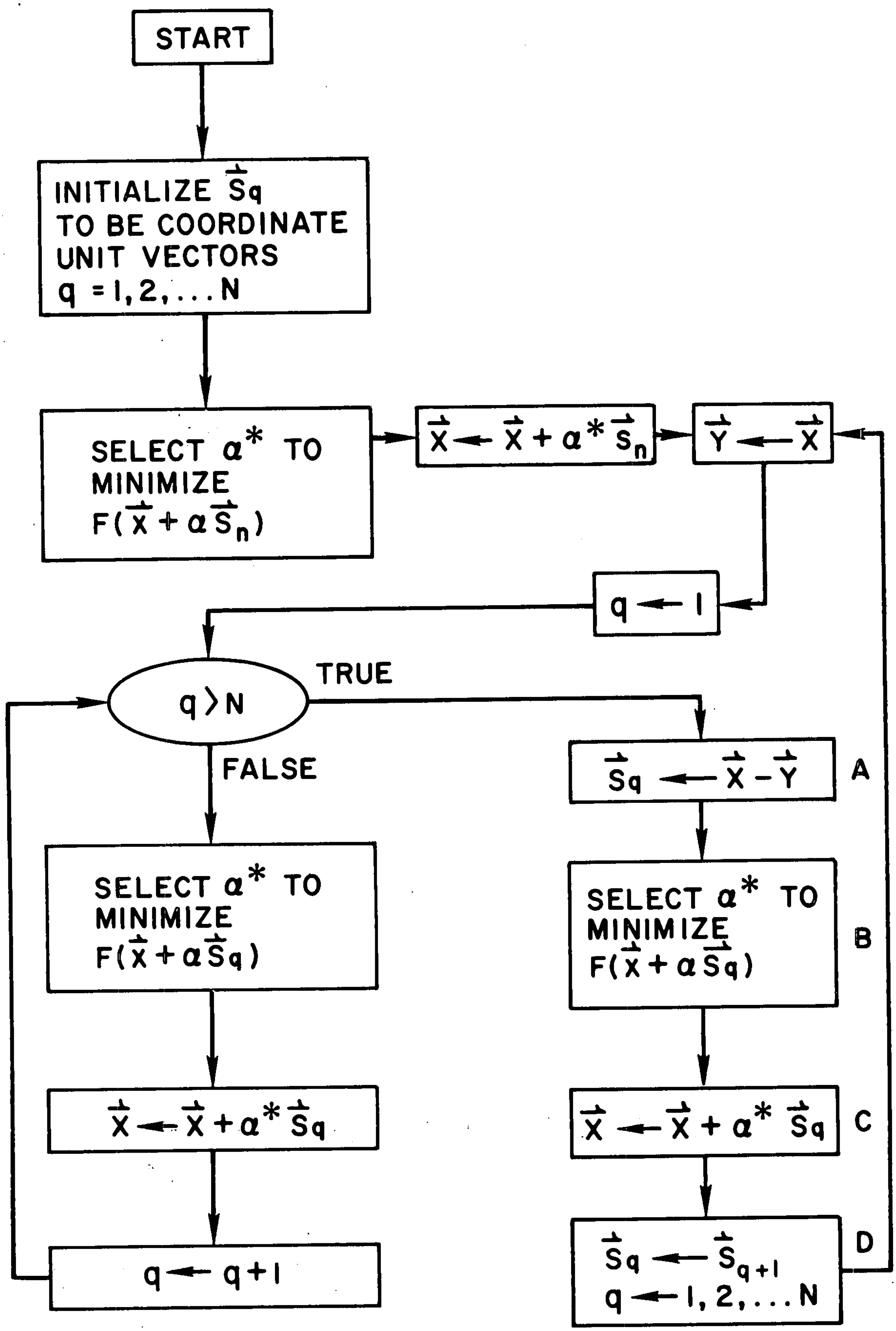


Fig.6.2 Flow Diagram for Powell's Method

We will now show that Powell's method generates conjugate directions. Given two vectors $\vec{X}_a$ and $\vec{X}_b$ and a direction $\vec{S}$; if $\vec{Y}_a$ is a minimum of Q from $\vec{X}_a$ along $\vec{S}$ and $\vec{Y}_b$ is a minimum from $\vec{X}_b$ along $\vec{S}$, i.e. if

$$\vec{Y}_a = \vec{X}_a + \alpha^*_a \vec{S} \tag{6-39}$$

$$\vec{Y}_b = \vec{X}_b + \alpha^*_b \vec{S} \tag{6-40}$$

then $\vec{Y}_a - \vec{Y}_b$ and $\vec{S}$ are A-conjugate. This fact is easily demonstrated starting with the definition of α^*. By definition

$$\frac{d}{d\alpha}\{Q(\vec{Y}_a + \alpha\vec{S})\} = 0 \ , \quad \text{at} \quad \alpha = 0 \tag{6-41}$$

and

$$\frac{d}{d\alpha}\{Q(\vec{Y}_b + \alpha\vec{S})\} = 0 \ , \quad \text{at} \quad \alpha = 0 \ . \tag{6-42}$$

Therefore, by substituting the above expressions into the equations of the quadratic, differentiating and then setting $\alpha = 0$, we obtain

$$\vec{S}^T(2\ A\ \vec{Y}_a + \vec{B}) = 0 \tag{6-43}$$

and

$$\vec{S}^T(2\ A\ \vec{Y}_b + \vec{B}) = 0 \tag{6-44}$$

and subtracting Eq. (6-44) from Eq. (6-43), we find

$$2\vec{S}^T\ A\ (\vec{Y}_a - \vec{Y}_b) = 0 \tag{6-45}$$

which demonstrates the conjugacy of $\vec{S}$ and $(\vec{Y}_a - \vec{Y}_b)$.

Returning now to the flow diagram of Fig.6.2, we see that in block A, both $\vec{X}$ and $\vec{Y}$ are minima along the direction $\vec{S}_N$ and therefore $\vec{S}_{N+1}$ is conjugate to $\vec{S}_N$. Thus, after N cycles, all of the $\vec{S}_q$ are mutually conjugate and a quadratic will theoretically be minimized in N^2 one-dimensional minimizations.

As is so often the case in these matters, things are not as good as they first seem. To begin with, the functions to be dealt with are not usually quadratics, and thus the number of iterations will ordinarily be greater than N. However, consider the least possible computational effort for N^2 minimizing steps. Suppose it requires at least three function evaluations per step, then for 50 variables it requires 7500 function evaluations to achieve minimization. In practice, moreover, it is found that even with luck, this can skyrocket to N^3 or more minimizations with 5 to 7 function evaluations each. This brings the number for 50 variables to around 700 000 evaluations!

In addition to the possibility of requiring a large number of function evaluations, the basic version of Powell's method described above can come to a halt before the minimum is reached. Both this complete failure and the previously described inefficiency are due to the fact that the $\vec{S}_j$ may become dependent or 'almost' dependent. The original set of $\vec{S}_j$ are, of course, independent and in theory each of the succeeding directions which are generated should be linear combinations of <u>all</u> of the preceding $\vec{S}_j$ unless some $\alpha^*_j = 0$ during the cycle. It has been found, however, that the basic method has a tendency to choose nearly dependent directions in ill-conditioned problems and for more than 5 variables the method can break down. One simple remedy is to reset the directions to the original coordinate vectors periodically and/or whenever there is some indication that the directions are no longer productive. This technique is sometimes useful but a procedure recommended by Powell [6.2] while somewhat more complicated, is very effective.

Powell recommends a termination criterion for ordinary use such that when a cycle produces a change in all variables of less than one-tenth of the required accuracy, the process is stopped. A safer (i.e. less likely to stop prematurely), but much more time consuming criterion also given by Powell is:

(i) Apply the normal procedure until a cycle causes a change of less than one-tenth of the desired accuracy. Call the resultant point $\vec{A}$.

(ii) Increase every variable by ten times the desired accuracy.

(iii) Apply the normal procedure until a cycle again causes a change of less than one-tenth of the desired accuracy. Call the resultant point $\vec{B}$.

(iv) Find the minimum on the line through $\vec{A}$ and $\vec{B}$; call it $\vec{C}$.

(v) Assume ultimate convergence if the components of $(\vec{A} - \vec{C})$ and $(\vec{B} - \vec{C})$ are less than one-tenth of the desired accuracy in the corresponding variables, otherwise

(vi) include the direction $(\vec{A} - \vec{C})$ in place of $\vec{S}_1$ (i.e. the x_1 direction) and restart the procedure from (i).

It should be mentioned that one of the most confounding problems in minimization, indeed of most iterative procedures, is that of termination. The preceding is a relatively safe rule, but it is expensive, (the problem must essentially be solved at least twice); in some problems, a more lax criterion may be appropriate and even other kinds of criteria may be reasonable. It is, however, difficult to set down general rules for termination with anything approaching confidence.

6.2.5 The Method of Conjugate Gradients

As has been mentioned already, the gradient, or steepest descent method when used with a minimizing step algorithm is not particularly efficient. The cause of inefficiency is a phenomenon called zigzagging. Note that in the iteration of Eq. (6-6) if the minimizing α (i.e. α^*) has been chosen, then the gradient at the new point, $\nabla F(\vec{X}_{q+1})$ is perpendicular to $\vec{S}_q$. To see this, observe that at α^*, $dF/d\alpha = 0$ and that $dF/d\alpha = \vec{S}_q^T \nabla F(\vec{X}_{q+1})$. This latter, of course, implies that $\vec{S}_q$ is orthogonal to $\nabla F(\vec{X}_{q+1})$. For eccentric functions, the process settles into an N-dimensional oscillation and convergence is often painfully slow. The convergence difficulties of the steepest descent method can be greatly reduced by a very simple modification which converts it to the conjugate gradient method [6.3], [6.4]. This consists of using an $\vec{S}_q$ in Eq. (6-7) defined by

$$\vec{S} = -\nabla F_{new} + \beta \vec{S}_{old} \tag{6-46}$$

where

$$\beta = \frac{|\nabla F_{new}|^2}{|\nabla F_{old}|^2} \tag{6-47}$$

or, writing the entire algorithm out,

$$\begin{array}{ll}
\text{(i)} & \vec{X}_o \leftarrow \text{arbitrary} \quad , \\
\text{(ii)} & \vec{G}_o \leftarrow \nabla F(\vec{X}_o) \quad , \\
\text{(iii)} & \vec{S}_o \leftarrow -\vec{G}_o \quad , \\
\text{(iv)} & \vec{X}_{i+1} \leftarrow \vec{X}_i + \alpha_i^* \vec{S}_i \quad , \\
\text{(v)} & \vec{G}_{i+1} \leftarrow \nabla F(\vec{X}_{i+1}) \quad , \\
\text{(vi)} & \beta_i \leftarrow |G_{i+1}|^2/|G_i|^2 \quad , \\
\text{(vii)} & \vec{S}_{i+1} \leftarrow -\vec{G}_{i+1} + \beta_i \vec{S}_i \quad .
\end{array} \tag{6-48}$$

Clearly from this definition $\vec{S}_{i+1}$ is a linear combination of $\vec{G}_{i+1}$ and $\vec{S}_o, \vec{S}_1, \ldots, \vec{S}_i$ and hence, it is a linear combination of $\vec{G}_o, \vec{G}_1, \ldots, \vec{G}_{i+1}$. Returning to the minimization of the quadratic $\vec{X}^T A \vec{X} + \vec{X}^T \vec{B} + c$, we have seen that if the $\vec{S}_i$ are A-conjugate, the minimum is attained in N or fewer steps. The process described by Eq. (6-48) is so constituted that the $\vec{S}_i$ satisfy the condition $\vec{S}_i^T A \vec{S}_j = 0$, $i \neq j$. This particular algorithm is derived from a Gram-Schmidt orthogonalization of the $\vec{G}_i$ [6.5]; for a different view see [6.3]. The conjugate gradient method was, in fact, originally proposed as a technique for solving any system of linear algebraic equations derived from the stationary conditions of a quadratic [6.4].

Theoretically, because the directions are A-conjugate, the process should converge in N or fewer cycles for a quadratic; however, for very badly conditioned quadratics, i.e. those with highly eccentric contours, it can take considerably more than N cycles. This phenomenon is due fundamentally to the finite digit arithmetic in which all actual calculations must be carried out. It manifests itself as a progressive contamination of $\vec{S}_i$, the only quantity carried over from iteration to iteration. All of the errors resulting from inaccuracies in the determination of α_i^* and the roundoff in accumulating the successive $\beta_i\vec{S}_i$ terms are carried forward in this vector. These difficulties lead to the need for occasionally 'restarting' the process, that is for setting $\vec{S}_q = -\nabla F(\vec{X}_q)$ and then continuing the standard process as before. In addition to a strategy for restarting, a great deal of improvement can be obtained by scaling the variables to reduce the eccentricity of the function. These and other topics are discussed in Fletcher and Reeves [6.3] and Fox and Stanton [6.6].

Essentially the conjugate gradient method is a good, efficient minimization technique which comes into its own for very large problems (say 150 variables and up) because of its modest storage and manipulative requirements. On the other hand, few design problems have this many design variables, and a more stable and reliable method, described in the next section, is more appropriate for the intermediate sized problem (10-50 variables).

6.2.6 The Davidon-Fletcher-Powell Variable Metric Method*

The conjugate gradient method is a quadratically convergent method but it suffers from a lack of stability when used on eccentric functions. In this section, we will describe a method which has much stronger stability although it involves a more elaborate computation to generate the steps of the iteration, which proceeds as follows:

(i) Start with an initial $\vec{X}_o$ and an initial positive definite symmetric matrix, H_o, (for example, the identity matrix) and set $\vec{S}_o \leftarrow -H_o \nabla F_o$.

(ii) Compute

$$\vec{X}_{q+1} \leftarrow \vec{X}_q + \alpha_q^* \vec{S}_q$$

where α_q^* minimizes $F(\vec{X}_q + \alpha \vec{S}_q)$.

(iii) Compute

$$H_{q+1} \leftarrow H_q + M_q + N_q \tag{6-49}$$

where, defining $\vec{Y}_q \equiv \vec{G}_{q+1} - \vec{G}_q \equiv \nabla F(\vec{X}_{q+1}) - \nabla F(\vec{X}_q)$,

$$M_q \leftarrow \alpha_q^* \frac{\vec{S}_q \vec{S}_q^T}{\vec{S}_q^T \vec{Y}_q}$$

and

$$N_q \leftarrow \frac{(H_q \vec{Y}_q)(H_q \vec{Y}_q)^T}{\vec{Y}_q^T H_q \vec{Y}_q} .$$

(iv) Compute

$$\vec{S}_{q+1} \leftarrow -H_{q+1} \vec{G}_{q+1}$$

and repeat from (ii).

The basic algorithm is extremely powerful for a first order method, i.e. one using only first derivatives of F, converging quadratically and possessing very good stability. By stability, we mean here that even in highly distorted and eccentric functions it continues to progress and needs little of the sort of special attention required by the conjugate gradient method. There is a plausible argument for this increase in stability in that with the conjugate gradient method, the entire history of the path is carried to $\vec{S}_{q+1}$ in the intelligence of $\beta_q \vec{S}_q$, a single vector. In the variable metric method, on the other hand, we carry the data in a full matrix which we carefully upgrade at each step. Another point of view is that the carryover term $\beta_q \vec{S}_q$ is only good if applied to $\nabla F(\vec{X}_q)$ and produces nonsense if applied to the gradient at some other point. On the other hand, it can be shown that H_q is a positive definite approximation to the matrix of second partial derivatives, the Hessian, and is applicable anywhere in the space.

*The method was essentially invented by Davidon, [6.7] and was further described and sharpened by Fletcher and Powell [6.8].

As will be seen, the positive definiteness is preserved in theory only if α_q^* is the true minimum point, i.e. if $\vec{G}_{q+1}^T\vec{S}_q = 0$, and furthermore, roundoff error may again dog our steps so that even this process can occasionally get into trouble. Before discussing modifications of the iteration to protect against this possible breakdown, we will state without proof (see Fletcher and Powell [6.8]) some important results concerning the theory of the method.

Again, returning to consideration of the quadratic

$$\tfrac{1}{2}\,\vec{X}^T A\vec{X} + \vec{X}^T\vec{B} + c$$

we state that for the iteration given by (i) through (iv):

(a) $\vec{S}_i^T A\vec{S}_j = 0; \quad i \neq j,$

(b) $\sum_{i=0}^{N-1} M_i = A^{-1}$,

(c) $H_o + \sum_{i=0}^{N-1} N_i = 0$,

(d) H_q is positive definite.

Thus (a) indicates that this is a conjugate direction method and hence is quadratically convergent while (b) and (c) show that $H_N = A^{-1}$ regardless of H_o. For the general nonquadratic problem (a), (b) and (c) have no exact meaning because there is no single A-matrix, but as the iteration nears the solution, $\vec{X}_m$, it is expected that H_q will tend to J_m^{-1}, where

$$J_m \equiv \left[\frac{\partial^2 F}{\partial x_i \partial x_j}\right]_{\vec{X}=\vec{X}_m} . \tag{6-50}$$

It may be shown that the matrix H_q is always positive definite even in the general problem and hence that the method is stable. Moreover, this matrix does not depend upon the form of F and its positive definiteness is influenced only by the accuracy with which α_q^* is determined. In applying the method, therefore, care must be exercised to insure that the H matrix is not updated with data arising from poor approximations to α_q^*. There are a number of approaches to this problem:

First, the algorithm used for computing α_q^* may be reapplied until $\vec{S}_q^T\vec{G}_{q+1}$ is sufficiently small; another alternative is simply to skip the 'update' cycle [step (iii)] when $\vec{S}_q^T\vec{G}_q$ is too large. In other words, if α_q^* is not close enough to the minimum along $\vec{S}_q$, set $H_{q+1} = H_q$ and $\vec{S}_{q+1} = H_{q+1}\vec{G}_{q+1}$ and continue as before. As long as $F_{q+1} < F_q$, the method will continue to progress towards the minimum.

It is difficult to choose between these approaches; the first may require excessive computation to refine α^* at points far from $\vec{X}_m$ while on the other hand, the second approach may miss valuable opportunities to improve the H-matrix. A reasonable compromise is to set a moderate criterion for $\vec{S}_q^T\vec{G}_{q+1}$, limit the number of refits to 1 or 2 and then skip the update if the criterion is not met after this.

Another area of numerical difficulty with the method has been identified [6.9]. This is a classical roundoff error problem. Suppose $H_o = I$; the elements of H_o are of the order of 1 and so are those of N_o but M_o is another matter. The elements of the latter matrix will be of order $|\alpha_o^*\vec{S}_o|/|\vec{Y}_o|$ which may be anything, depending upon the scale factors on F and $\vec{X}$. Consider minimizing bF where b is some positive scalar; M_o will be scaled by b but N_o will be unchanged. On the other hand, consider working in the space $a\vec{X}$ where a is a positive scalar; M_o will be scaled by a^2. The numerical significance of these relationships is that if the scaling turns out to be bad then in finite arithmetic, either

(a) $H_1 \approx H_o + N_o$

or (b) $H_1 \approx M_o$

and the latter form is singular. There is then little hope of recovery. Bard [6.9] recommends overcoming this problem by either increasing the precision of the arithmetic or scaling the variable appropriately. The initial scaling should, for these purposes, be such that the diagonal elements of M_o are approximately 1. The scaling should be rechecked and revised as necessary either if the method bogs down or if it is observed that the magnitude of the elements of H, M and N are consistently disparate.

In practice, the method is so powerful that difficulties seldom arise except on very badly distorted or eccentric functions. In such problems, however, the H-matrix will occasionally become indisposed in spite of all precautions and it will occur that $\vec{G}^T\vec{S}_q$ is positive, indicating that $\vec{S}_q$ is not a direction of descent. When this happens, the most efficacious remedy seems to be to set H back to H_o, or some other predetermined positive definite matrix, and proceed as if starting over again. The previously mentioned rescaling would have to be done in conjunction with a resetting of H. Of course, if this has to be done repeatedly and in many fewer cycles than N, the method would not be expected to work well.

Finally, we note that as with any gradient method, the computation of ∇F by finite difference can be considered for the variable metric method. Stewart [6.10] develops some special techniques for this purpose. Briefly, these involve the fact that since H_q is an approximation to $[\partial^2 F/\partial x_i \partial x_j]^{-1}$, we can extract an approximation to $\partial^2 F/\partial x_i^2$ from it. With this and an a priori estimate of the accuracy with which $F(\vec{X})$ itself can be computed, Stewart develops a solid estimate for the finite difference increment to produce maximum accuracy. With Stewart's modifications, this method becomes competitive with Powell's method for situations where formulas for the gradient components are not easily obtained.

6.3 Penalty Functions

The unconstrained minimization methods of the previous section are quite general and reliable for finding the unconstrained minimum of a function but are not usable for constrained problems without modification. Their reliability has, however, prompted the use of a variety of so-called penalty function formulations for solving the constrained problem. In this section, we will discuss a subclass of these formulations employing interior penalty functions, but first we note briefly the nature of the other main subclass, which employs exterior penalty functions. In this latter, the penalty term [$P(h_1,h_2,\ldots r)$ in Eq. (6-2)] is constructed so that when $\vec{D}$ is a point not satisfying the constraints, then P takes on some positive value which increases as the constraints are approached from outside the feasible region.

Usually at points inside the feasible region, P is zero. In the most common form of exterior penalty function, the parameter r is a simple multiplier of the penalty so that as r is increased P changes proportionally. The operation of the method is to choose a value of r, minimize ϕ and then check the constraints. If the constraints are sufficiently well satisfied, then terminate the method; otherwise increase r and minimize ϕ again. This sequence of unconstrained minimizations is continued until an optimum is found.

Some advantages of the method are that it allows the solution sequence to be started from an infeasible point, eliminating the need for a preliminary procedure to find an initial acceptable design as do most other methods. It provides a reasonably well-conditioned function to minimize, and the sequential nature of the method yields a set of starting points for the individual minimizations which are good initial approximations to the minima if r is changed a moderate amount each time.

The most serious disadvantage of the method is a need for careful weighting of the component parts of P for each h_j and no general procedure is available to select a satisfactory weighting. This failing can, in many problems, cause the method to be inoperable. For details of the exterior penalty function method see Zangwill [6.11].

6.3.1 An Interior Penalty Function

The exterior penalty method seeks to obtain an optimum feasible point by minimizing a penalty function for an increasing sequence of values of the penalty parameter. This technique forces the minimum point of $\phi(\vec{D},r)$ toward the feasible region from the outside. In this section, we discuss a penalty function, also for inequality constraints, which always has its minimum inside the feasible region and which, for a decreasing sequence of values of the penalty parameter r_i forces the minimum point $\vec{D}_{min}(r_i)$ towards the constrained optimum from the interior. This approach has a number of computational, as well as engineering advantages which will be discussed.

As with the exterior penalty function, the idea here is quite simple. The objective function is augmented with a penalty term which is small at points away from the constraints in the feasible region, but which 'blows up' as the constraints are approached. The most commonly used such function is:

$$\phi(\vec{D},r) = M(\vec{D}) - r \sum_{j=1}^{J} \frac{1}{h_j(\vec{D})} \tag{6-51}$$

where M is to be minimized over all $\vec{D}$ satisfying $h_j(\vec{D}) \leq 0$, $j = 1,2,\ldots J$.
Note that if r is positive, then since at any interior point all of the terms in the sum are negative, the effect is to add a positive penalty to $M(\vec{D})$. As a boundary is approached, some h_j will approach zero and the penalty will 'explode'. The penalty parameter, r, will be made successively smaller in order to obtain the constrained minimum of M.

To show how such a function looks, we consider the two bar truss optimization problem shown in Fig.6.3. The members are of tubular steel and the yield stress constraint is represented by

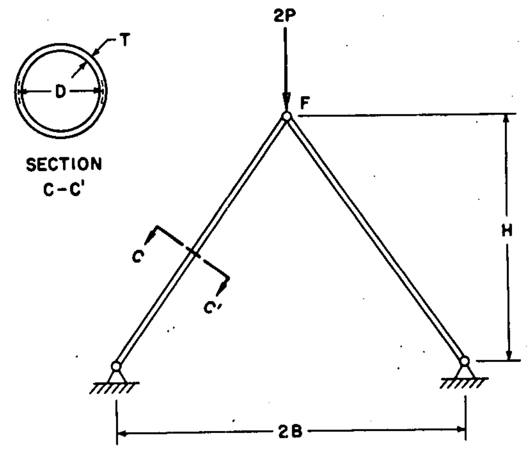


Fig.6.3 The Two Bar Truss

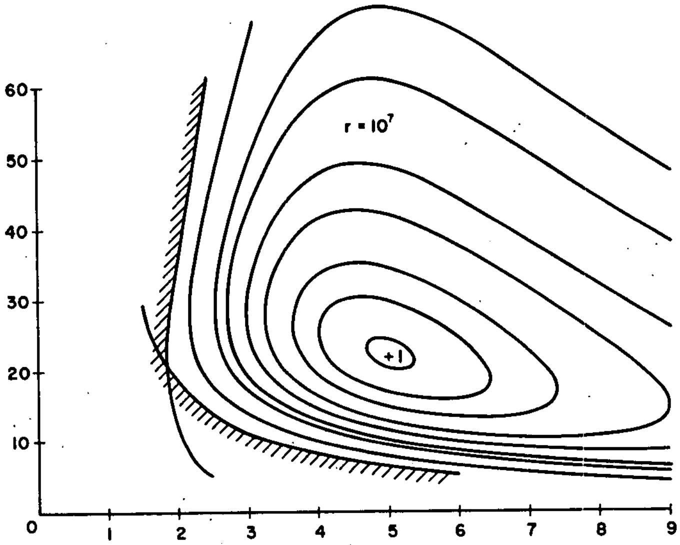


Fig.6.4 Interior Penalty Function for the Two Bar Truss, $r = 10^7$

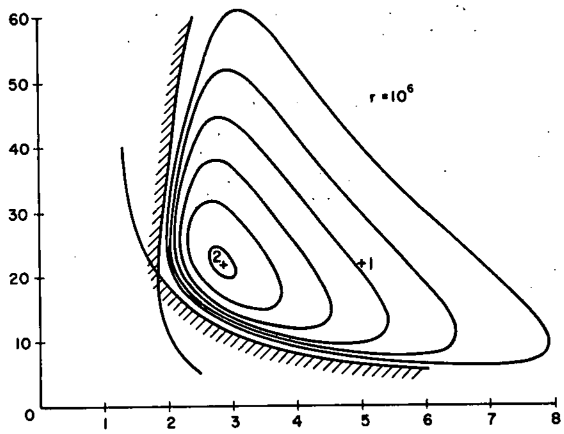


Fig.6.5 Interior Penalty Function for the Two Bar Truss, $r = 10^6$

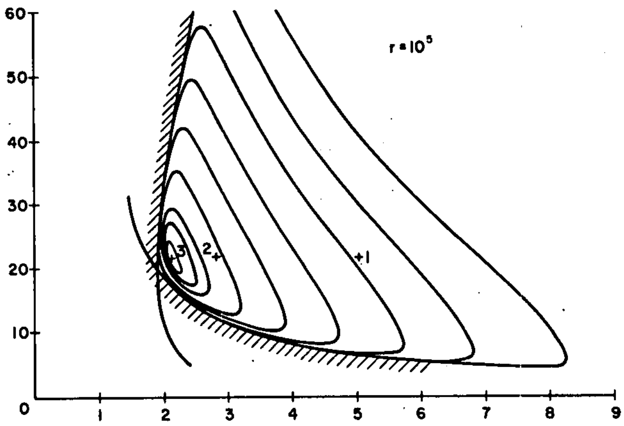


Fig.6.6 Interior Penalty Function for the Two Bar Truss, $r = 10^5$

$$h_1 \equiv \frac{P}{\pi t} \cdot \frac{(B^2 + H^2)^{\frac{1}{2}}}{Hd} - 100\ 000 \leqslant 0$$

for a material with a strength of 100 000 psi. The Euler buckling constraint is

$$h_2 \equiv \frac{P}{\pi t} \cdot \frac{(B^2 + H^2)^{\frac{1}{2}}}{Hd} - \frac{\pi^2 E}{8} \cdot \frac{(d^2 + H^2)}{(B^2 + H^2)} \leqslant 0 \quad .$$

The volume, which is to be minimized, is given by

$$M = 2\pi\ td\ (B^2 + H^2)^{\frac{1}{2}} \quad .$$

The penalty function plotted in Figs.6.4, 6.5 and 6.6 for this problem is thus

$$\phi = M - r\left[\frac{1}{h_1} + \frac{1}{h_2}\right] \quad . \tag{6-52}$$

The interior minima, indicated by '+' in the figures, for successively smaller values of r tend towards the constrained optimum of the original problem. It is also observed that the closer to the constrained optimum of M the minimum of ϕ is forced, the more eccentric the function becomes. This again leads to the necessity for sequential minimization of ϕ.

An algorithm which possesses the steps most commonly used is as follows:

(i) Given a starting point $\vec{D}_o$ satisfying all $h_j(\vec{D}) < 0$ and an initial value for r, minimize ϕ.

(ii) Check for convergence of $\vec{D}$ to the optimum.

(iii) If the convergence criterion is not satisfied, reduce r by $r \leftarrow rc$ where $c < 1$.

(iv) Compute a new starting point for the minimization, initialize the minimization algorithm and repeat from (ii).

There are a number of points to be considered in applying the method;

(a) The starting design, $\vec{D}_o$ required by (i) is usually available in engineering problems, but sometimes finding such a point may cause difficulty.

(b) A proper initial value for r must be selected.

(c) The possibilities for the convergence criteria of (ii) are numerous and there are choices to be made.

(d) Because of the sequential nature of the process, it is possible to improve the starting points for the third and subsequent minimizations.

(e) In some cases, considerable improvement in efficiency of the minimization method itself is possible by taking advantage of the special nature of the process.

6.3.1.1 Starting point

Starting with the first of these, we note that in many engineering situations, particularly in the structural and mechanical design areas, it is easy to find a point satisfying $h_j(\vec{D}) \leqslant 0$ at the expense of large values of M. For example, in structural design if cost or weight of the structure is ignored, it is usually easy to propose many designs which fulfill the basic requirements of strength and rigidity for the particular application. In other design situations, it may not be at all obvious what the acceptable designs are. In these situations, the initial acceptable design required by the interior penalty function method can be obtained as follows.

Suppose an engineering assessment of the situation has produced the design $\vec{D}_o$ which satisfies $h_j(\vec{D}_o) < 0$, $j = 1,2,\ldots m$, but has $h_j(\vec{D}_o) > 0$, $j = m+1,\ldots J$ where the expressions have been renumbered so that the last $J-m$ inequalities are the unsatisfied ones. Select k for which $h_k(\vec{D}_o)$ is a maximum where $k = m+1,\ldots J$ and temporarily define $h_k(\vec{D})$ to be the objective function for the problem:

Find $\vec{D}$ such that $h_k(\vec{D}) \to \text{Min}$ and

$$\text{(i)}\quad h_j(\vec{D}) \leq 0 \;, \quad j = 1,2,\ldots m$$

$$\text{(ii)}\quad h_j(\vec{D}) - h_j(\vec{D}_o) \leq 0 \;, \quad j = m+1, m+2, \ldots J \;.$$

Whenever, during the process of solving this problem by the penalty function method, the value of $h_k(\vec{D})$ drops below zero, the procedure is halted. The point so obtained satisfies at least one more constraint than the original $\vec{D}_o$. The procedure can be repeated until all the constraints have been satisfied and a $\vec{D}_o$ is obtained for which $h_j(\vec{D}_o) < 0$, $j = 1,2,\ldots J$. Ordinarily this approach should yield a point $\vec{D}_o$, if one exists, although there are circumstances where it will converge to a constrained or unconstrained local minimum of some $h_k(\vec{D})$ which is *positive*. Some ingenuity is required in such situations to select new starting points from which to repeat the sequence.

6.3.1.2 An Initial Value for r

The matter of selecting an initial value of the penalty parameter r has been the subject of discussion in the literature [6.12], but while there is some theory available, the task is still mainly an art. The problem is similar to one encountered with exterior penalty functions. If r is large, the function is easy to minimize, but the minimum may lie far from the desired solution to the original constrained minimization problem. On the other hand, if r is small the function will be hard to minimize.

A feeling for the problem can be developed by considering a few simple ideas. If the initial design is conservative, i.e. not near any constraints, one would like to pick the initial $r = r_o$ so that $M_{min}(r_o)$ would not increase drastically over the original design. In other words, r ought to be chosen small enough so that in the neighborhood of the initial design the $-r \sum 1/h_j$ terms do not completely dominate ϕ. A rule which might follow from this observation is that if $\vec{D}_o$ is a conservative design, pick r_o so that $-r \sum 1/h(\vec{D}_o)$ approximately equals $M(\vec{D}_o)$. In practice, this approach usually yields reasonable initial values for r.

If $\vec{D}_o$ happens to be a near-critical but nonoptimal design, i.e. such that one or more of the h_j are small but negative quantities, the situation becomes more complicated because the r value dictated by the above rule might be too small to allow the first minimization to be carried out. In this case, a proper value of r_o will probably be large enough so that in minimizing $\phi(\vec{D},r_o)$, M will increase from its value at $\vec{D}_o$. While this is distressing, it probably cannot be helped with this form of penalty function without a good deal of complex logic. Furthermore, unless something really drastic happens, very little is lost since r can be reduced quite quickly in this method.

Another approach to this latter problem which seems appealing in some cases, is to pick a relatively large value of r but to temporarily add a new constraint to the problem in the form of

$$h_{J+1} = M(\vec{D}) - M(\vec{D}_o) \leq 0 \tag{6-53}$$

or, to make it easier to get a starting point

$$h_{J+1} = M(\vec{D}) - [M(\vec{D}_o) + \varepsilon] \leq 0 \tag{6-54}$$

where ε is some small amount of increase which will theoretically be permitted in M during the first minimization. The penalty function for this revised problem is then

$$\phi(\vec{D},r) = M(\vec{D}) - r\left\{\sum_{j=1}^{J} \frac{1}{h_j} + \frac{1}{M(\vec{D}) - [M(\vec{D}_o) + \varepsilon]}\right\} \;. \tag{6-55}$$

The minimum for large values of r is approximately the point where the term in brackets alone is a minimum. As r is decreased, the fictitious constraint term can be removed or left in as desired since it will ultimately vanish.

6.3.1.3 Convergence Criterion

As the ϕ-function is minimized for various decreasing values of r, the sequence of minima, $\vec{D}_{min}(r_i)$, $i = 1,2,\ldots$ should converge to the solution of the constrained minimization problem and a means is needed to ascertain this convergence without an unnecessarily large number of minimizations. One simple criterion is to compute the relative difference

$$\delta = \frac{|M_{min}(r_{i-1}) - M_{min}(r_i)|}{|M_{min}(r_i)|} \tag{6-56}$$

and stop when δ drops below a specified value. It can require clever logic in some cases to prevent premature termination in situations where the process temporarily bogs down. Furthermore, the magnitude of δ must bear some relation to c, the fraction by which r is reduced each cycle. In general, however, this concept can form the basis for a useful convergence criterion.

An equally appealing group of convergence test numbers are contained in various norms of the vector

$$\vec{\Delta} \equiv \vec{D}_{min}(r_{i-1}) - \vec{D}_{min}(r_i) \quad . \tag{6-57}$$

For example, we could impose as a test for convergence

$$|\Delta_j| \leqslant \varepsilon_j \quad , \quad j = 1,2,\ldots N \tag{6-58}$$

or

$$\max(|\Delta_j|) \leqslant \varepsilon \tag{6-59}$$

or

$$|\vec{\Delta}| \equiv \left[\sum_{j=1}^{N} \Delta_j^2\right]^{\frac{1}{2}} \leqslant \varepsilon \tag{6-60}$$

and all of these have been used to advantage in various problems. The choice of norm and the proper value for ε depend upon the problem.

Another level of sophistication in methods of termination follows from the observation that $M_{min}(r_i)$ is merely a point on what one would expect to be a continuous function of r, namely, $M_{min}(r)$. This function can be approximated by a function $g(r)$ from data accumulated in two or more minimizations and then $g(o)$ will serve as an approximation to the true solution $M_{min}(0) \equiv M_{opt}$. If this approximation appears to be reliable and if the latest solution $M_{min}(r_i)$ is acceptably close to the latest approximation $g_i(o)$, then the process is terminated.

Computational experience and some theoretical support [6.12] suggest the use of an extrapolation function in the form of a polynomial in $r^{\frac{1}{2}}$. In particular, the most commonly [6.12] used form is

$$M_{min}(r) \simeq a + br^{\frac{1}{2}} \equiv g(r) \tag{6-61}$$

where the ith approximation is determined from interpolating

$$g_i(r_{i-1}) = a_i + b_i r_{i-1}^{\frac{1}{2}} = M_{min}(r_{i-1}) \tag{6-62}$$

$$g_i(r_i) = a_i + b_i(cr_{i-1})^{\frac{1}{2}} = M_{min}(r_i) \tag{6-63}$$

which leads to

$$a_i = [c^{\frac{1}{2}} M_{min}(r_{i-1}) - M_{min}(r_i)]/(c^{\frac{1}{2}} - 1) = g(0) \tag{6-64}$$

$$b_i = [M_{min}(r_{i-1}) - a_i]/r_{i-1}^{\frac{1}{2}} \quad . \tag{6-65}$$

This approximation scheme essentially fits a parabola to the data.

6.3.1.4 Improving the Starting Points, Extrapolation

The sequential process which converges the point $\vec{D}_{min}(r_i)$ toward the solution, $\vec{D}_{opt}$, is essentially a means for finding a sequence of good starting points for an ever more difficult sequence of minimization problems. It is possible to improve these further by using an extrapolation scheme similar to that given by Eq. (6-61) for extrapolating $M_{min}(r_i)$.

Writing a vector extrapolation for $\vec{D}_{min}(r)$ as

$$\vec{D}_{min}(r) \simeq \vec{A} + r^{\frac{1}{2}} \vec{B} \equiv \vec{Z}(r) \tag{6-66}$$

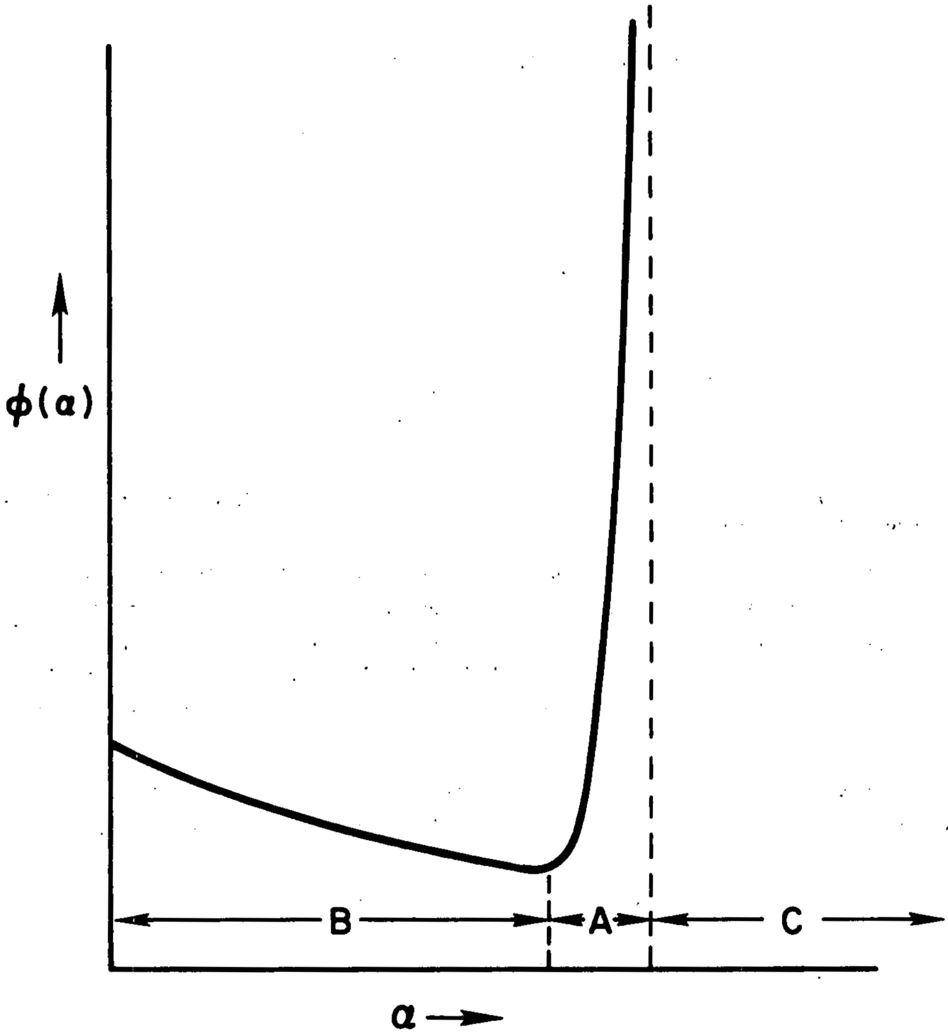


Fig.6.7 One-Dimensional Minimization of a Penalty Function

we can interpolate two known points $\vec{D}_{min}(r_{i-1})$ and $\vec{D}_{min}(r_i)$ from

$$\vec{Z}(r_{i-1}) = \vec{A} + r_{i-1}^{\frac{1}{2}}\,\vec{B} = \vec{D}_{min}(r_{i-1}) \tag{6-67}$$

and

$$\vec{Z}(r_i) = \vec{A} + (cr_{i-1})^{\frac{1}{2}}\,\vec{B} = \vec{D}_{min}(r_i) \tag{6-68}$$

which lead to

$$\vec{A} = \frac{c^{\frac{1}{2}}\,\vec{D}_{min}(r_{i-1}) - \vec{D}_{min}(r_i)}{(c^{\frac{1}{2}} - 1)} \tag{6-69}$$

and

$$\vec{B} = \frac{\vec{D}_{min}(r_{i-1}) - \vec{A}}{r_{i-1}^{\frac{1}{2}}} \quad . \tag{6-70}$$

From these, an improved starting point for the next value of r can be estimated:

$$\vec{Z}(r_{i+1}) = \vec{A} + c(r_{i-1})^{\frac{1}{2}}\,\vec{B} \tag{6-71}$$

or

$$\vec{Z}(r_{i+1}) = (c^{\frac{1}{2}} + 1)\,\vec{D}_{min}(r_i) - c^{\frac{1}{2}}\,\vec{D}_{min}(r_{i-1}) \quad . \tag{6-72}$$

It is, of course, necessary to check the extrapolated point $\vec{Z}(r_{i+1})$ against the constraints. If the constraints are satisfied, the vector may be used as a starting point. If not, and there is no guarantee that it will be, it must be abandoned. We can, however, attempt to salvage something of the extrapolation in these cases by taking a minimizing step in either the direction $\vec{S} \equiv \vec{D}_{min}(r_i) - \vec{D}_{min}(r_{i-1})$ or the direction $\vec{S} \equiv \vec{D}_{min}(r_i) - \vec{Z}(r_{i+1})$ from $\vec{D}_{min}(r_i)$. This will certainly produce a feasible point and will generally yield a good starting point for minimizing $\phi(\vec{D},r_{i+1})$.

6.3.1.5 Minimizing-Step Difficulties

The function defined by Eq. (6-51) cannot be minimized over the whole $\vec{D}$-space, but only in the interior of the feasible region $h_j < 0$. The ϕ-function is actually unbounded in both the positive and negative directions on the boundary of the feasible region and special steps must be taken to keep the minimization process in the proper portion of the space. An effective strategy for accomplishing this requires some ingenuity and it is not always clear what the best approach is.

The problem centers about finding the minimum when taking the step $\vec{D}_{q+1} = \vec{D}_q + \alpha\,\vec{S}_q$. In applying interpolation methods, the sample points should all be in the domain of definition and should preferably bracket the minimum. Fig.6.7 illustrates a hypothetical plot of ϕ vs α along some $\vec{S}_q$. From this figure it can be seen that the task involves two difficulties: (1) finding at least one sample point in Zone "A", and (2) getting a reasonable interpolation of this perverse function.

Approaches to the first part of the problem must take into account the nature of the search problem at hand: we seek a point in the narrow region, A, which is bordered on one side by the unacceptable region, C, and on the other by the negative slope region, B. Simple interval splitting schemes may be appropriate for this problem. That is, given a point in B and a point in C, take a point midway between them; if this point is in B, use it to replace the current B point and repeat, and similarly if the point falls in C use it to replace the current C point. This technique is hampered because zone B is distinguished from zone A by a difference in the sign of the slope of ϕ. When ϕ is of a nature where its derivatives are too difficult to compute, it may be necessary to use a crude finite difference scheme to locate the point.

Moe [6.13] has suggested some efficient approaches for coping with the difficulties in the one-dimensional minimization problem associated with interior penalty function methods. These techniques are based upon employing interpolated approximations for the h_j functions themselves rather than working with their reciprocals.

6.3.1.6 Engineering Implications of the Interior Penalty Function Method

An appealing feature of the interior penalty function method is the fact that, given an initial acceptable design, an improving sequence of acceptable designs is produced.

Moreover, the constraints are approached in this sequence in such a way that they become critical only near the very end of the procedure. This is a desirable feature in a structural design process because instead of taking the optimum design, a suboptimal but less critical design can be chosen if desired. Such designs are often said to have 'reserve capacity' to absorb overload or abuse and are prepared in advance for the performance upgrading processes which so often occur. The interior penalty function method is said to 'funnel the optimum design process down the middle', keeping the designs away from the constraint surfaces until final convergence.

In spite of the appeal of its simplicity, this approach to true safety is not endorsed here and the more direct methods of reliability based optimum design (see Chapter 10) should be used if these considerations are a factor. On the other hand, these ideas can sometimes be useful if applied intelligently and with a proper recognition of their true nature.

6.3.2 Penalty Functions for Equality Constraints

In many engineering design problems, a complicated or at least time consuming analysis must be performed to relate a set of values for the h_j to a particular set of values of the design variables $\vec{D}$. Often this analysis involves the solution of a system of algebraic equations of the form

$$\ell_i(\vec{D},\vec{Y}) = 0 \quad ; \quad i = 1,2,\ldots I \tag{6-73}$$

for the analysis variables $\vec{Y}$ for a given $\vec{D}$ and then computing the h_j from their explicit dependence upon $\vec{Y}$. If the penalty function method is applied to the direct formulation, each computation of the ϕ-function would require a new solution of the equations. For problems of the size considered practical from the analysis point of view in the aerospace industries, a large number of repetitions of such simultaneous equation solutions is expensive. Furthermore, in an increasing number of situations, the simultaneous equations are non-linear in the analysis variables $\vec{Y}$ and they require the application of iterative solution methods.

The fact that iterative solution methods can or must be used has motivated the development of penalty functions which include equality constraints. Almost all such methods are based upon the idea that one way of solving the equations

$$\ell_i(\vec{Y}) = 0 \quad ; \quad i = 1,2,\ldots I \tag{6-74}$$

for $\vec{Y}$ is to solve a minimization problem:

$$\text{Find } \vec{Y} \text{ such that: } \sum_{i=1}^{I} \ell_i^2 \rightarrow \text{Min} \quad .$$

If the above minimum is zero, then the corresponding $\vec{Y}$ is a solution to Eq. (6-74). The term to be minimized is sometimes referred to as the residual of those equations and is expressed as

$$R(\vec{X}) = \sum_{i=1}^{I} R_i(\vec{X}) \equiv \sum_{i=1}^{I} \ell_i^2 \tag{6-75}$$

where the dependence of R upon $\vec{X}$, i.e. $(\vec{D},\vec{Y})$, reflects that it is a function of both the design variables, $\vec{D}$, and the analysis variables, $\vec{Y}$.

It should be noted that solving the equation for $\vec{Y}$ by minimizing R is not generally the most efficient approach if the only purpose is to obtain a solution. This is because the residual is often a poorly conditioned function in $\vec{Y}$-space [6.6]. In linear problems, $A\vec{Y} = \vec{B}$, the 'conditioning' or measure of difficulty in obtaining accurate solutions is ordinarily related to the ratio of largest to smallest eigenvalue of the matrix of coefficients A. However, in residual minimization, it is related to the ratio of largest to smallest eigenvalues of A^TA, assuming A is symmetric. Thus, if A has a conditioning number of 100, then the residual has one of 10000 which is much worse.

A number of penalty functions for equality constraints have been described in the literature and some of these will be briefly presented here.

Fiacco and McCormick [6.13] report some success with the formulation

$$\phi(\vec{X},r) = M - r \sum_{j=1}^{J} \frac{1}{h_j} + r^{-\frac{1}{2}} \sum_{i=1}^{I} \ell_i^2 \tag{6-76}$$

where ϕ is minimized for a sequence of decreasing values of r. As r is made small, the second term does its familiar job of allowing the minimum to approach the constraints from the inside and the third term successively forces a satisfaction of $R = 0$. The reasons for the $-\frac{1}{2}$ power on r in the third term are given in [6.14]. The method works in principle and it has been used successfully on a number of problems. However, in many cases it presents an extremely difficult minimization problem and scale disparities between the terms $M - r \sum 1/h_j$, and $r^{-\frac{1}{2}} \sum \ell_i^2$ are hard to resolve.

An exterior penalty function of the same type has been proposed as

$$\phi(\vec{X},r) = M + r\left\{\sum_{j=1}^{J} \langle h_j\rangle^2 + \sum_{i=1}^{I} \ell_i^2\right\} \tag{6-77}$$

where

$$\langle z\rangle^2 = \begin{cases} 0, & z \leq 0 \\ Z, & z \geq 0 \end{cases} \tag{6-78}$$

which would be minimized for a sequence of increasing values of r. This formulation would doubtless suffer from the same scaling problem as the interior function.

A different approach to the problem is to consider the residual as the function to be minimized subject to the usual constraints $h_j \leq 0$, $j = 1,2,\ldots J$ plus a new constraint $M - M_o < 0$ where M_o is a constant selected as a goal for the objective function in a particular cycle of minimization. Thus, the problem is posed as:

$$\left.\begin{array}{ll} \text{Find } \vec{X} \text{ such that: } R(\vec{X}) \rightarrow \text{Min subject to:} & \\ \text{(i)} \quad h_j(X) \leq 0 & j = 1,2,\ldots J \\ \text{(ii)} \quad M - M_o \leq 0 \quad . & \end{array}\right\} \tag{6-79}$$

If an $\vec{X}$ for which $R(\vec{X}) = 0$ is obtained as a solution to this problem, then we have an acceptable design and its correct analysis and one which has a value of the objective function which is less than M_o. Optimization is carried out by solving the problem for a succession of decreasing values of M_o until one is chosen for which $R_{min}(\vec{X}) > 0$. The optimum design lies between the last two values of M_o and if the steps taken in M_o are small enough, the last successful design can be taken as a reasonable approximation to the optimum.

The alternative formulation given by Eq. (6-79) which treats the residual of the analysis equations $R(\vec{X})$ as an objective function can be attacked using either external [6.15], [6.16] or internal [6.17] penalty function methods.

There are many possibilities for the different segments of a program for the unconstrained minimization approach to equality and inequality constrained problems. It is definitely a situation where the algorithm must be tailored to the problem in order to be successful. These approaches for the general equality constrained problem represent a state-of-the-art situation; the problem is not really solved, but some useful approaches are available.

List of References

Ref.

6.1 Wilde, D. J. and Beightler, C. S., *Foundations of Optimization,* 1st ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1967

6.2 Powell, M. J. D., "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives," *The Computer Journal,* Vol.7, No.2, July, 1964, pp.155-162

6.3 Fletcher, R. and Reeves, C. M., "Function Minimization by Conjugate Gradients," *Computer Journal* (British), Vol.7, 1964, pp.149-154

6.4 Hestenes, M. R. and Stiefel, E., "Methods of Conjugate Gradients for Solving Linear Systems," *Journal Res. Natl. Bureau Standards,* Vol.49, No.6, December 1952, pp.409-436

6.5 Kowalik, J. "Iterative Methods for Large Systems of Linear Equations in Matrix Structural Analysis," Inter, Shipbuilding Progress, Vol.13, No.138, 1966

6.6 Fox, R. L. and Stanton, E., "Developments in Structural Analysis by Direct Energy Minimization," *AIAA Journal,* Vol.6, No.6, June 1968, pp.1036-1042

6.7 Davidon, W. C., "Variable Metric Method for Minimization," ANL-5990 Rev., November 1959, Argonne National Laboratory, University of Chicago, Lemont, Illinois

6.8 Fletcher, R. and Powell, M. J. K., "A Rapidly Convergent Descent Method for Minimization," *Computer Journal* (British), Vol.6, 1963, pp.163-168

6.9 Bard, Y., "On a Numerical Instability of Davidon-Like Methods," *Mathematics of Computation,* Vol.22, No.103, July 1968, pp.665-666

6.10 Stewart, III, G. W., "A modification of Davidon's Minimization Method to Accept Difference Approximations of Derivatives," *Journal ACM,* Vol.14, No.1, January 1967, pp.72-83

6.11 Zangwill, W. I., "Nonlinear Programming via Penalty Functions," *Management Science, Series A,* Vol.13, No.5, January 1967, pp.344-358

6.12 Fiacco, A. V. and McCormick, G. P., "Programming Under Nonlinear Constraints by Unconstrained Minimization: A Primal-Dual Method," RAC-TP-96, September 1963, Research Analysis Corporation. Bethesda, Maryland

6.13 Moe, J., "Design of Ship Structures by Means of Non-linear Programming Techniques," Proceedings AGARD Symposium on Structural Optimization, Istanbul, October 1968

6.14 Fiacco, A. and McCormick, G. P., "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming, *Manag. Sci. 10,* No.4, 1964, pp.601-617

6.15 Schmit, L. A. and Fox, R. L., "An Integrated Approach to Structural Synthesis and Analysis," *AIAA Journal,* Vol.3, No.6, June 1965, pp.1104-1112

6.16 Fox, R. L. and Schmit, L. A., "Advances in the Integrated Approach to Structural Synthesis," *Journal of Spacecraft and Rockets,* Vol.3, No.6, June 1966, pp.858-866

6.17 Schrader, M. J., "An Algorithm for the Minimum Weight Design of the General Truss," Case Western Reserve University, Master's Thesis, June 1968

Chapter 7

FEASIBLE DIRECTION METHODS

by

J. S. Kowalik

7.1 Introduction

The methods which are designed to solve a general non-linear programming problem

$$\text{minimize} \quad M(\vec{D}) \quad \text{subject to}$$

$$h_j(\vec{D}) \leq 0 \quad ; \quad j = 1,2,\ldots J \tag{7-1}$$

fall into the following two categories: methods which handle side constraints explicitly and those where formulation (7-1) is transformed to a sequence of unconstrained optimizations. Within the first category we can distinguish between (a) the methods where the non-linear problem is replaced by its linear approximation and solved in the repetitive manner by the simplex method, (b) the feasible directions methods which are discussed in this Chapter and (c) methods handling problems of a special nature, such as: separable programming, geometric programming, etc. We focus our attention on the second group and, in particular, the following three algorithms are presented here which have proved to be successful and applicable to structural optimization problems: the methods of Zoutendijk [7.1], Rosen [7.2] and Gellatly [7.3].

As far as theoretical validation is concerned the first two algorithms have been shown to converge to the global optimum for convex problems. In the general case where we cannot test our problem on convexity, we expect that these methods will find local solutions.

The question as to which of these three methods is preferable is difficult to answer without considering various aspects in conjunction with the problems which are being solved. Some of the most important aspects are: restrictions imposed on problems which the methods can handle, speed of convergence, ability to solve nonconvex or highly non-linear problems, ability to solve large scale problems, simplicity of code, etc. We will attempt to compare some of the merits of these methods and emphasize their advantages and disadvantages, from the theoretical and computational point of view.

The reader interested in a comparative numerical study of non-linear programming, restricted to the computational aspects of several methods tested on a few selected problems, is referred to a recent paper of Colville [7.4].

7.2 Zoutendijk's Usable Feasible Directions Method

7.2.1 Preliminary Considerations

The feasible directions method of Zoutendijk [7.1], [7.5] starts and operates inside the feasible region. It generates a sequence of feasible points $\vec{D}_1, \vec{D}_2, \ldots \vec{D}_{q+1}, \ldots$ such that for all q

$$M(\vec{D}_{q+1}) \leq M(\vec{D}_q) \quad , \tag{7-2}$$

where

$$\vec{D}_{q+1} = \vec{D}_q + \alpha_q \vec{S}_q \quad , \tag{7-3}$$

and

$$\alpha_q > 0 \quad .$$

The move from $\vec{D}_q$ to $\vec{D}_{q+1}$ is accomplished in two stages. In the first stage the direction finding problem is solved, i.e., the vector $\vec{S}_q$ is computed. In the second stage the step length α_q is found.

Assuming that the current approximation to the solution $\vec{D}_q$ is a feasible point (interior or located at the boundary) we say that a direction vector $\vec{S}_q$ is feasible if we do not immediately violate any constraint when making a sufficiently small step along this direction. Clearly, any direction $\vec{S}_q$ is feasible if $\vec{D}_q$ is an internal feasible point. If, however, $\vec{D}_q$ is a boundary point, then some vectors are directed to the outside of the feasible region and we cannot take a step of any length in these directions without violating some constraints.

We say that $h_j(\vec{D})$ is a critical constraint with respect to $\vec{D}_q$ if $h_j(\vec{D}_q) = 0$ and denote a set of all the critical constraints by J_c. Feasibility of $\vec{S}_q$ is assured if $\vec{S}_q$ satisfies the inequality

$$(\vec{S}_q)^T \nabla h_j(\vec{D}_q) \leq 0 \quad ; \quad j \in J_c \tag{7-4}$$

and all critical constraints are linear. If, however, some of the constraints are non-linear then (7-4) will not be, in general, sufficient, and we have to require that for the non-linear $h_j(\vec{D})$

$$(\vec{S}_q)^T \nabla h_j(\vec{D}_q) < 0 \quad ; \quad j \in J_c \quad . \tag{7-5}$$

Introducing a slack variable and individual scaling coefficients we get from (7-5),

$$(\vec{S}_q)^T \nabla h_j(\vec{D}_q) + C_j \sigma \leq 0 \quad ; \quad j \in J_c \tag{7-6}$$

where

$$\sigma > 0 \; , \quad C_j > 0 \quad .$$

Furthermore, we want the direction $\vec{S}_q$ to be usable, i.e., to be able to yield a reduced objective function value in the vicinity of $\vec{D}_q$. This requirement is: for $\alpha_q = 0$,

$$\frac{dM(\vec{D}_q + \alpha_q \vec{S}_q)}{d\alpha_q} = (\vec{S}_q)^T \nabla M(\vec{D}_q) \leq 0 \quad . \tag{7-7}$$

Any direction vector $\vec{S}_q$ satisfying the last two relations is usable-feasible and could serve our purposes.

Once we have obtained the direction $\vec{S}_q$ we have to find a step length $\alpha_q^* > 0$ which minimizes $M(\vec{D}_q + \alpha_q \vec{S}_q)$ and at the same time gives a new approximation to the solution $\vec{D}_{q+1}$ in the feasible region. The problem of finding α_q^* is a one-dimensional optimization problem and is solved by various search techniques. In some special cases, for example, when the objective function is quadratic and the constraints are linear then α_q^* can be found easily from explicit formulas. In more general cases this problem has to be solved by iterative techniques. The two methods most frequently used are; the golden section method and interpolation by low-order polynomials. In the first method the minimum is bracketed in an interval which is then systematically narrowed by comparing function values computed at the optimally chosen points inside the interval. The golden section method has a guaranteed convergence to the minimum but its rate of convergence is very slow if the minimum has to be found with high precision. In the second type of method the function is evaluated at several points and a low-order polynomial (typically quadratic or cubic) is fitted to it and the minimum of this interpolant is sought (see Chapter 6). Certain precautions are necessary to avoid divergence or convergence to unwanted stationary points. A comparison of these two approaches to the one-dimensional optimization can be found in [7.6].

7.2.2 Determination of Usable, Feasible Directions

To take into account two different feasibility requirements (7-5) and (7-6) we define our optimization problem as follows:

$$\begin{aligned} &\text{minimize } M(\vec{D}) \text{ subject to} \\ &h_j(\vec{D}) \leq 0 \; , \quad j \in J_N \\ &\vec{a}_j^T \vec{D} \leq b_j \; , \quad j \in J_L \end{aligned} \tag{7-8}$$

where $h_j(\vec{D}) \leq 0$ and $\vec{a}_j^T \vec{D} \leq b_j$ are the non-linear and linear constraints respectively. Let us also denote by J_{CN} and J_{CL} the sets of indices of the non-linear and linear critical constraints. The direction finding problem can now be formulated in the following manner:

given $\vec{D}_q$, find $\vec{S}_q$ and $\sigma > 0$ such that

$$\text{(i)} \quad (\vec{S}_q)^T \nabla h_j(\vec{D}_q) + C_j \sigma \leqslant 0 \ , \quad j \in J_{CN} \ , \tag{7-9}$$

$$\text{(ii)} \quad (\vec{S}_q)^T \vec{a}_j \leqslant 0 \ , \quad j \in J_{CL} \ , \tag{7-10}$$

$$\text{(iii)} \quad (\vec{S}_q)^T \nabla M(\vec{D}_q) + \sigma \leqslant 0 \ , \tag{7-11}$$

(iv) $\vec{S}_q$ is normalized by an additional requirement such as one of the following:

$$\text{(a)} \quad (\vec{S}_q)^T \vec{S}_q = 1 \ , \tag{7-12}$$

$$\text{(b)} \quad -1 \leqslant S_{qi} \leqslant 1 \ , \quad \text{all } i \ , \tag{7-13}$$

$$\text{(c)} \quad (\nabla M(\vec{D}_q))^T \vec{S}_q \leqslant 1 \ , \quad \text{etc.} \ , \tag{7-14}$$

$$\text{(v)} \quad \sigma \text{ is maximum} \ . \tag{7-15}$$

Any solution of (i)-(v) with $\sigma > 0$ gives a usable-feasible direction $\vec{S}_q$. If we select all $C_j = 1$ then we can interpret our auxiliary optimization problem (i)-(v) as an attempt to find a direction $\vec{S}_q$ in which the constraint functions $h_j(\vec{D})$ decrease about the same amount as the objective function in the vicinity of $\vec{D}_q$. It is desirable that this decrease be maximal.

In the case when only linear constraints are critical the auxiliary optimization problem reduces to:

given $\vec{D}_q$, find $\vec{S}_q$ such that

$$\text{(i)} \quad (\vec{S}_q)^T \vec{a}_j \leqslant 0 \ , \quad j \in J_{CL} \ , \tag{7-16}$$

(ii) an $\vec{S}_q$ normalization condition is satisfied and

$$\text{(iii)} \quad (\vec{S}_q)^T \nabla M(\vec{D}_q) \text{ is minimized} \ . \tag{7-17}$$

Both auxiliary problems are linear provided that a linear S-normalization is selected. Furthermore, if this condition is chosen to be $-1 \leqslant S_{qi} \leqslant 1$, which can be transformed to $0 \leqslant S_{qi} \leqslant 2$, then both auxiliary problems are linear programming problems with upper bounded variables. They can be solved by an efficient, special simplex method subroutine without the necessity of storing these normalization constraints. If the auxiliary problem leads to $\sigma > 0$ then $M(\vec{D})$ can be improved within the feasible region. If, however, we obtain $\sigma = 0$ then it can be demonstrated that $\vec{D}_q$ is the optimal solution.

7.2.3 Special Acceleration Techniques

Special precautions are necessary to guarantee and speed up the convergence of the feasible direction method. Careful investigation shows that the process described in Sections 7.2.1 and 7.2.2 may be very slow or nonconvergent due to so-called jamming which occurs when the algorithm generates a sequence of $(\vec{D}_q)$ which converge to a non-solution point. This happens when the sequence $(\vec{D}_q)$ becomes caught in a corner of the feasible region and is unable to leave it. This phenomenon was first observed by Zoutendijk [7.1] and numerical examples of the feasible direction procedures which lead to jamming when used to solve certain sample problems can be found in papers by Wolfe [7.7] and Zangwill [7.8].

Another common feature of all gradient methods is that sometimes a large number of very short steps are taken in strongly alternating directions. This is caused by a rapid change of the gradient vector in the direction of the feasible region (zigzagging). Small steps may also occur when the algorithm progresses along the boundaries. To prevent these inefficiences and secure convergence we can try to stabilize the search directions and keep an iterative solution away from the boundaries by including in the set of 'critical' constraints those nearly critical constraints which are likely to be approached. Let $J_{CN}(\vec{D},\varepsilon)$ denote the set of integers identifying those non-linear constraints for which the $h_j(\vec{D})$ are within ε of zero, i.e. $-\varepsilon \leqslant h_j(\vec{D}) \leqslant 0$. Similarly let $J_{CL}(\vec{D},\varepsilon)$ denote the set of integers identifying those linear constraints for which

$$- \varepsilon \leqslant \vec{a}_j^T \vec{D} - b_j \leqslant 0 \quad .$$

Then $J(\vec{D},\varepsilon)$ is the concatenation of these two sets of integers. These sets include as a particular case the sets of critical constraints, $J_{CN} = J_{CN}(\vec{D},0)$ and $J_{CL} = J_{CL}(\vec{D},0)$. Since we want to avoid the phenomenon of slow creeping along the boundaries we may solve a modified direction finding problem where $J_{CN}(\vec{D},\varepsilon)$ and $J_{CL}(\vec{D},\varepsilon)$ replace J_{CN} and J_{CL} respectively. The parameter ε should be reduced when small values of σ in the direction finding problem indicate that $\vec{D}_q$ approaches the optimal solution.

In a more refined procedure the constraints which have been encountered twice during the optimization process are kept in the critical set for a certain number of iterations. The following strategy has been successfully used in practice [7.9]:

(a) If at the current step of the iterative process the approximate solution $\vec{D}_q$ is on the boundary of a linear constraint (j) which has been met at least twice before, then the condition

$$(\vec{S}_q)^T \vec{a}_j \leqslant 0 \tag{7-18}$$

is added in the determination of $\vec{S}_q$ in subsequent problems. If, however, the variable σ has not improved by a significant amount from the previous step then only critical constraints are entered and the antijamming entries are deleted.

(b) If at the current step of the iterative process the point $\vec{D}_q$ is on the non-linear boundary j which has been approached previously then we require

$$(\vec{S}_q)^T \nabla h_j(\vec{D}_q) \leqslant 0 \tag{7-19}$$

in all auxiliary problems following the first one in which

$$(\vec{S}_q)^T \nabla h_j(\vec{D}_q) + \sigma \leqslant 0 \tag{7-20}$$

has to be required. We delete this requirement as soon as we arrive again at this constraint.

(c) In both cases (a) and (b), the antijamming inequalities are deleted if the current point is within the feasible region or if σ is less than some predetermined number (which can be gradually reduced).

The danger of zigzagging inside the feasible region can be avoided by introducing the principle of conjugate directions as an additional requirement in the direction finding subproblem, which may be expressed in the form

$$\{\nabla M(\vec{D}_{\ell+1}) - \nabla M(\vec{D}_\ell)\}^T \vec{S}_q = 0 \quad , \tag{7-21}$$

where $\ell = r, r+1, \ldots q-1$ and $\vec{D}_r$ is the last step located on the boundary. All the subsequent points $\vec{D}_{r+1}, \ldots, \vec{D}_q$ are interior-feasible.

The condition (7-21) is taken from the quadratic programming problems where the conjugacy of the search directions gives a computational procedure with a finite number of steps. In a more general problem it may be expected that the application of this principle improves the convergence properties of the algorithm.

7.2.4 Algorithm

This sample algorithm shows the essential computations which are executed to perform a single iteration step from $\vec{D}_q$ to $\vec{D}_{q+1}$, using the Zoutendijk method of feasible directions.

(i) *If* $\vec{D}_q$ *is a feasible interior point then*

$$\vec{S}_q = - \nabla M(\vec{D}_q)$$

is used as a usable-feasible direction.

A superior strategy would be to generate a conjugate direction using equations (7-21).

(ii) Otherwise the auxiliary subproblem (7-9)-(7-15) or (7-16)-(7-17), which can also include the antijamming precautions, is solved as described in Section 7.2.3.

(iii) If the auxiliary subproblem leads to a solution with $\sigma = 0$ to the required accuracy then $\vec{D}_q$ is assumed to be the optimum.

(iv) If $\sigma > 0$ then $\vec{S}_q$ is usable-feasible and the objective function will decrease in this direction.

(v) Determine the best step size α^*_q in the direction $\vec{S}_q$, i.e.

$$\alpha^*_q = \text{Min } M(\vec{D}_q + \alpha_q \vec{S}_q)$$

and set

$$\vec{D}_{q+1} = \vec{D}_q + \alpha^*_q \vec{S}_q .$$

(vi) If

$$M(\vec{D}_q) - M(\vec{D}_{q+1}) < \eta \tag{7-22}$$

where η is a preset small positive number, then the computations are terminated. Otherwise repeat from (i).

7.2.5 Summary of the Zoutendijk Method of Feasible Directions

Zoutendijk's method offers an efficient way of reducing the non-linear programming problem to a sequence of linear programming problems if a linear normalization of $\vec{S}_q$ is used. Furthermore, the method is finite for quadratic programming problems and can handle nonconvex problems. From the practical point of view, it offers an additional advantage of generating feasible intermediate approximations to the solution. The method has been used successfully to solve realistic problems [7.10]. Following Zoutendijk's critique [7.5] we indicate the following disadvantages of this method:

(a) The determination of the steplength α^*_q is a time consuming process which has to be performed in every step.

(b) The computer program is rather complicated and has to include antijamming precautions.

There are several questions which can be investigated and answered only on the basis of extensive computational experience, such as:

(a) What is an appropriate choice of the parameters $C_j > 0$ and of antijamming devices (both are probably heavily formulation and problem dependent).

(b) What type of $\vec{S}$ - bounding gives the most efficiently solvable subproblems.

(c) To achieve the best overall efficiency should we take the optimal steps α^*_q, i.e. $\alpha^*_q = \underset{\alpha_q}{\text{Min}}\, M(\vec{D}_q + \alpha \vec{S}_q)$ or just try to satisfy relation:

$$M(\vec{D}_q + \alpha_q \vec{S}_q) < M(\vec{D}_q) .$$

7.2.6 Modified Feasible Directions Method

It is worth-while to sketch briefly a recent version of the feasible directions method suggested by Zoutendijk [7.5] and referred to by him as MFD (Modified Feasible Directions). The original non-linear programming problem (7-8) is converted to a form with a linear objective function by adding $M(\vec{D}) + h_o \leqslant 0$ to the constraints and maximizing h_o. The method uses the linearization technique extensively and generates three sequences of points:

(a) Interior feasible points $\vec{D}_q$, such that $M(\vec{D}_{q+1}) < M(\vec{D}_q)$, which converge to the solution.

(b) Infeasible points $\vec{A}_q$ with nondecreasing values of $M(\vec{A}_q)$ giving a lower bound for a minimum.

(c) Boundary points $\vec{B}_q$ giving at each step an upper bound for the minimum.

To start the computation a feasible initial point $\vec{D}_o$ is needed. The algorithm consists of two phases.

Initial Phase

(i) Solve the linear auxiliary problem L_o which is minimize $M(\vec{D})$, or $-h_o$ if $M(\vec{D})$ is nonlinear, subject to the linear constraints of the problem

$$\vec{a}_j^T \vec{D} - b_j \leq 0 \tag{7-23}$$

and the additional restriction

$$|D_i| \leq \alpha$$

where α is a sufficiently large positive number. Call the solution $\vec{A}_o$ and proceed to (iii).

Iteration Phase

(ii) Solve the subsequent auxiliary linear problem L_q $(q \geq 1)$ and call the solution $\vec{A}_q$.

(iii) Find the boundary point $\vec{B}_q$ which is located on the line joining $\vec{A}_q$ and $\vec{D}_q$, i.e.

$$\vec{B}_q = \vec{D}_q + \alpha_q(\vec{A}_q - \vec{D}_q) \tag{7-24}$$

with the maximum α_q for which $\vec{D}_q$ is feasible.

(iv) All constraints for which $h_j(\vec{B}_q) = 0$ are linearized with respect to $\vec{B}_q$,

$$\nabla h_j(\vec{B}_q)\ (\vec{D} - \vec{B}_q) \leq 0 \tag{7-25}$$

and these linear inequalities are added to the constraints of the current linear problem L_q. This enlarged set of constraints will be used in the auxiliary problem L_{q+1}.

(v) A new point $\vec{D}_{q+1}$ is computed which is interior feasible and is located between $\vec{D}_q$ and $\vec{B}_q$, that is

$$\vec{D}_{q+1} = a\,\vec{D}_q + (1 - a)\,\vec{B}_q \quad , \quad 0 < a < 1 \quad . \tag{7-26}$$

(vi) If $M(\vec{B}_q) - M(\vec{A}_q) < \varepsilon$ then stop. Otherwise, $q \leftarrow q+1$ and the process is repeated from (ii).

7.2.7 Summary of the Modified Feasible Directions Method

(a) The modified Zoutendijk algorithm utilizes some of the ideas of the cutting plane method of Kelly [7.11]. However, in contrast to that method which produces infeasible points, the MFD method generates the feasible sequence $(\vec{D}_q)$.

(b) Computational performance of the method is not known to the author of this paper, but the method should be efficient for problems with nearly linear constraints.

(c) It is possible to foresee some computational problems similar to those encountered in the cutting plane method. We may have bad conditioning of linear problems due to near-dependency of constraints, which occurs close to the solution. This may probably be prevented by removing nonactive linearizations from the linear subproblems.

(d) In the problems where the feasible region defined by the constraints is nonconvex there is a possibility that some portions of the feasible region can be cut off by the tangential planes. A simple rule enables us to avoid this danger. From time to time all the constraints are checked and if for some of them $h_j(\vec{A}_q) < 0$ then the linearizations of $h_j(\vec{D})$ which determine the solution $\vec{A}_q$ are taken out in the next auxiliary linear problem.

(e) The method can be speeded up by using the principle of conjugate directions.

7.3 The Gradient Projection Method

7.3.1 Preliminary Considerations

The gradient projection method of Rosen [7.2] in contrast to Zoutendijk's method does not require the solution of auxiliary linear optimization problems. It uses projections of the objective function gradient into the manifold defined by constraints which are currently active. The method works with vectors $\vec{S}$ which are feasible and usable, that is vectors which satisfy the relationships $\vec{S}^T \nabla M < 0$ and $\vec{S}^T \nabla h_j = 0$. The latter is required for all active constraints. We assume here, that all the constraints are linear $h_j(\vec{D}) = \vec{a}_j^T \vec{D} - b_j \leq 0$, and the critical constraints have indices $j = j_1, j_2, \ldots j_k$. It is convenient to introduce the matrix of constraint gradients

$$N_k = [\nabla h_{j_1}, \nabla h_{j_2}, \ldots, \nabla h_{j_k}] \tag{7-27}$$

so that the feasibility condition can be stated concisely by

$$N_k^T \vec{S}^T = 0 \quad . \tag{7-28}$$

In the iterative process we move from $\vec{D}_q$ to $\vec{D}_{q+1}$ using the relationship

$$\vec{D}_{q+1} = \vec{D}_q - \alpha_q P \nabla M(\vec{D}_q) \tag{7-29}$$

where the matrix P projects $\nabla M(\vec{D})$ into the manifold formed by the active constraints. The projected vector $P \nabla M(\vec{D})$ can be obtained from $\nabla M(\vec{D})$ by subtracting from it the vector $N_k \vec{V}$, where $\vec{V}$ is such that

$$N_k^T (P \nabla M(\vec{D})) = N_k^T (\nabla M(\vec{D}) - N_k \vec{V}) = 0 \tag{7-30}$$

which leads to

$$\vec{V} = (N_k^T N_k)^{-1} N_k^T \nabla M(\vec{D}) \quad , \tag{7-31}$$

and

$$P = I - N_k (N_k^T N_k)^{-1} N_k^T \quad . \tag{7-32}$$

Matrix P is called a projection matrix and it projects every vector into the intersection of the $k \leq n$ hyperplanes (linear critical constraints). It is assumed that all columns of N_k are linearly independent from which it follows that $(N_k^T N_k)$ is nonsingular and can be inverted.

The normalized directions $\vec{S}_q$ can be found from

$$\vec{S}_q = - P \nabla M(\vec{D}_q)/|P\nabla M(\vec{D}_q)| \quad . \tag{7-33}$$

If $\vec{S}_q \neq 0$ then it is possible to find $\vec{D}_{q+1}$ such that $\vec{D}_{q+1}$ is feasible and $M(\vec{D}_{q+1}) < M(\vec{D}_q)$. If, however, $\vec{S}_q = 0$ then from (7-30) we have

$$- \nabla M(\vec{D}_q) = - N_k \vec{V} \quad , \tag{7-34}$$

i.e., the negative gradient of the objective function can be expressed as a linear combination of the gradients of the active constraints. If all components of $-V$ are nonnegative then the first order necessary conditions of Kuhn-Tucker for $\vec{D}_q$ to be the minimum are satisfied and the computation is terminated. In the case when this condition does not hold, then the computation is continued after the projection matrix is modified by deleting from N the column which corresponds to the most negative component of $-V$. By releasing a critical constraint which corresponds to the negative component of $-V$, a lower value of $M(D)$ can be obtained. It may also occur that $\vec{D}_{q+1}$ which gives the manifold optimal value of $M(\vec{D})$ is located at a new constraint (hyperplane). We then have to form a new manifold by adding this constraint to the set of critical constraints. In consequence, a considerable computational effort is involved in the periodical updating of the projection matrix P. This problem will be discussed in Section 7.3.3.

7.3.2 Algorithm

The following are the steps to compute $\vec{D}_{q+1}$ from $\vec{D}_q$ using the gradient projection method:

(i) Compute $\vec{S}_q = - P \nabla M(\vec{D}_q)/|P\nabla M(\vec{D}_q)|$,

where $P = I - N_k (N_k^T N_k)^{-1} N_k^T$

and N_k includes all currently critical (linearly independent) constraints.

(ii) If $\vec{S} \neq 0$ the one-dimensional minimization problem is stated as follows:

$$\alpha_q^* = \text{Min } M(\vec{D}_q + \alpha_q \vec{S}_q), \quad 0 \leq \alpha_q \leq \alpha^{max}$$

where α^{max} is the largest step which may be taken from $\vec{D}_q$ along $\vec{S}_q$ without violating any constraints. This value is computed from

$$\vec{a}_j^T(\vec{D}_q + \alpha_j^{max} \vec{S}_q) - b_j = 0$$

for those j for which $\vec{a}_j^T \vec{S}_q > 0$ and $j \in J_n$ where J_n denotes the current set of noncritical constraints. We have to accept the smallest α_j^{max} from the set of all these values, i.e.,

$$\alpha^{max} = \text{Min} \left[\frac{\vec{a}_j^T \vec{D}_q - b_j}{- \vec{a}_j^T \vec{S}_q} \right]$$

clearly $\alpha^{max} > 0$ since $\vec{a}_j^T \vec{D}_q - b_j < 0$ ($\vec{D}_q$ is feasible) and $- \vec{a}_j^T \vec{S}_q < 0$.

Two cases should now be considered:

(a) If $\alpha_q^* = \alpha^{max}$ then *some* new constraints (one or more) become active $[h_j(\vec{D}_{q+1}) = 0]$ and should be added to the matrix N_k. The projection matrix is modified and the computation returns to (i).

(b) If $\alpha_q^* < \alpha^{max}$ then matrix P remains unaltered and the computation returns to (i).

(iii) If $\vec{S} = 0$ then we compute vector $-\vec{V}$ from (7-31)

$$-\vec{V} = -(N_k^T N_k)^{-1} N_k^T \nabla M(\vec{D}_q) .$$

Two cases are possible:

(a) All components of $-\vec{V}$ are nonnegative, which indicates that a minimum has been found and the computation is terminated.

(b) If some components of $-\vec{V}$ are negative then the column ∇h_j corresponding to the most negative component is deleted from N_k, matrix P is modified and the computation returns to (i).

Remark

The method can easily handle linear equality constraints. Suppose our constraints are $h_j(\vec{D}) = \vec{a}_j^T \vec{D} - b_j \leq 0$, $j = 1,\ldots,m$ and $h_j(\vec{D}) = \vec{a}_j^T \vec{D} - b_j = 0$, $j = m+1,\ldots,J$. We reduce the N-dimensional space E of the original problem to the manifold determined by the intersection of the J-m hyperplanes

$$h_j(\vec{D}) = 0 , \quad j = m+1,\ldots,J .$$

That means that all the feasible points $\vec{D}_q$ must lie in the manifold defined above. With this restriction the problem with equality and inequality constraints can be treated as one having inequalities only. Computationally this can be accomplished by forming initially the matrix $(N_k^T N_k)^{-1}$ where $N_k = [\nabla h_{m+1}, \ldots, \nabla h_J]$, $k = J-m$ and keeping vectors $\nabla h_{m+1}, \ldots, \nabla h_J$ in N_k throughout the whole computing process.

7.3.3 Computational Aspects of the Gradient Projection Method

A considerable computational problem is introduced by the periodical updating of the projection matrix. Fortunately the subsequent matrices N differ usually by only one column ∇h_j which is either dropped from the set of active constraints or is added to it. It is possible to avoid the complete recomputation of $(N_k^T N_k)^{-1}$ from its definition, which takes $O(k^3)$ multiplications, and use a more efficient recursive procedure which generates the new inverse in only $O(k^2)$ multiplications. The technique is based on the partitioned form of an inverse. Suppose that the inverse $(N_k^T N_k)^{-1}$ is known and that ∇h_{j_k} is to be deleted from $N_k = [\nabla h_{j_1}, \nabla h_{j_2}, \ldots, \nabla h_{j_k}]$ and the new inverse $(N_{k-1}^T N_{k-1})^{-1}$ is sought where $N_{k-1} = [\nabla h_{j_1}, \nabla h_{j_2}, \ldots, \nabla h_{j_{k-1}}]$.

Let

$$N_k^T N_k = \begin{bmatrix} A & \vec{z} \\ \vec{z}^T & a \end{bmatrix} \tag{7-35}$$

where

$$A = N_{k-1}^T N_{k-1} \quad . \tag{7-36}$$

The desired inverse of A can be computed from the available submatrices of

$$(N_k^T N_k)^{-1} = \begin{bmatrix} B & \vec{u} \\ \vec{u}^T & b \end{bmatrix} \quad . \tag{7-37}$$

The relationship

$$\begin{bmatrix} B & \vec{u} \\ \vec{u}^T & b \end{bmatrix} \begin{bmatrix} A & \vec{z} \\ \vec{z}^T & a \end{bmatrix} = I \tag{7-38}$$

gives

$$BA + \vec{u}\,\vec{z}^T = I \tag{7-39}$$

$$B\vec{z} + a\,\vec{u} = \vec{0} \tag{7-40}$$

$$\vec{u}^T A + b\,\vec{z}^T = \vec{0} \tag{7-41}$$

$$\vec{u}^T \vec{z} + a\,b = 1 \quad . \tag{7-42}$$

From Eq. (7-39) and (7-41) we get

$$A^{-1} = B + \vec{u}\,\vec{z}^T A^{-1} \tag{7-43}$$

and

$$A^{-1} = B - b^{-1}\,\vec{u}\,\vec{u}^T \quad . \tag{7-44}$$

This procedure can be generalized in the case when a ∇h_{j_ℓ} other than ∇h_{j_k} is dropped from N. It is sufficient to interchange the ℓth and kth row and column of $(N_k^T N_k)^{-1}$ before relationship (7-44) is applied. In a similar way we can obtain a procedure for computing $(N_k^T N_k)^{-1}$ when a column is added to N_{k-1}. We assume the inverse $(N_{k-1}^T N_{k-1})^{-1}$ and ∇h_{j_k} are known. We have

$$(N_k^T N_k) = (N_{k-1}, \nabla h_{j_k})^T (N_{k-1}, \nabla h_{j_k}) = \begin{bmatrix} A & \vec{z} \\ \vec{z}^T & a \end{bmatrix} \tag{7-45}$$

where

$$A = N_{k-1}^T N_{k-1} \quad , \tag{7-46}$$

$$\vec{z} = N_{k-1}^T \nabla h_{j_k} \quad , \tag{7-47}$$

$$a = \nabla h_{j_k}^T \nabla h_{j_k} \quad . \tag{7-48}$$

From Eq. (7-39) and (7-41) we get

$$B = A^{-1} - \vec{u}\,\vec{z}^T A^{-1} = A^{-1} + b\,A^{-1}\vec{z}\,\vec{z}^T A^{-1} + b\,\vec{w}\,\vec{w}^T \tag{7-49}$$

where

$$\vec{w} = A^{-1}\vec{z} \tag{7-50}$$

$$\vec{u} = -\,b\,A^{-1}\vec{z} = -\,b\,\vec{w} \quad . \tag{7-51}$$

Scalar b can be found from Eq. (7-42), i.e.

$$b = a^{-1}(1 - \vec{u}^T\vec{z}) = a^{-1}(1 + b\,\vec{z}^T A^{-1}\vec{z}) \tag{7-52}$$

and

$$b = (a - \vec{z}^T A^{-1}\vec{z})^{-1} = (\nabla h_{j_k}^T\,(I - N_{k-1}(N_{k-1}^T N_{k-1})^{-1} N_{k-1}^T)\,\nabla h_{j_k})^{-1}$$
$$= (\nabla h_{j_k}^T P_{k-1}\nabla h_{j_k})^{-1} = |P_{k-1}\nabla h_{j_k}|^{-2} \quad . \tag{7-53}$$

The last equality holds because

$$P_{k-1}\nabla h_{j_k} = P_{k-1}(P_{k-1}\nabla h_{j_k}) \tag{7-54}$$

which is an obvious property of the projection matrix.

The computational procedure can be summarized as follows.

(i) An auxiliary vector is computed,

$$\vec{w} = (N_{k-1}^T N_{k-1})^{-1} N_{k-1}^T \nabla h_{j_k} \quad ,$$

together with the scalar

$$c = |P_{k-1}\nabla h_{j_k}|^2 \quad .$$

(ii) The segments of the matrix

$$(N_k^T N_k)^{-1} = \begin{bmatrix} B & \vec{u} \\ \vec{u}^T & b \end{bmatrix}$$

are given by the relations

$$b = c^{-1} \quad ,$$

$$\vec{u} = -\,c^{-1}\vec{w} \quad ,$$

$$B = (N_{k-1}^T N_{k-1})^{-1} + c^{-1}\vec{w}\,\vec{w}^T \quad . \tag{7-55}$$

This procedure can also be used recursively to obtain the initial inversion of $(N_k^T N_k)^{-1}$ and P_k from the set of active constraints. An additional advantage of using this recursion is its ability to select the largest set of linearly independent critical constraints from the set of all critical constraints.

It is clear from Eq. (7-55) that $(N_k^T N_k)^{-1}$ cannot be obtained if $P_{k-1}\nabla h_{j_k} = 0$. This equation reveals that ∇h_{j_k} is linearly dependent on the set of the vectors $\nabla h_{j_1}, \nabla h_{j_2}, \ldots \nabla h_{j_{k-1}}$, and should be ignored. Unfortunately the matrix $(N_k^T N_k)$ is frequently very ill-conditioned (with respect to the inverse problem) and, if $(N_k^T N_k)^{-1}$ is computed without special precautions, it may be greatly influenced by round-off errors. This is a well known numerical difficulty which appears in linear least squares problems. It is therefore desirable to compute $(N_k^T N_k)^{-1}$ without forming the numerical product matrix $(N_k^T N)$. To do this, the matrix N_k is decomposed in the following manner,

$$N_k = QR \tag{7-56}$$

where Q is an orthogonal matrix (product of unitary elementary matrices),

$$R = \begin{pmatrix} \hat{R} \\ 0 \end{pmatrix}_{(n-k)\times(k)} \tag{7-57}$$

and $\hat{R}(k\times k)$ is upper triangular. Thus, we have

$$N_k^T N_k = R^T Q^T Q R = R^T R = \hat{R}^T \hat{R} \tag{7-58}$$

which is the Choleski decomposition of $N_k^T N_k$. Now it is easy to compute $(N_k^T N_k)^{-1}$ since $\hat{R}$ is triangular and this inverse can be computed directly from $\hat{R}$. An essential gain is that we do not work with $(N_k^T N_k)$ but with $\hat{R}$ which is better conditioned. There are a number of ways to achieve the decomposition Eq. (7-56) and a very effective one is by using the Householder transformation [7.12], [7.13]. This type of inversion procedure is very important because it usually secures numerical stability (accuracy) in computations, and is highly recommended. Kalfon et al., [7.14], [7.15] implemented such techniques in their version of the gradient projection method and achieved a very stable inversion process.

7.3.4 Problems with Special and Non-linear Constraints

Further simplifications in computing the projection matrix P can be achieved if some of the constraints have a special form $D_i \leqslant$ constant [7.16]. Let us assume that, for example, $h_{j_p} = D_r \leqslant 0$, and let $\vec{e}_r$ be a unit vector which has all components equal to zero except component number r which is 1. If this constraint is critical then

$$N = [\nabla h_{j_1}, \nabla h_{j_2}, \ldots, (\nabla h_{j_p} = \vec{e}_r), \ldots \nabla h_{j_k}] \tag{7-59}$$

and the projection matrix P has the null row and column number r. This property of P follows from the observation that the projected vector $S = PV$ must have $S_r = 0$ for all possible vectors V and that P is symmetric. It can also be shown that the reduced matrix $\hat{P}$ (which is the P matrix without the null row and column) is $\hat{P} = I - \hat{N}(\hat{N}^T \hat{N})^{-1} \hat{N}^T$ where $\hat{N}$ is obtained from N by deleting column p and row r from N. This simplification reduces the size of P thus reducing computer storage required and decreases the computational effort.

In general, the gradient projection method has been found efficient if used for solving problems with linear constraints. There are, however, at least two ways in which this method can handle non-linear constraints. One possibility is via the Fiacco and McCormick transformation where the non-linear constraints are absorbed by the redefined objective function and the linear constraints remain as side restrictions. This transformation reduces the original problem with non-linear constraints to the formulation with linear constraints.

Another technique is to linearize locally the critical non-linear constraints and consider a sequence of approximate problems with the linearized constraints. There are unfortunately at least two reasons why this last technique is not as efficient as it is in cases where all the constraints are linear. A major computational problem is introduced by the fact that we cannot in general, use the recurrence formulas which relate $(N_k^T N_k)^{-1}$ and $(N_{k-1}^T N_{k-1})^{-1}$. When the new solution $\vec{D}_{q+1}$ has been obtained it is very likely that several columns of N_k will have to be replaced by the new linear approximation to the constraints. Thus, the old inverse becomes almost useless and a completely new one has to be computed. This is true even when the new set of critical constraints remains unchanged or differs only by one constraint from the last one. Another difficulty is introduced by the problem of returning back to the convex constraints after a move has been performed along the projected gradient on the intersection of the hypersurfaces tangential to the critical set of constraints. Such a correction move (iterative) to the feasible region may be relatively easy if the steps performed in the infeasible region are short enough. On the other hand this would cause the growth of the total number of steps which are necessary to obtain the solution of the problem. There is therefore an obvious trade-off between the length of step in each iteration and the effort of returning to the feasible region.

7.3.5 Conjugate Gradient Version of the Method for Problems with Linear Constraints

The Rosen projection method can be viewed as the steepest descent method with the ability to handle constraints. It is therefore reasonable to expect that the method may be improved by using the conjugate gradient vectors instead of gradients. Design of such a refinement has been attempted by Goldfarb and Lapidus [7.17] and their method has proved to be quite successful. The capability of the method is limited to cases with linear constraints and its derivation is based on the quadratic objective function

$$M(\vec{D}) = M_o + \vec{B}^T \vec{D} + \tfrac{1}{2} \vec{D}^T A \vec{D} \quad . \tag{7-60}$$

It follows from Eq. (7-60) that

$$\vec{D}_{q+1} - \vec{D}_q = A^{-1} (\nabla M(\vec{D}_{q+1}) - \nabla M(\vec{D}_q)) \tag{7-61}$$

and, if $\vec{D}_{q+1}$ gives a minimum of $M(\vec{D})$ on the cross-section of the hyperplanes h_{j_p} ; $p = 1,2,\ldots k$ then

$$\nabla M(\vec{D}_{q+1}) = N_k \vec{\gamma} \quad , \tag{7-62}$$

where N_k is defined as in Section 7.3.1, and $\vec{\gamma}$ is a vector to be determined.

From Eq. (7-61), (7-62) and the condition that $N_k(\vec{D}_{q+1} - \vec{D}_q) = 0$ we get

$$\vec{D}_{q+1} = \vec{D}_q - \hat{P} A^{-1} \nabla M(\vec{D}_q) \quad , \tag{7-63}$$

where

$$\hat{P} = I - A^{-1} N_k (N_k^T A^{-1} N_k)^{-1} N_k^T \quad . \tag{7-64}$$

Formula (7-63) is an extension of the Newton method, where by using matrix $\hat{P}$ the search for the minimum is restricted to the feasible region defined by the linear constraints of the problem. Due to the well known disadvantages of the Newton method it is preferred to implement the conjugate directions method of Davidon. This idea leads to a version of the variable metric method (Davidon) which is capable of optimizing a non-linear function subject to the linear constraints. The method uses positive definite matrices H_q which approximate $-\hat{P} A^{-1}$ and are updated whenever a hyperplane is added or dropped from the constraints. In addition the matrices H_q are modified as in the unconstrained version of Davidon's method and this modification is applied if the minimum of $M(\vec{D})$ is found along $\vec{S}_q = - H_q \nabla M(\vec{D}_q)$ before a new constraining hyperplane is reached.

The same method can be used if the objective function is non-linear and non-quadratic. This is motivated by the assumption that in the neighborhood of the solution the non-linear function can be adequately approximated by a positive definite quadratic form.

7.3.6 Summary of the Gradient Projection Method

In the gradient projection method, the linear optimization subproblems are replaced by matrix inversion schemes. These schemes have to be able to handle the ill-conditioned matrices N_k via special decomposition techniques. The method is computationally efficient if all the constraints in the problem are linear and becomes less practical if non-linear constraints are involved. There have been, however, reported successful applications of the method to structural optimization problems with non-linear constraints [7.18].

The method has the advantage of being able to deal with nonconvex constraints. The disadvantages include: rather complex computer code, computational difficulties in inverting $(N_k^T N_k)$ and the expensive process of correcting iterations back to the feasible region if problems involve non-linear constraints.

7.4 Gellatly's Optimum Vector Method

7.4.1 Concept of the Method

Gellatly [7.3] has suggested a feasible direction method where the direction of search is determined from a set of simultaneous linear equations. First note that the direction vector $\vec{S}_q$ can be expressed by a linear combination of the gradients of the objective function and critical constraints at the current iteration point $\vec{D}_q$,

$$\vec{S}_q = - \alpha \nabla M(\vec{D}_q) + \sum_{j \in J_c} \beta_j \nabla h_j(\vec{D}_q) \quad , \tag{7-65}$$

where J_c is the set of critical constraints. Gellatly distinguishes between the two components of Eq. (7-65) which correspond to the two types of travel modes; the steepest descent and side-step mode. In the first case we have all $\beta_j = 0$ and $\vec{S}_q = -\alpha \nabla M(\vec{D}_q)$. In the second we demand that

$$(\vec{S}_q)^T \nabla M(\vec{D}_q) = 0 \quad , \tag{7-66}$$

$$(\vec{S}_q)^T \nabla h_j(\vec{D}_q) + \varepsilon_j = 0 \quad , \quad j \in J_c \tag{7-67}$$

where the ε_j are some preset positive constants. If $M(\vec{D})$ is a linear function then any vector satisfying Eq. (7-66), (7-67) is feasible and a step can be taken along $\vec{S}_q$ which holds the value of the objective function constant. System (7-66), (7-67) is symmetric, positive definite as is easily demonstrated when a more uniform notation is introduced. Let $\alpha = \beta_o$, $\nabla M = \nabla h_o$, $\vec{\varepsilon} = [0, -\varepsilon_1, -\varepsilon_2, \ldots, -\varepsilon_m]$ and $j \in J_c$ if $j = 1,2,\ldots m.$, We also define the matrix $H = [\nabla h_o, \nabla h_1, \ldots, \nabla h_m]$. From Eq. (7-66), (7-67) it follows that

$$\left(\sum_{i=0}^{m} \beta_i \nabla h_i^T \right) \nabla h_j = -\varepsilon_j \quad , \quad j = 0,1,2,\ldots m \tag{7-68}$$

and in matrix form

$$H^T H \vec{\beta} = -\vec{\varepsilon} \quad . \tag{7-69}$$

Linear set (7-68) can be considered as being a side condition of an optimization subproblem (as in Zoutendijk's method) or it can be solved for some fixed values of ε_j. Gellatly takes the latter approach and selects arbitrarily the unit values for ε_j, $j > 0$. In the more general case when $M(\vec{D})$ is a non-linear function, conditions (7-66), (7-67) are not sufficient for determining a usable-feasible direction and the problem has to be reformulated if the same method is to be used.

In order to obtain an equivalent problem with a linear objective function an additional variable is introduced which replaces the objective function. The modified optimization problem becomes:

$$\text{Min } D_{n+1}$$

subject to the original constraints and in addition

$$M(\vec{D}) - D_{n+1} \leqslant 0 \quad . \tag{7-70}$$

With this modification the method of Gellatly can be used without any substantial changes except that the first equation of the set (7-66), (7-67) drops out from the set. Due to the particular formulation of the new objective function, the steepest descent step can be obtained simply by reducing D_{n+1}. In the side-step the variable D_{n+1} is kept constant but the non-linear weight function $M(\vec{D})$ may change.

7.4.2 <u>Computational Problems</u>

Some comments should be made on the solvability of the linear set of equations (7-66), (7-67) which determines the direction $\vec{S}$. There are three cases where the coefficient matrix of this set becomes singular (or nearly singular) and special actions must be taken to circumvent this difficulty.

The most obvious case of singularity occurs when the number of vectors ∇M, ∇h_j, $j \in J_c$ exceeds the dimension of the multivariable space n, so that these vectors cannot be linearly independent and consequently α, β_j are not uniquely defined. A similar difficulty occurs when there is a linear dependence between some of the vectors ∇M, ∇h_j, $j \in J_c$ (whose total number can be less than n). Finally, the system matrix also becomes singular when the optimum solution is reached where $-\nabla M$ becomes a nonnegative linear combination of the gradients to the active constraints (Kuhn-Tucker optimum condition).

A straightforward procedure can be used to remove any linearly dependent equation from the system, Eq. (7-69), which is solved by Cholesky decomposition. During the decomposition process we obtain a zero on the main diagonal of the triangular matrix due to the dependence of the linear equations. The first zero appears in the row corresponding to the first dependent equation. To remove this equation the complete row and corresponding column is set to zero (including ε_i) with the exception of the main diagonal element where the unit value is inserted. This operation results in computing $\beta_i = 0$ for the corresponding linearly dependent vector which eliminates this vector from Eq. (7-65). It is, furthermore, necessary to detect the case when the linear dependence is caused by optimality. To do this we check the vector products, Eq. (7-67), after determining the solution. If the products are negative we have a feasible $\vec{S}_q$ and optimization is continued. If, however, some of them do not satisfy this condition we assume that the optimal solution has been reached and the computation is terminated.

7.4.3 Summary of the Optimum Vector Method

In the Optimum Vector Method the feasible direction finding problem is reduced to the solution of linear equations. Since these equations involve the positive definite matrix $H^T H$ the efficient and stable Choleski decomposition method can be used to solve them. We may, however, expect numerical difficulties if H is not well-conditioned unless special techniques are used to decompose $H^T H$ (see Section 7.3.3). Another feature of the method which we should consider as being disadvantageous is the arbitrary choice of the $\vec{\varepsilon}$-vector. This method has the ability to handle nonconvex problems.

7.5 Conclusion

Table 1 summarizes briefly some of the important features of the methods discussed in this Chapter. It has to be pointed out that the methods have not yet been compared by numerical experimentation.

Table 1

	Feasible Direction Method	Gradient Projection Method	Optimum Vector Method
Feasible direction subproblem	Linear or quadratic programming	Matrix inversion and updating	Solution of linear equations
Efficient for problems with non-linear constraints	yes	no	yes
Ability to handle nonconvex problems	yes	yes	yes
Unstable numerical process involved	no	yes	yes
Generates strictly feasible directions	yes	no (nonlin. constr.)	yes
Simplicity of computer code	no	no	no
Successful applications to structural optimization problems	yes large size	yes small size	yes large size

List of References

Ref.

7.1 Zoutendijk, G., *Methods of Feasible Directions,* Elsevier Publishing Co., Amsterdam, 1960

7.2 Rosen, J. B., "The Gradient Projection Method for Nonlinear Programming, Part I, Linear Constraints," *Journal SIAM,* Vol.8, March 1960, pp.181-217

7.3 Gellatly, R. A., "Development of Procedures for Large Scale Automated Minimum Weight Structural Design," AFFDL-TR-66-180, December 1966

7.4 Colville, A. R., "A Comparative Study on Nonlinear Programming Codes," IBM Technical Report No. 320-2949, 1968

7.5 Zoutendijk, G., "Nonlinear Programming: A Numerical Survey," *SIAM Journal on Control,* Vol.4, February 1966, pp.194-210

7.6 Kowalik, J. and Osborne, M. R., *Methods for Unconstrained Optimization Problems,* American Elsevier Publishing Co., Inc., New York, 1968

7.7 Wolfe, P., "On the Convergence of Gradient Methods under Constraints," IBM Research Paper RZ-204, Zurich, Switzerland, 1966

7.8 Zangwill, W. I., "A Decomposable Nonlinear Programming Approach," Operations Research, Vol.15, November-December 1967, pp.1068-1087

7.9 Bendsen, R. L. and D'Hondt, I. W., "A Modified Application of a Method of Feasible Directions to the Extremum Problems," The Boeing Company, Technical Note No.D6-19340 TN, 1966

7.10 Karnes, R. M. and Tocher, J. L., "Automatic Design of Optimum Hole Reinforcement," The Boeing Company, Technical Report No.D6-23359, May 1968.

7.11 Kelley, J. E., Jr., "The Cutting Plane Method for Solving Convex Programs," *Journal SIAM,* Vol.8, December 1960, pp.703-712

7.12 Businger, P. and Golub, G. H., "Linear Least Squares Solutions by Householder Transformations," *Numerische Mathematik,* Vol.7, May 1965, pp.269-276

7.13 Golub, G. H., "Numerical Methods for Solving Linear Least Squares Problems," *Numerische Mathematik,* Vol.7, May 1965, pp.206-216

7.14 Kalfon, P., Ribieve, G., and Sogno, J. C., "A Method of Feasible Directions Using Projection Operators," Proceedings IFIP Congress 1968, Math. Booklet A, p.123

7.15 Kalfon, P., Ribieve, G., and Sogno, J. C., "Methode du Gradient Projete Utilisant la Triangularisation Unitaire," Publication No.FT/11.3.8/AI Centre National de la Recherche Scientifique, Institute Blaise Pascal, 1968

7.16 Fiacco, A. V. and McCormick, C. P., *Nonlinear Programming: Sequential Unconstrained Optimization Techniques,* John Wiley, New York, 1968

7.17 Goldfarb, D. and Lapidus, L., "Conjugate Gradient Method for Nonlinear Programming Problems with Linear Constraints," *Industrial and Engineering Chemistry Fundamentals,* Vol.7, February 1968, pp.142-151

7.18 Brown, D. M. and Ang, A. H. S., "Structural Optimization by Nonlinear Programming," J. of the Structural Division, *ASCE,* Vol.92, No.ST6, December 1966, pp.319-340

SECTION III

SAMPLE APPLICATIONS

Chapter 8

COMPUTER PROGRAMS FOR THE OPTIMUM DESIGN OF COMPLEX ELASTIC STRUCTURES

by

G. G. Pope

8.1 Introduction

This Chapter describes a number of computer programs which have been developed for the optimum design of idealised aerospace structures of arbitrary geometry, and which include not only optimization algorithms but also segments for the efficient finite element analysis of structures of this class. These programs are concerned mainly with the choice of member cross-sectional areas and thicknesses, but some of them include facilities which, in principle, permit the lengths and spacings of members to be varied within a prescribed topology.

Early direct applications of finite element methods to the design of efficient structures concentrated on the generation of fully-stressed designs in which every member is either fully-stressed under at least one of the applied loadings or has a minimum permissible cross-section or thickness. Such designs, which usually approximate to or coincide with a least weight design in applications where no constraints are imposed on the displacements, can normally be deduced iteratively by repeatedly modifying members on the basis of the local stress level, and by re-analysing the resulting structures. The computations involved in this process are relatively short compared with those usually associated with a rigorous search for a least weight design, although the efficiency of techniques for the computation of the latter is continually being improved (see, for example, Section 8.4). The fully-stressed design approach continues to find useful applications and some relevant recent developments are described in Section 8.5. It also provides a means of generating useful initial trial designs for a more general class of optimization problems.

The main portion of this Chapter is concerned with more rigorous optimization procedures. Section 8.2 describes computer programs developed at the Bell Aerosystems Company as the culmination of the first major exercise in the application of mathematical programming techniques to the design of complex structural components, and Sections 8.3 and 8.4 describe major subsequent contributions from the Boeing Company and the Philco-Ford Corporation.

8.2 Bell/AFFDL Programs for the Least Weight Design of Stressed-Skin Structures

The Bell Aerosystems Company, working under contract to the U.S. Air Force Flight Dynamics Laboratory, has developed several computer programs [8.1], [8.2], [8.3] for the least weight design of stressed-skin structures of arbitrary geometry. Two of these programs are described in this Section. Both are written for use on IBM 7090/7094 computers or equivalent machines with a core store of 32K words. The first is directly applicable only in situations where the basic configuration of the structure is fixed, where the design variables consist solely of skin thicknesses and member cross-sectional areas, and where consequently the merit function is linear. Structural dimensions within a prescribed topology may be treated as variables in the second program described, which also permits the study of larger problems of fixed geometry; this more powerful program is, however, less efficient in applications where either program could be used.

8.2.1 Analysis Procedure

Both programs employ the finite element displacement method for analysis purposes and include as a basic facility the following types of element: axially-loaded bar, shear web, quadrilateral shear panel, triangular region in plane stress, quadrilateral panel in plane stress. Displacements are assumed to vary linearly along the edges of all these elements. The bar elements have uniform cross-sectional areas and the plane elements are of uniform thickness. The modular form of the programs enables additional elements to be added with a minimum of modification. An option is included to take account of the symmetry of lifting surfaces of symmetric cross-section. Several independent load conditions may be considered and temperature variations may be prescribed over the structure to correspond to the load conditions. Buckling effects are not included but restraints may be imposed on the amplitudes of the displacement components and on the minimum permissible values of the design variables. The analysis segments have a nominal capacity of 200 discrete elements and 170 degrees of freedom in the fixed geometry program; the nominal capacity of the corresponding segments in the larger varying geometry program is 600 discrete elements and 450 degrees of freedom. The size of idealisation which can be handled in practice by either program depends on the detailed specification of the problem under consideration and is influenced by such factors as the bandwidth of the non-zero terms in the stiffness matrix. The Choleski method is used to solve the analysis equations, storing the intermediate triangular matrix in the computer core locations previously occupied by the stiffness matrix.

8.2.2 Optimization Procedure employed in the Fixed Geometry Program

In the fixed geometry program where the weight is, by definition, a linear function of the design variables, the optimum design is sought by a direct application of Gellatly's optimum vector method which is described in detail in Chapter 7. Starting from a known feasible design, a search is first made along a 'steepest descent' path in design space, normal to the planes representing structures of equal weight, to find a design in which one constraint at least is active. At this constrained design a direction of search is selected within the relevant constant weight plane, pointing into the feasible region and away from the current critical constraints. A further constrained design is found following this direction of search, and a design midway between the two constrained designs of the same weight is used as a starting point for a repetition of the whole process.

The following procedure is adopted in this program to find the appropriate distance of travel along each path in design space. First the structure is re-analysed after the design has been modified by a specified amount. The step length between successive designs is then doubled as many times as is necessary to achieve a design which is not feasible. This design is then modified by an increment which is half that used in the final step of the preceding process but which is of opposite sign. The step length between successive designs is then halved repeatedly, the sign of each step being always chosen to be such that the direction of travel is towards the edge of the feasible region. The process is continued until a constrained design is obtained to the required accuracy.

In order to select a suitable direction of search in a typical constant weight plane it is necessary to evaluate at the relevant starting point the partial derivatives with respect to each of the design variables of the stress and displacement components which are subject to active constraints. These derivatives are calculated directly from an analytical expression with significantly less effort than would be involved in the application of first order difference techniques in which the structure is re-analysed for small changes in each of the design variables in turn (see also Fox [8.4]).

Large savings in computing time can often be achieved by generating iteratively a design which is approximately fully stressed, before entering the above search procedure. The program includes facilities for the automatic generation of such designs.

8.2.3 Optimization Procedure employed in the Varying Geometry Program

In this more powerful program where the weight is a non-linear function of the design variables, the optimization problem is reformulated in terms of a linear merit function by introducing an additional variable and an additional constraint, i.e. the basic problem

$$\text{minimize} \quad M(\vec{D}) \quad \text{subject to}$$

$$h_j(\vec{D}) \leqslant 0 \qquad j = 1,2,\ldots J$$

is replaced by

$$\text{minimize} \quad A \quad \text{subject to}$$

$$h_j(\vec{D}) \leqslant 0$$

and

$$M(\vec{D}) - A \leqslant 0 \quad .$$

With this reformulation the steepest descent searches in the Gellatly optimum vector method become trivial and the computational task is then concentrated in the searches conducted at constant values of A; it should be noted that a constant value of this variable does not correspond to a constant structural weight. The required partial derivatives of the actively constrained stress and displacement variables and of the weight function M with respect to the design variables are calculated by a first order difference procedure, as it did not prove practicable to adapt the analytical procedure used in the fixed geometry program. Facilities are included for the generation of fully-stressed designs when the structural geometry is specified.

8.2.4 Applications

A number of applications of the fixed geometry program have been reported. These include a re-sizing of the members of the idealised fin of the Bell X-22A ducted fan VTOL aircraft [8.3], [8.5]. This application involved 141 degrees of freedom and 136 design variables; multiple load conditions were specified and both strength and stiffness requirements had to be satisfied. A weight saving of the order of 35% relative to the idealised structure of the actual fin was obtained at a computing cost of less than 500 dollars. Another interesting application has been to the design of the horizontal stabiliser of a supersonic aircraft [8.6]. Here the avoidance of binary flutter contributed an active constraint which was represented approximately by a limitation on the ratio of the overall flexural and torsional rigidities; the program was modified to incorporate constraints of this type. Applications of the varying geometry program reported so far have been limited in the main to pin-jointed trusses of relatively simple geometry.

8.3 Boeing Program for the Least Weight Design of Stressed-Skin Structures

Karnes and Tocher [8.7] describe a computer program which they have developed at the Boeing Company to search for the least weight design of stressed-skin structures, with emphasis on regions containing holes and cut-outs, in circumstances where buckling effects can be neglected. The program permits the design to be influenced by a number of independent load conditions and also enables the user to specify limitations on the maximum and minimum permissible thicknesses.

8.3.1 Analysis Procedure

The sheet is idealised as an assembly of triangular membrane elements, each of which is assumed to be in a state of uniform strain, and corresponding flanges which can carry axial loads only. The thicknesses of the individual membrane elements and the cross-sectional areas of the individual flange elements are prescribed uniform. An efficient routine, in which all the non-zero elements required to specify the stiffness matrix are carried simultaneously in the core store, is used to analyse the idealised structure by the direct stiffness (displacement) method. The equilibrium equations are solved iteratively using block over-relaxation. Such iterative techniques often prove very efficient in optimization problems since the changes in the design parameters between analyses are usually relatively small; thus the displacements before a typical redesign are usually a good first approximation to those after the redesign has taken place.

The number of design variables can easily become very large when finite element idealisations are used in optimization studies. Karnes and Tocher therefore express the distribution of sheet thickness in terms of the thicknesses of a limited number of elements only; the program defines the thicknesses of the remaining elements automatically by a linear interpolation technique. It has been demonstrated that the intelligent use of this approach leads to a dramatic reduction in problem size without materially influencing the optimum design.

8.3.2 Optimization Procedure

The optimization problem is solved by a version of Zoutendijk's method of feasible directions, following the procedure outlined in Section 7.2.4. A known feasible design is used as a starting point and a single search is made in a direction of steepest descent to find a feasible design in which at least one constraint is active; this procedure is of course unnecessary if the initial design is itself of this type. The constrained design is used as a starting point in a search for a lighter constrained design in a direction established by solving the linear sub-problem which is formulated in Eq. (7-9) to (7-15). The latter process is then repeated starting each time from the lighter constrained design obtained in the preceding application, until the least weight design has been found to an acceptable standard of accuracy. The setting up of each ancilliary sub-problem involves the choice on the basis of experience of a set of constants denoted by C_j in Chapter 7. Using a slightly different formulation, Karnes and Tocher choose these constants, in effect, to be equal; the actual value is selected on a basis of experience to prevent rapid changes in the direction of search (zigzagging) and excessively small steps along the boundary of the feasible region. The following procedure is adopted to establish the distance of travel along each search path:

(1) Assuming that the partial derivatives of the design variables with respect to relevant stress and displacement components are constant, an estimate is made of the changes in the design variables necessary to reach a lighter constrained design.

(2) The modified design is analysed and the lighter constrained design is computed more accurately by linear interpolation (or extrapolation) between the modified design and the previous critical design.

(3) This interpolated design is analysed and if it does not represent the critical design to an acceptable standard of accuracy it is used together with the two preceding designs to obtain a better approximation by parabolic interpolation.

(4) The parabolic interpolation procedure is repeated if necessary, using each time the three most recently analysed designs, until a constrained design is obtained to a specified standard of accuracy.

The same interpolation technique is employed in obtaining the initial critical design along a steepest descent path, once a design outside the feasible region has been obtained by a simple step-doubling process.

The partial derivatives of the design variables with respect to the stress and displacement components subject to active constraints are calculated in this program by a first order difference procedure which involves re-analysis of the structure for small changes in each of the design variables in turn. The user specifies the amplitude of design modification which is likely to lead to a significant variation in these derivatives; when the design modifications are below this level the derivatives are assumed constant in the interest of computational efficiency.

8.3.3 Application

The Boeing program is written for the CDC 6600 computer and permits the employment of up to 100 design variables and 700 degrees of freedom. It has been used to study possible improvements in the design of a window panel for the 747 aircraft. An initial application to the whole panel, which includes three windows, employed a finite element idealisation involving 600 elements and 300 nodes, under five independent loadings. A design was produced after a computing time of 3½ hours which was lighter than any that had previously been generated by hand. In a second stage a more detailed application was made to the local region between adjacent windows; a finer grid was employed involving 144 nodes and 267 finite elements. It only proved necessary to consider two load conditions, and the running time on the CDC 6600 was 45 minutes. The configuration obtained in this way was 10% lighter than the best hand-generated design based on the first stage of the optimization study.

8.4 Approximate Multiple Configuration Analysis and Allocation Procedure (Philco-Ford/AFFDL)

Melosh and Luik [8.8], [8.9], working at the Philco-Ford Corporation under a contract from the U.S. Air Force Flight Dynamics Laboratory, have developed a technique for the design of least weight structures which has proved very efficient in a number of trial examples and which is particularly well suited to applications where the design variables can take a series of discrete values only. The current implementation is limited to pin-jointed trusses, but stressed-skin structures have been optimized with its aid, using the Hrennikoff analogy [8.10] to deduce an equivalent framework. Stress limitations are the only constraints considered, and the design variables consist solely of the cross-sectional areas of the members; variations in geometry have, however, been included in one application where it proved possible, without imposing specious strain restraints, to incorporate a sufficient number of members in the initial idealisation to include to an adequate degree of accuracy, any member which might be present in the optimum design. The computer program, which is written for the Philco 212 computer, is capable of handling a maximum of 1000 truss elements, 1000 sizing variables, 450 degrees of freedom, and up to five independent load conditions.

8.4.1 Analysis Procedure

The search technique is made practicable by the use of an efficient approximate procedure to estimate, without repeating the analysis of the entire structure, the influence on the internal force system of a change in a single design variable. The effect of such a modification is estimated by a complementary energy analysis in which three force systems only are considered, namely:

(1) the internal forces in the structure before any modifications were incorporated,

(2) the self-equilibrating system obtained by subtracting the above system from the internal force system immediately prior to the modification under consideration,

(3) a self-equilibrating system corresponding to a self-straining of the member to be modified.

If the above procedure is applied repeatedly with self-straining of each of the members in turn, but without any design modifications, it can easily be seen that an exact analysis will be obtained of the idealised structure.

8.4.2 Optimization Procedure

A series of permissible discrete values is assigned to each design variable. A typical variable is then decreased tentatively from its value in an initial feasible design to the next permissible smaller value, and the structure is re-analysed approximately by the above technique to see whether any stress constraints are violated. The design change is rejected immediately if the modified member is over-stressed; if the stress in this member remains within the permitted range but the stress limit is exceeded in another member or members, a trade-off calculation is performed to see whether any weight saving is achieved if the critical members are appropriately re-sized. Tentative decreases are made in all the design variables in turn, and the procedure is repeated until no significant modification results from a cycle involving attempted changes in all the design variables.

8.4.3 Applications

Melosh and Luik [8.8], [8.9] describe the application of the above procedure to a number of design problems and show that it is comparable in efficiency with an iteration to a fully-stressed design when the latter is relevant. They also show that the efficiency of their computer program compares favourably in several applications with existing programs based on more conventional non-linear programming techniques.

8.5 Application of Iterative Procedures for the Generation of Fully-Stressed and Similar Designs

8.5.1 Contributions of the Grumman Aircraft Corporation

Some investigations have been conducted at the Grumman Aircraft Corporation into practical techniques for the generation of fully-stressed designs in the airframe context [8.11], [8.12]. A number of structural configurations typical of aircraft lifting surfaces have been studied and fully-stressed designs have been obtained. The conventional displacement method was employed for analysis purposes and the average equivalent stress in each structural panel was used in the initial study as a basis for factoring the thickness after each iteration. Since many of the panels were relatively large from an analysis viewpoint, individual panels sometimes included significant variations in stress. Consequently it was found that designs evolved by straightforward iteration sometimes involved erratic thickness variations between individual elements which no designer would accept.

Recognising that this difficulty arose because average panel stresses were employed in the iteration rather than the peak stresses which are likely to occur, for example, in regions of load diffusion, the Grumman investigators re-interpreted the results of the individual displacement method analyses in a format typical of the force method, by re-idealising the structure as an assembly of flange elements with linearly varying end load, and panels in a state of pure shear. Members were subsequently re-sized using the results in this form, direct stresses at the panel corners being deduced from the loads in the adjacent flanges. It was found that more satisfactory fully-stressed designs were obtained in this manner which were of virtually the same weight as those derived by the more direct approach. This reintroduction of a force method idealisation does, of course, complicate the programming of the redesign procedure and simpler techniques might produce an equivalent improvement. This idealisation is, however, valued in its own right by designers who need to interpret the results of overall structural analyses in the context of the design of structural details and an automated sequence of computer programs has been developed for its use in this way in the generation of fully-stressed designs.

Lansing et al. [8.12] have recently adapted this kind of approach to the design of structures in fibre-reinforced composite materials. Such structures are usually fabricated from layers of unidirectionally-reinforced material which each have a prescribed thickness and volume fraction in their cured state. Each skin thickness parameter associated with design in isotropic materials is replaced therefore by the numbers of layers of composite with fibres orientated in each of the prescribed directions; a free variation of fibre direction is usually impracticable from the fabrication viewpoint. In this Grumman procedure the structure is first analysed with assumed values for the design variables and the results are interpreted using a force method idealisation as described above. Stress fields which may be critical are identified in each composite panel, and with the aid of these a rigorous optimum lay-up is calculated for the panel, allowing for practical restrictions on thickness and fibre orientation; elements in conventional materials are re-sized in the customary manner. The structure is then re-analysed and the process is repeated iteratively until no significant change in weight occurs between successive cycles.

A successful trial application to composite construction has been made in the design of a horizontal stabiliser for a supersonic aircraft. Boron epoxy composite was selected as the skin material, supported by full-depth aluminium alloy honeycomb; other internal structure and attachments were designed in titanium alloy. The boron fibres were permitted to lie in four directions, i.e. at 0°, 90° and $\pm 45^{\circ}$ to a datum direction. The structural idealisation which took account of the symmetry of the structure and of the loading about the mid surface, employed approximately 1000 structural elements and 1100 degrees of freedom; four independent load conditions were considered. Starting from arbitrary but intelligently chosen member sizes, the structure was redesigned five times by an automated version of the above procedure; it was found that the structure weight was sensibly constant after the second redesign.

8.5.2 Generation of Structures with Uniform Strain Energy Density

An alternative semi-intuitive method for the generation of near optimum designs, which has been developed by Venkayya et al. [8.13], is closely related to the fully-stressed design procedure and in some applications is, in effect, identical to it. This method is based on the hypothesis that the strain energy density is uniform throughout a least weight structure designed to withstand a single load system when instability constraints are inactive and displacements are unrestrained. If more than one loading is involved, the strain energy due to each is evaluated in turn and the maximum value of the strain energy density is found at every point in the structure. It is then postulated that the least weight design is one in which the maximum strain energy density is uniform.

When displacement constraints are active, a uniform maximum strain energy design is obtained first by the above procedure and the member sizes (e.g. cross-sections in the case of a pin-jointed truss) are factored up, if necessary, so that none of the critical displacement components exceed their permissible amplitudes by more than about 20%. The first order sensitivity of the various restrained displacements to unit changes in the volumes of the individual members is then calculated and the increases in member sizes proportional to these sensitivities are derived which would be necessary to satisfy each displacement constraint in turn; whenever an individual sensitivity is such that an increase in volume results in an increase in the critical displacement, the size of the member concerned is held constant. The increases in the individual member sizes required to satisfy the various displacement constraints are compared, and the structure is modified on the basis of the largest values, resulting in a feasible design in which the displacement constraints are not necessarily critical. The re-sizing procedures are repeated using starting points each time based on the results of the proceding applications, as described in Venkayya et al. [8.13], until no further reduction occurs in the structure weight.

A computer program for an IBM 7094-II-7044 DCS has been prepared for the implementation of the above process in the context of pin-jointed trusses. The largest applications reported have been to a geodesic dome (61 nodes, 132 bars, 4 load conditions) and a plane truss involving 77 nodes, 200 bars and 5 load conditions; active displacement constraints were present in both these examples. Of particular interest is an application to the design of a ten node twenty-five bar transmission tower under two independent loadings, with upper bounds imposed on all the displacements. This design problem had been studied previously by Fox and Schmit [8.14] and by Gellatly [8.3]. Venkayya et al. obtained, after a computing time of 24 seconds, a structure of virtually identical weight to the least weight design obtained by Gellatly; the latter employed the fixed geometry program described in Section 8.2 with a computing time of 20 minutes on an IBM 7090. Both Vankayya et al. and Gellatly have indicated improvements that might be incorporated in their programs to improve efficiency; the above computing times are, however, convincing evidence of the effectiveness of the Venkayya approach in this application.

Acknowledgement - This Chapter is British Crown Copyright reproduced with the permission of the Controller, Her Majesty's Stationery Office.

List of References

Ref.

8.1 Gellatly, R. A., Gallagher, R. H. and Luberacki, W. A. 'Development of a Procedure for Automated Synthesis of Minimum Weight Structures,' USAF, FDL-TDR-64-141, October 1964

8.2 Gellatly, R. A. and Gallagher, R. H., 'A Procedure for Automated Minimum Weight Structural Design, Part I - Theoretical Basis, Part II - Applications,' Aeronautical Quarterly, Vol.17, No.3, August 1966, pp.216-230 and No.4, November 1966, pp.332-342

8.3 Gellatly, R. A., 'Development of Procedures for Large Scale Automated Minimum Weight Structural Design,' USAF, AFFDL-TR-66-180, December 1966

8.4 Fox, R. L., 'Constraint Surface Normals for Structural Synthesis Techniques,' AIAA Journal, Vol.3, No.8, 1965, pp.1516-1517

8.5 Gellatly, R. A., 'The Role of Optimisation in the Design of Aircraft Structures,' Proc. AGARD Symposium on Structural Optimisation, Istanbul, October 1969, AGARD-CP-36-70

8.6 Johnson, J. R. and Warren, D. S., 'Structural Optimization of a Supersonic Stabilizer,' Proc. AGARD Symposium on Structural Optimisation, October 1969, AGARD-CP-36-70

8.7 Karnes, R. N. and Tocher, J. L., 'Automatic Design of Optimum Hole Reinforcement,' Boeing Report D6-23359, June 1968

8.8 Melosh, R. J. and Luik, R., 'Approximate Multiple Configuration Analysis and Allocation for Least Weight Structural Design,' USAF, AFFDL-TR-67-59, April 1967

8.9 Luik, R. and Melosh, R. J., 'An Allocation Procedure for Structural Designs,' AIAA Paper 68-329, April 1968

8.10 Hrennikoff, A., 'Solution of Problems in Elasticity by the Framework Method,' J. Appl. Mech. Vol.8, December 1941, pp.A169-A175

8.11 Dwyer, W., Rosenbaum, J., Shulman, M. and Pardo, H., 'Fully-Stressed Design of Airframe Redundant Structures,' Proc. of the 2nd Conference on Matrix Methods in Structural Mechanics, WPAFB, October 1968, AFFDL-TR-68-150, 1968, pp.155-181

8.12 Lansing, W., Dwyer, W., Emerton, R. and Ranalli, E., 'Application of Fully-Stressed Design Procedures to Wing and Empennage Structures,' Proc. AIAA/ASME 11th Structures, Structural Dynamics and Materials Conference, Denver, April 1970, pp.97-111

8.13 Venkayya, V. B., Khot, N. S. and Reddy, V. S., 'Optimization of Structures based on the Study of Strain Energy Distribution,' Proc. of the 2nd Conference on Matrix Methods in Structural Mechanics, WPAFB, October 1968, AFFDL-TR-68-150, 1968, pp.111-153

8.14 Fox, R. L. and Schmit, L. A., An Integrated Approach to Synthesis and Analysis. Summer Course on Structural Synthesis, Case Institute of Technology, July 1965

Chapter 9

SPECIAL PURPOSE APPLICATIONS

by

L. A. Schmit

9.1 Introduction

The previous Chapter describes some general purpose structural optimization capabilities for relatively large scale systems. In this Chapter, a few examples of mathematical programming applications to specific structural design problems are described.

It is suggested that the cost of developing a special purpose structural optimization capability may be justified when a particular design problem can be identified as fundamental and recurring. Problems in this category often require complicated failure mode analyses. When developing a special purpose structural optimization capability, it is possible to carefully tailor the analysis and optimization scheme together. Exploitation of physical insight with respect to the analysis and familiarity with the characteristics of the various mathematical programming formulations and the associated algorithmic tools, facilitate the development of tractable optimization capabilities based upon careful and detailed failure mode analyses. The examples to be discussed point up the important role structural optimization can play in evaluating alternative design concepts and materials based upon a comparison of optima. In Section 9.2, the stiffened cylindrical shell optimization capability reported in [9.1] is reviewed in some detail. The extension of this capability to shells with slight meridional curvature [9.2] is briefly discussed and two recently reported special purpose applications to fiber composite structures are noted [9.3], [9.4]. In Section 9.3 application of an integrated penalty function approach (see Figs.2.10 and 2.11) to the optimum design of an ablating composite type heat shield [9.5] is described.

9.2 Integrally Stiffened Cylindrical Shell Example

The frequent occurrence of stiffened cylindrical shell configurations in aerospace structural applications is well known. This example represents a state-of-the-art special purpose application of mathematical programming in structural design optimization as of 1968.

9.2.1 Problem Statement

Consider an integrally stiffened cylindrical shell of radius R and length L such as that shown in Fig.9.1. The stiffeners are assumed to be integral and of rectangular cross section. There are two sets of stiffeners, one in the longitudinal direction and one in the circumferential direction. Each set of stiffeners may be entirely inside or entirely outside the shell. The radius R of the shell wall middle surface, the total length L, and the material properties of the skin and stiffeners are preassigned parameters. It should be noted that the influence of a different but uniform structural temperature (in each of several load conditions) can be introduced by preassigning different values to the material properties in each load condition.

Seven design variables (see Fig.9.2) are dealt with by the optimization procedure namely: (1) the skin thickness t_s, (2) the thickness of the longitudinal stiffeners (t_x), (3) the thickness of the circumferential stiffeners (t_ϕ), (4) the depth of the longitudinal stiffeners (d_x)*, (5) the depth of the circumferential stiffeners (d_ϕ), (6) spacing of the circumferential stiffeners (ℓ_x) and (7) spacing of the longitudinal stiffeners (ℓ_ϕ). Any particular design is represented by a point in the design space located by a vector $\vec{D}$ such that

$$\vec{D}^T = [t_s, t_x, t_\phi, d_x, d_\phi, \ell_x, \ell_\phi] \quad . \tag{9-1}$$

The option to preassign any subset of design variables is available and the stiffener depths may optionally be linked as follows

$$d_x = d_\phi \tag{9-2}$$

which in effect requires that the stiffeners be flush and on the same side of the shell wall.

Side constraints on the design variables limiting the range of admissible values and insuring geometric realizability are considered. The upper bounds on the design variables $D_j \leqslant U_j$: $j = 1,2,\ldots 7$ are expressed in the following normalized form

$$h_j(\vec{D}) = \frac{D_j - U_j}{U_j - L_j} \leqslant 0 \quad ; \quad j = 1,2,\ldots 7 \tag{9-3}$$

---

*Note that the stiffener depth is taken positive for internal stiffening and negative values of d_x and d_ϕ denote external stiffening.

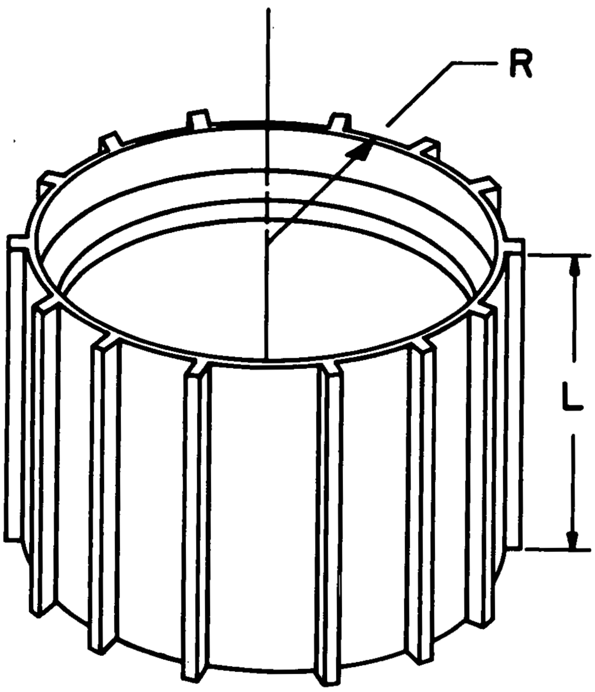


Fig.9.1 Integrally Stiffened Cylindrical Shell

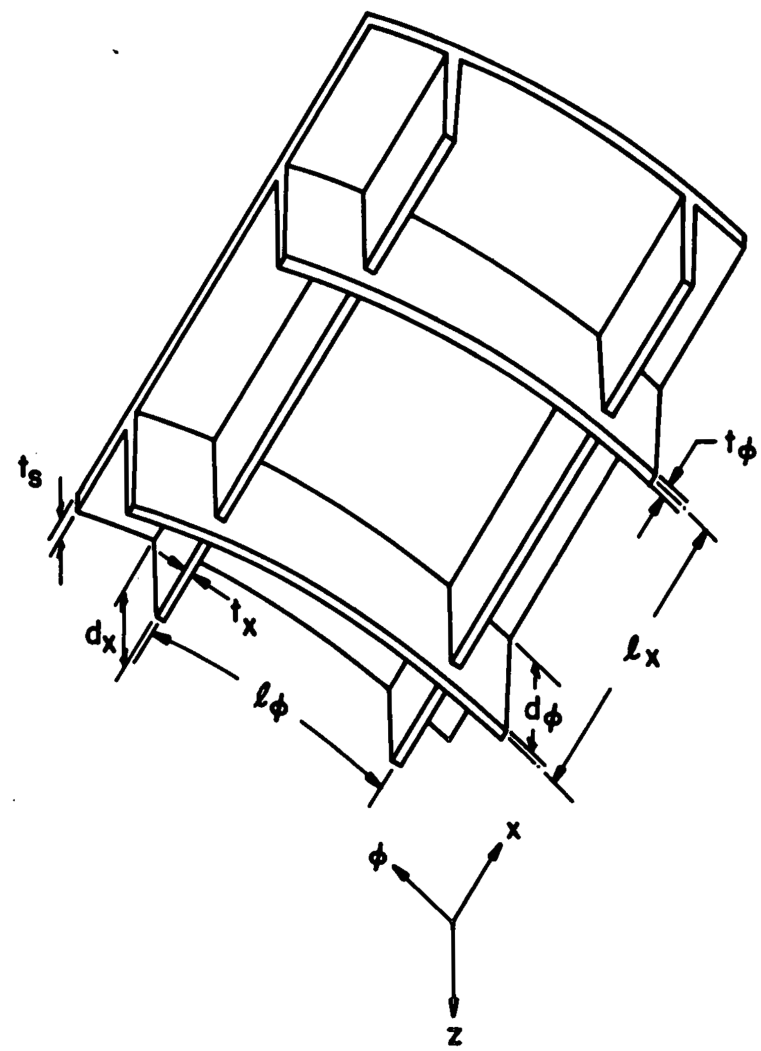


Fig.9.2 An Element of the Stiffened Cylinder

where U_j and L_j denote the upper and lower limits on the value of the jth design variable (D_j) for $j = 1,2,\ldots 7$. The lower bounds on the design variables $L_{j-7} \leq D_{j-7}$; $j = 8,9,\ldots 14$ are expressed in normalized form as follows

$$h_j(\vec{D}) = \frac{L_{j-7} - D_{j-7}}{U_{j-7} - L_{j-7}} \leq 0 \quad ; \quad j = 8,9,\ldots 14 \quad . \tag{9-4}$$

The geometric requirements that the stiffener thicknesses must not exceed the corresponding stiffener spacings, are expressed as follows:

$$t_x \leq \ell_\phi \quad , \quad \text{that is} \quad D_2 \leq D_7 \tag{9-5}$$

or in normalized form

$$h_j(\vec{D}) = \frac{D_2 - D_7}{U_7 - L_7} \leq 0 \quad ; \quad j = 15 \tag{9-6}$$

and

$$t_\phi \leq \ell_x \quad , \quad \text{that is} \quad D_3 \leq D_6 \quad , \tag{9-7}$$

or in normalized form

$$h_j(\vec{D}) = \frac{D_3 - D_6}{U_6 - L_6} \leq 0 \quad ; \quad j = 16 \quad . \tag{9-8}$$

Note that all the side constraints represented by Eq. (9-3), (9-4), (9-6) and (9-8) are normalized so that for acceptable designs

$$-1 \leq h_j(\vec{D}) \leq 0 \quad ; \quad j = 1,2,\ldots 16 \quad . \tag{9-9}$$

The stiffened cylinder is subject to a multiplicity of K distinct load conditions and the maximum number of load conditions that can be handled by the program reported in [9.1] is ten (i.e. $K_{max} = 10$). Each load condition (k) is specified by giving the applied uniform axial load per unit length of circumference (N_k, compression positive, tension negative), the net uniform radial pressure (p_k, inward positive, outward negative), and material properties for the shell and stiffeners corresponding to a given uniform temperature (T_k).

The automated minimum weight optimization procedure reported in [9.1] guards against unsatisfactory structural behavior by considering eleven independent failure modes as follows:

(1) buckling of the entire stiffened cylinder (Gross Buckling - G.B.)

(2) buckling of the stiffened cylinder between the circumferential stiffeners (Panel Buckling - P.B.)

(3) buckling of the cylindrical skin between longitudinal and circumferential stiffeners (Skin Buckling - S.B.)

(4) buckling of the longitudinal stiffeners (Longitudinal Stiffener Buckling - L.S.B.)

(5) buckling of the circumferential stiffeners due to contraction of the cylinder (Circumferential Stiffener Buckling Contraction - C.S.B.C.)

(6) buckling of the circumferential stiffeners due to expansion of the cylinder (Circumferential Stiffener Buckling Expansion - C.S.B.E.)

(7) yield failure under biaxial stress in the skin (Skin Yield - S.Y.)

(8) yield failure in the longitudinal stiffeners under uniaxial tensile stress (Longitudinal Stiffener Yield Tension - L.S.Y.T.)

(9) yield failure in the longitudinal stiffeners under uniaxial compressive stress (Longitudinal Stiffener Yield Compression - L.S.Y.C.)

(10) yield failure in the circumferential stiffeners under uniaxial tensile stress (Circumferential Stiffener Yield Tension - C.S.Y.T.)

(11) yield failure in the circumferential stiffeners under uniaxial compressive stress (Circumferential Stiffener Yield Compression - C.S.Y.C.).

Let the index i refer to the ith failure mode where $i = 1,2,\ldots 11$ and let the index k refer to the kth load condition where $k = 1,2,\ldots K \leq 10$. Each of the failure modes is characterized in terms of a behavior variable Y_{ik} such as a force resultant, a stress, or a strain. Each behavior variable in each load condition (Y_{ik}) is checked against its critical or limiting value to determine whether or not the structural behavior is acceptable. The failure mode constraints may be expressed as follows

$$h_{ik}(\vec{D}) = \frac{Y_{ik}}{(Y_{ik})_{cr}} - 1 \leq 0 \quad ; \quad \begin{array}{l} i = 1,2,\ldots 11 \\ k = 1,2,\ldots K \\ K \leq 10 \end{array} \quad . \tag{9-10}$$

Note that the behavior variables (Y_{ik}) and their limiting values $(Y_{ik})_{cr}$ may in general depend upon both the design $(\vec{D})$ and the load condition (k). The behavior constraints of Eq. (9-10) can be written in the following alternative form

$$h_j(\vec{D}) \leq 0 \tag{9-11}$$

where

$$j = 16 + k + (i - 1)K \quad ; \quad \begin{array}{l} i = 1,2,\ldots 11 \\ k = 1,2,\ldots K \end{array} \tag{9-12}$$

so that the behavior constraints are represented by

$$j = 17,18,\ldots \tilde{J} \tag{9-13}$$

where

$$\tilde{J} = 16 + 11K \quad . \tag{9-14}$$

Note that the behavior constraints have also been normalized (Eq. (9-10)) so that for acceptable designs

$$-1 \leq h_j(\vec{D}) \leq 0 \quad ; \quad j = 17,18,\ldots \tilde{J} \quad . \tag{9-15}$$

The objective of the optimization procedure is taken to be minimization of the total weight of the cylinder. The objective function (M) in terms of the design variables and preassigned parameters is

$$\begin{aligned} M(\vec{D}) = {} & 2\pi R L t_s \gamma_s + L\,|d_x|\,t_x \gamma_x n_x + \left|2R\,d_\phi - d_\phi^2 - t_s\,|d_\phi|\right| \pi\, t_\phi \gamma_\phi n_\phi \\ & - \mathrm{Min}\left(|d_x|, |d_\phi|\right) \delta_{x\phi}\, t_x t_\phi (\gamma_x \delta_{xw} + \gamma_\phi \delta_{\phi w})\, n_\phi n_x \end{aligned} \tag{9-16}$$

where

$$n_\phi = \frac{L - \ell_x}{\ell_x} \quad , \tag{9-17}$$

$$n_x = \frac{2\pi R}{\ell_d} \quad , \tag{9-18}$$

$$\delta_{x\phi} = \begin{cases} 0 & \text{stiffener sets on opposite sides of skin} \\ 1 & \text{stiffener sets on same side of skin} \end{cases} \quad , \tag{9-19}$$

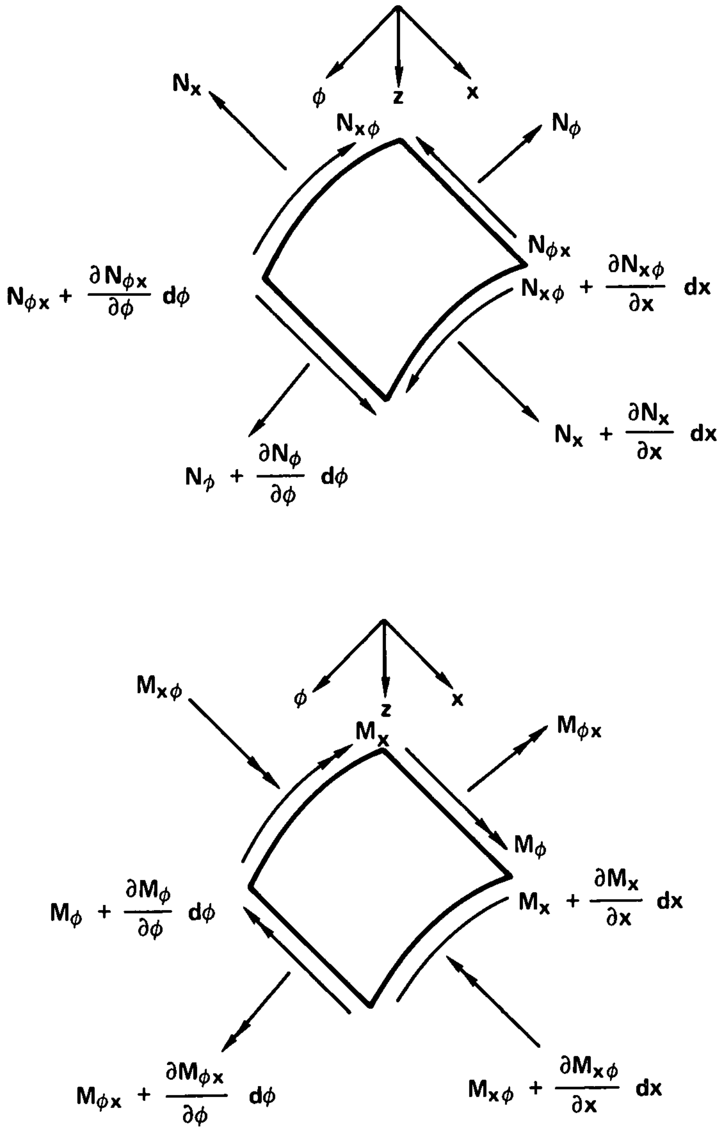


Fig.9.3 Force and Moment Resultants

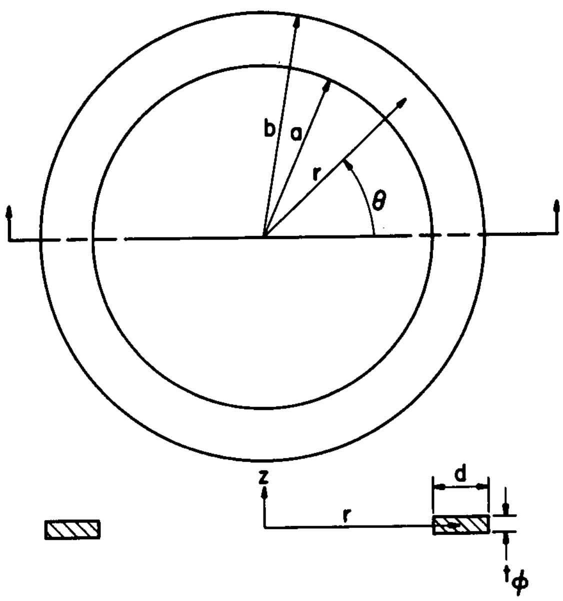


Fig.9.4 Circumferential Stiffener

$$\delta_{xw} = \begin{cases} 0 & \text{longitudinal stiffeners continuous} \\ 1 & \text{circumferential stiffeners continuous} \end{cases} , \tag{9-20}$$

and

$$\delta_{\phi w} = \begin{cases} 0 & \text{circumferential stiffeners continuous} \\ 1 & \text{longitudinal stiffenrs continuous} \end{cases} . \tag{9-21}$$

The first term in Eq. (9-16) represents the weight of the shell skin, the second term adds the weight of the longitudinal stiffeners, the third term introduces the weight of the circumferential stiffeners, and the fourth term accounts for the fact that the stiffeners may cross when they are on the same side of the cylinder and the material at this intersection must not be counted twice.

The problem statement can be summarized as follows :

Find $\vec{D}$

such that $h_j(\vec{D}) \leqslant 0 \quad ; \quad j = 1,2,\ldots\tilde{J}$

and $M(\vec{D}) \rightarrow \text{Min}$

where $\vec{D}$ is defined by Eq. (9-1), the $h_j(\vec{D})$ are given by Eq. (9-3), (9-4), (9-6), (9-8) and (9-11) and $M(\vec{D})$ is given by Eq. (9-16)

9.2.2 Features of the analysis

The first three failure modes involve determining the buckling load values for a cylindrical shell and comparing these with the corresponding applied load. The same basic analysis can be used to determine the critical loads for gross (G.B.), panel (P.B.), and skin (S.B.) buckling provided appropriate shell stiffness properties and buckling mode displacement patterns are employed. A linear small displacement buckling analysis is used and a uniform prebuckled membrane force distribution as well as simply supported boundary conditions are assumed. Bending and torsional stiffness of the stiffeners is taken into account as well as stiffener eccentricity; however initial imperfection sensitivity is neglected. In both the gross (G.B.) and panel (P.B.) buckling analyses the effects of the stiffeners are averaged over stiffener spacing (smeared).

The uniform prebuckled membrane force distribution is given by the following expressions

$$N_x = -N , \tag{9-22}$$

and

$$N_\phi = -pR . \tag{9-23}$$

The positive sign convention for force and moment resultants is indicated in Fig.9.3. The buckling equilibrium equations are those given by Flügge [9.6] but they contain only the buckling force terms recommended by Hedgepeth and Hall [9.7] and they are

$$\frac{\partial N_x}{\partial x} + \frac{1}{R}\frac{\partial N_{\phi x}}{\partial \phi} = 0 \tag{9-24}$$

$$\frac{1}{R}\frac{\partial N_\phi}{\partial \phi} + \frac{\partial N_{x\phi}}{\partial x} - \frac{1}{R^2}\frac{\partial M_\phi}{\partial \phi} - \frac{1}{R}\frac{\partial M_{x\phi}}{\partial x} - N\frac{\partial^2 v}{\partial x^2} = 0 \tag{9-25}$$

$$\frac{\partial^2 M_\phi}{\partial \phi^2} + R\frac{\partial^2 M_{x\phi}}{\partial x \partial \phi} + R\frac{\partial^2 M_{\phi x}}{\partial x \partial \phi} + R\frac{\partial^2 M_x}{\partial x^2} + RN_\phi - NR^2\frac{\partial^2 w}{\partial x^2} - pR\left(\frac{\partial^2 w}{\partial \phi^2} + w\right) = 0 . \tag{9-26}$$

The buckling equlibrium equations (Eq. (9-24), (9-25) and (9-26)) can be expressed in terms of displacements u, v and w using the force-displacement relations given by Eq. (A-2) and (A-3) of [9.1]. The force-displacement relations are obtained from the force resultant definitions in terms of the stresses by relating the stresses to the displacements using the elastic stress-strain law and the strain-displacement relations. Thus for example, the force resultant N_x may be expressed as a function of the displacements (u, v and w), the design variables $(\vec{D})$ and the material properties. The three buckling equilibrium equations in terms of the displacements u, v and w are homogeneous linear coupled partial differential equations [see Eq. (A-16) of [9.1]].

Substituting the following displacement functions

$$u = A \sin \eta\phi \ \cos \lambda x \tag{9-27}$$

$$v = B \cos \eta\phi \ \sin \lambda x \tag{9-28}$$

$$w = C \sin \eta\phi \ \sin \lambda x \tag{9-29}$$

into the three buckling equilibrium equations in terms of the displacements leads to a 3 × 3 stability determinant. Note that the assumed displacements given by Eq. (9-27), (9-28) and (9-29) satisfy the simply supported boundary conditions assumed. The same basic buckling analysis may be used for the gross buckling (G.B.), panel buckling (P.B.) and skin buckling (S.B.) analyses provided appropriate wave length parameters η and λ are chosen for each of the three failure modes as follows:

$i = 1$ gross buckling (G.B.)

$$\lambda = \frac{m\pi}{L} \quad ; \quad m = 1,2,\ldots \tag{9-30}$$

$$\eta = n \quad ; \quad n = 0,1,2,\ldots \quad , \tag{9-31}$$

$i = 2$ panel buckling (P.B.)

$$\lambda = \frac{m\pi}{\ell_x} \quad ; \quad m = 1,2,\ldots \tag{9-32}$$

$$\eta = n \quad ; \quad n = 0,1,2,\ldots \quad , \tag{9-33}$$

$i = 3$ skin buckling (S.B.)

$$\lambda = \frac{m\pi}{\ell_x} \quad ; \quad m = 1,2,\ldots \tag{9-34}$$

$$\eta = \frac{n\pi}{\ell_\phi} \quad ; \quad n = 1,2,\ldots \quad . \tag{9-35}$$

In each case, setting the 3 × 3 stability determinant to zero gives an expression for the buckling load of the form

$$N_{ik} = f(\vec{D}, PP, i, k, m, n) \quad . \tag{9-36}$$

Thus given a design $\vec{D}$ and the preassigned parameters PP (R, L and the material properties), buckling loads for the ith failure mode $i = 1,2,3$ in the kth load condition $(N_{ik})_{cr}$ can be obtained by seeking the minimum of (N_{ik}) over a range of integer values for m and n; that is

$$(Y_{ik})_{cr} = (N_{ik})_{cr} = \underset{m}{\mathrm{Min}}\ \underset{n}{\mathrm{Min}}\ (N_{ik}) = f(\vec{D}, PP, i, k, n^*, m^*)$$

where m^* and n^* denote the integer values of m and n that make N_{ik} a minimum.

It is useful to sort out, order and store the first M most nearly critical combinations of m and n. The first M most nearly critical combinations of m and n provide a basis for conducting approximate buckling analyses in failure modes $i = 1,2,3$, that is in gross (G.B.), panel (P.B.) and skin buckling (S.B.). As modest changes in the design are made during the optimization procedure shifting of the critical buckling mode shape is to be expected, but it is very likely that the new critical mode shape will be amongst the previously identified M most nearly critical modes. This characteristic is used to advantage subsequently in constructing the optimization procedure (see Section 9.2.3).

The buckling of the longitudinal stiffeners is guarded against using a failure analysis that treats the stiffeners as a long plate simply supported on three edges and free on the fourth. Because the longitudinal and circumferential stiffeners can have different depths and because they may indeed not even be on the same side of the shell, provision is made for using various combinations of plate planform dimensions in this analysis. The longitudinal stiffener buckling failure mode is represented by the following inequality

$$h_{4k}(\vec{D}) = \frac{Y_{4k}}{(Y_{4k})_{cr}} - 1 \leqslant 1 \tag{9-38}$$

where

$$Y_{4k} = (\sigma_{xs})_k = E_{xs}\left[\frac{H_\nu p_k R - (H_{s2} + H_\phi) N_k}{(H_{s1} + H_x)(H_{s2} + H_\phi) - H_\nu^2}\right], \tag{9-39}$$

and

$$(Y_{4k})_{cr} = (\sigma_c)_k = -\frac{\pi^2 E_{xs}}{12(1 - \nu_{xs}^2)}\left(\frac{t_x}{d}\right)^2\left[\left(\frac{d}{\ell}\right)^2 + 0.425\right] \quad . \tag{9-40}$$

In Eq. (9-39) the H's are section properties that depend upon the design variables and the material properties, p_k and N_k are mechanical loads for the kth load condition, R is the shell radius, and E_{xs} is the modulus of elasticity of the longitudinal stiffeners. In Eq. (9-40) ν_{xs} represents the Poisson's ratio and t_x is the thickness of the longitudinal stiffeners. Selection of the length (ℓ) and the depth (d) to be used in computing the longitudinal stiffener buckling stress is carried out according to the following prescription:

(1) stiffeners on opposite sides of the shell

$d = |d_x|$, $\ell = L$

(2) stiffeners on the same side of the shell and

$|d_x| < |d_\phi|$, then let

$d = |d_x|$, $\ell = \ell_x$

(3) stiffeners on the same side of the shell but

$|d_x| > |d_\phi|$, then $(\sigma_c)_k$ is given by Eq. (9-40) with

either

(A) $d = |d_x|$, $\ell = \ell_x$

or

(B) $d = |d_x| - |d_\phi|$, $\ell = L$

whichever gives $|(\sigma_c)_k|$ smallest.

It should be noted that analogous situations are encountered with respect to the determination of the critical buckling strain in the failure mode analysis of the circumferential stiffeners.

The circumferential stiffener failure mode analysis treats the stiffener as a circular plate with a concentric circular hole in the middle, simply supported along the edge that forms the shell (see Fig.9.4). Due to their curvature external circumferential stiffeners can buckle when the cylinder expands. Two separate failure modes are considered: one associated with contraction of the cylinder $\varepsilon_\phi < 0$ (C.S.B.C.) the other associated with expansion $\varepsilon_\phi > 0$ (C.S.B.E.). In the case of contraction (C.S.B.C.) there are six possibilities that must be considered:

circumferential stiffener inside $d_\phi > 0$	(1) longitudinal stiffener outside
	(2) inside and $|d_x| \geq |d_\phi|$
	(3) inside and $|d_x| < |d_\phi|$
circumferential stiffener outside $d_\phi < 0$	(1) longitudinal stiffener inside
	(2) outside and $|d_x| \geq |d_\phi|$
	(3) outside and $|d_x| < |d_\phi|$.

In the case of <u>expansion</u> (C.S.B.E.) of the shell, circumferential stiffener buckling can only occur when it is on the outside of the shell and only three possibilities need to be considered:

circumferential stiffener outside $d_\phi < 0$

(1) longitudinal stiffener inside

(2) outside and $|d_x| \geq |d_\phi|$

(3) outside and $|d_x| < |d_\phi|$.

The remaining failure modes $i = 7,8,9,10,11$ deal with yield stress constraints and they need not be elaborated on here. It may be noted, however, that the yield constraint for the skin considers the biaxial stress condition. This failure mode was found to be of particular importance in the case of barrel shells [9.2].

9.2.3 Features of the optimization procedure

The problem is formulated using the Fiacco-McCormick interior penalty function approach (see Sections 2.6.2 and 6.3.1). This formulation transforms the basic inequality constrained minimization problem into a sequence of unconstrained minimizations that are carried out using the variable metric algorithm described in Section 6.2.6. The constraint repulsion characteristics of the Fiacco-McCormick interior penalty function facilitate the use of approximate analyses. In particular the three cylindrical shell buckling analyses (G.B., P.B. and S.B.) are carried out using a drastically reduced number of possible buckling mode shapes (m,n). At the beginning of each unconstrained minimization stage a full buckling analysis is executed and the M most nearly critical combinations of (m,n) are ordered and stored. Then, within that unconstrained minimization stage, the shell buckling analyses are approximate in the sense that the search for the critical buckling mode shape is carried out over only the M combinations of (m,n) identified at the beginning of the stage.

The Fiacco-McCormick penalty function formulation for this problem can be expressed as follows

$$\phi(\vec{D}, r_p) = M(\vec{D}) - r_p[P_s(\vec{D}) + P_b(\vec{D})] \tag{9-41}$$

where

$$P_s(\vec{D}) = \sum_{j=1}^{16} \frac{1}{h_j(\vec{D})} \qquad \text{side constraints,} \tag{9-42}$$

and

$$P_b(\vec{D}) = \sum_{j=17}^{J} \frac{1}{h_j(\vec{D})} \qquad \text{behavior constraints} \quad . \tag{9-43}$$

The gradient to the function $\phi(\vec{D}, r_p)$ has the following form

$$\nabla\phi = \nabla M - r_p[\nabla P_s + \nabla P_b] \tag{9-44}$$

and the gradients ∇M and ∇P_s are determined from analytic expressions for the partial derivatives while the gradient ∇P_b is obtained using first order forward finite difference approximations for the partial derivatives, i.e.

$$\frac{\partial P_b}{\partial D_i}(\vec{D}) \approx \frac{1}{\Delta D_i}\left[P_b(\vec{D})\Big|_{(D_i + \Delta D_i)} - P_b(\vec{D})\Big|_{D_i}\right] \quad . \tag{9-45}$$

In Eq. (9-45) it is assumed that the critical buckling mode shape is the same at $\vec{D}$ and $\vec{D} + \Delta\vec{D}$. The selection of the finite difference increment sizes ΔD_i can be guided by some foreknowledge of the gross proportions of the design.

The unconstrained minimization of the function $\phi(\vec{D}, r_p)$ for each stage is carried out using the variable metric algorithm. The (q + 1)th design is obtained from the qth design through a design modification defined by a direction $\vec{S}_q$ and a magnitude α_q, i.e.

$$\vec{D}_{q+1} = \vec{D}_q + \alpha_q \vec{S}_q \tag{9-46}$$

where

$$\vec{S}_q = - H_q \, \phi(\vec{D})_q \tag{9-47}$$

and α_q is the distance to the minimum of $\phi(\vec{D}_q + \alpha\vec{S}_q) = f(\alpha)$ along $\vec{S}_q$. The matrix H_q is initially taken as the identity matrix and is then systematically updated according to the prescription given in Section 6.2.6.

For a specified direction $\vec{S}_q$ within an r_p stage the problem reduces to a one-dimensional minimization problem. To find the minimum of $f(\alpha)$ along a line, an incrementation scheme, with the slope as a test, is used to locate two points such that the minimum lies between them. Then, using the function value and slope at these two points, a cubic interpolation is made to estimate the location of the minimum. It should be noted that the H_q matrix (see Eq. (9-47)) is not updated unless the one-dimensional minimum has been found within a prescribed tolerance. Also the H_q matrix is reset to I whenever the number of one-dimensional minimizations equals the number of independent design variables.

A maximum of five cubic interpolations is made in order to obtain convergence of the one-dimensional minimization. Convergence is said to have occurred if either the dot product test is satisfied, i.e.

$$\frac{\nabla\phi}{|\nabla\phi|} \cdot \frac{\vec{S}_q}{|\vec{S}_q|} \leqslant 0.005 \qquad (9\text{-}48)$$

or the distance between the two points straddling the minimum is less than a specified minimum.

Three alternative criteria are used to test for convergence of each n dimensional unconstrained minimization stage in the sequence. Convergence of the pth stage is assumed when any one of the following three criteria is satisfied:

(1) absolute value of the gradient $|\nabla\phi| < \varepsilon$ where $\varepsilon = \dfrac{|\nabla\phi|_{\text{initial}}}{10^n}$ and $n = 3$ or 4

(2) estimated amount by which ϕ exceeds its minimum is less than 2% (after n one-dimensional minimizations, just prior to resetting H_q matrix to I), i.e. $\frac{1}{2}\,\dfrac{\nabla\phi_q^T\, H_q\, \nabla\phi}{\phi_q} < 0.02$

(3) minimum move distance test converged if a move in the negative gradient direction $(\vec{S}_q = -\nabla\phi_q)$ which is twice the minimum move distance causes violation of any constraint $[h_j(\vec{D}_q + 2T_{\min}\,\vec{S}_q) > 0]$ or if the sign of the slope is reversed, i.e. if $\nabla\phi(\vec{D}_q + 2T_{\min}\,\vec{S}_q) \cdot \vec{S}_q$ has its sign opposite to $\nabla\phi(\vec{D}_q) \cdot \vec{S}_q$.

Convergence of the sequence of n dimensional unconstrained minimization stages is usually based upon a criterion that depends upon the primal-dual nature of the Fiacco-McCormick method. It is noted that this criterion given in [9.8] depends upon the convexity of the programming problem. An option to terminate the SUMT procedure after converging a user prescribed number of stages is also provided in the computer program. Once a minimum is obtained for one value of the parameter r_p, bounds can be placed on the value of the minimum weight. The minimum weight value is bounded below by the value of the dual objective function and above by the current value of the weight. This leads to the following convergence criterion [9.8]

$$\frac{M - \Theta}{\Theta} \leqslant \varepsilon \qquad (9\text{-}49)$$

where ε is a small number to be assigned and Θ is the value of the dual objective function given by

$$\Theta(\vec{D}, r_p) = M(\vec{D}) + \sum_{j=1}^{\tilde{J}} \frac{1}{h_j(\vec{D})} \quad . \qquad (9\text{-}50)$$

There are several control parameters, in addition to the convergence criteria, that influence the operational efficiency of this design optimization procedure in application. Some suggestions for the selection of these parameters based upon operational experience with the program are:

(1) select the initial value of r_p such that

$$M(\vec{D}_o) \approx -\, r_1 \sum_{j=1}^{\tilde{J}} \frac{1}{h_j(\vec{D}_o)} \qquad (9\text{-}51)$$

(2) set the cut factor applied to r_p after each stage equal to $\frac{1}{2}$ (i.e. let $c = \frac{1}{2}$ so that $r_{p+1} = \frac{1}{2}\, r_p$),

(3) let the number of near critical ordered modes saved for the approximate shell buckling analyses be

(a) gross buckling, 40 modes (except for cases with external pressure, then 10 modes),

(b) panel buckling, 20 modes,

(c) skin buckling, 10 modes,

(4) let the number of modes examined in the 'complete' shell buckling analyses be

(a) gross buckling, longitudinal $m_{max} = 30 \rightarrow 50$, circumferential $n_{max} = 30$,

(b) panel buckling, longitudinal $m_{max} = 10 \rightarrow 20$, circumferential $n_{max} = 50 \rightarrow 150$,

(c) skin buckling, longitudinal $m_{max} = 20 \rightarrow 30$, circumferential $n_{max} = 15 \rightarrow 20$.

9.2.4 Sample Results

A substantial body of experience has been gained with this capability and results for over 30 cases were reported in [9.1]. These numerical results illustrated the following points:

(1) the effectiveness of the penalty function approach when used in conjunction with analysis approximations,

(2) the influence of various combinations of internal and external stiffening,

(3) the sensitivity of the minimum weight design to loading intensity and minimum gage limitations,

(4) the importance of considering multiple load conditions,

and (5) the existence of relative minima in the design space associated with design subconcepts embedded within the basic problem statement.

Consider the following example, Case 1-I' taken from [9.1]. The preassigned parameters are R = 60 in, L = 165 in; the material is aluminium with the following properties:

$$E = 10 \times 10^6 \text{ lb/in}^2 ,$$

$$\nu = 0.333 ,$$

$$\rho = 0.101 \text{ lb/in}^3 ,$$

$$\sigma_y = 50000 \text{ lb/in}^2 .$$

The initial trial design has all internal stiffening and the following minimum gage requirements are stipulated;

$$t_s \geqslant 0.19 \text{ in} ,$$

$$t_x \geqslant 0.050 \text{ in} ,$$

$$t_\phi \geqslant 0.050 \text{ in} .$$

The stiffened shell is subject to a set of three distinct load conditions summarized as follows:

Load condition \ Loads	N lb/in + compressive	p lb/in^2 + external pressure
1	700	0
2	940	-2.0
3	212	+0.4

The initial trial design and the final proposed optimum design are depicted graphically in Fig.9.5. The weight is reduced from 715 lb to 293 lb. It may be noted that the stiffener thicknesses are essentially minimum gage. There are five other constraints that are critical or near critical for the final design shown in Fig.9.5 and they are

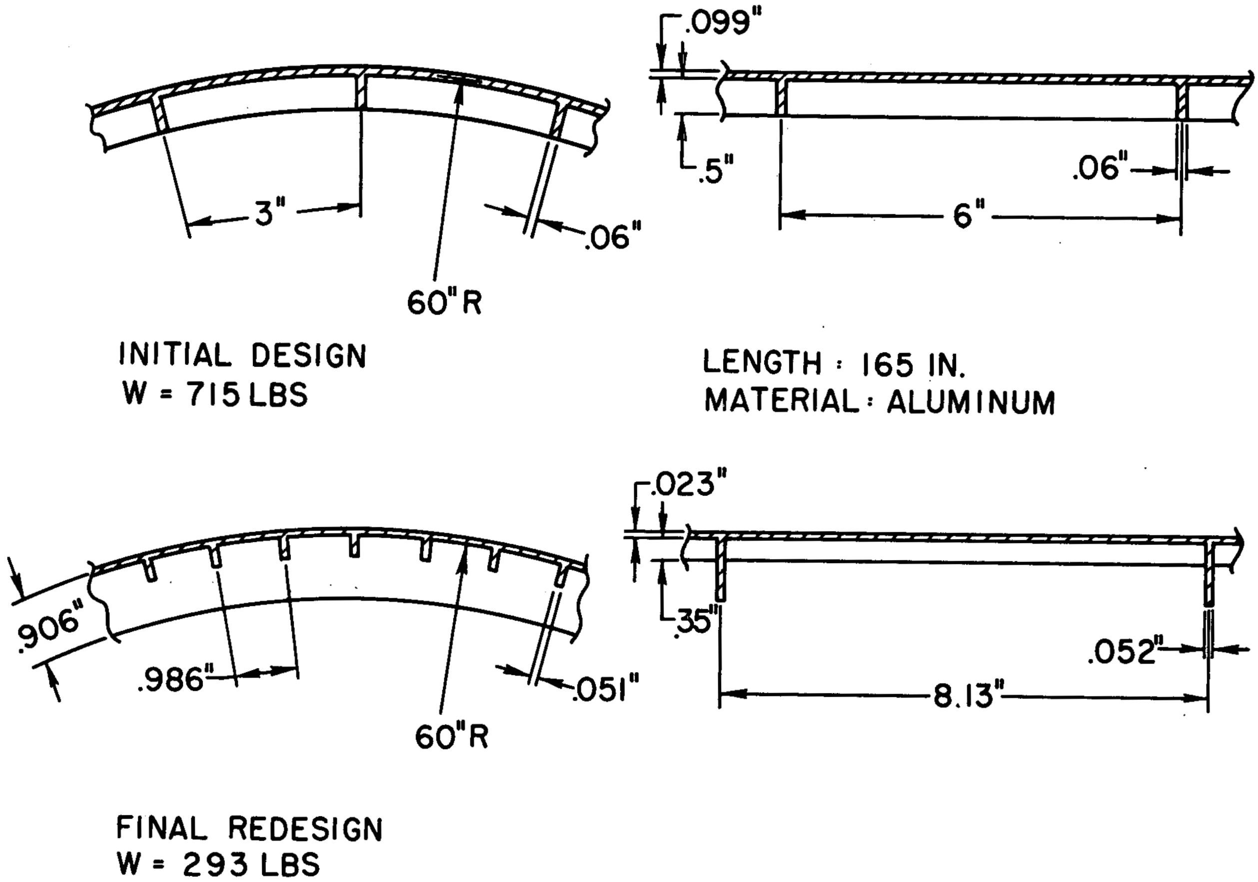


Fig.9.5 Initial and Final Design (Case 1–I′)

(1) gross buckling in load condition 2,

$$\frac{Y_{12}}{(Y_{12})_{cr}} = 0.999 \quad ,$$

(2) skin buckling in load condition 2,

$$\frac{Y_{32}}{(Y_{32})_{cr}} = 0.999 \quad ,$$

(3) panel buckling in load condition 2,

$$\frac{Y_{22}}{(Y_{22})_{cr}} = 0.975 \quad ,$$

(4) skin yield in load condition 3,

$$\frac{Y_{73}}{(Y_{73})_{cr}} = 0.968 \quad ,$$

(5) skin buckling in load condition 1,

$$\frac{Y_{31}}{(Y_{31})_{cr}} = 0.915 \quad ,$$

The design improvement depicted in Fig.9.5 was achieved in twelve unconstrained minimization stages in which r_p was reduced by a factor of $\frac{1}{2}$ for each subsequent stage i.e. $[r_{p+1} = \frac{1}{2} r_p,$ $p = 1,2,\ldots 12]$. The total run time for the Fortran IV program on the Univac 1107 computer was approximately 15 minutes. Essentially the same results have been obtained on a Univac 1108 and a CDC 6600 computer with run times less than 5 minutes. It is interesting to note that for this particular three load condition example, Case 1-I' from [9.1], the 1107 machine time required for a complete analysis was 35 seconds while an approximate analysis required 0.5 second. The efficiency gained as a result of using approximate analyses for the cylindrical shell buckling mode analyses is very significant.

A collection of twelve examples based on this one basic problem was studied and reported in [9.1]. The twelve cases examined can be generated by considering all combinations of four structural concepts and three load levels. The structural concepts are:

(1) all inside stiffening, no minimum gage restrictions,

(2) circumferential stiffening inside and longitudinal stiffening outside, no minimum gage restrictions,

(3) all outside stiffening, no minimum gage restrictions,

and (4) all inside stiffening, with minimum gage restrictions as discussed previously.

The increasing levels of load intensity are given by θN_k, θp_k where $\theta = 1, 2$, and 3. The minimum weights obtained in pounds for each of the twelve cases are summarized as follows:

Concept \ Θ	1	2	3
Inside	231	340	445
Inside Outside	235	358	459
Outside	240	363	468
Inside min. gage	293	389	490

The minimum weight for Case 1-I' previously discussed in some detail is underlined. The foregoing summary of results show the strong influence on the optimum weight of minimum gage limitations. It can also be observed that there is a higher percentage penalty for imposing minimum gage limitations on lightly loaded structures than on more heavily loaded structures. It is also apparent, from these results, that there is only a moderate weight reduction associated with the various combinations of internal and external stiffening examined, in this instance. The optimization capability provides a means of evaluating alternative stiffening concepts based on a comparison of optima. While the best concept was found to be load condition dependent in [9.1], it should be noted that the maximum weight reduction associated with alternative stiffener locations (inside-outside) did not exceed 12% for any of the examples studied.

The results reported in [9.1] reinforce the contention that subconcepts contained within the basic problem statement are often associated with relative minima pockets in the design space. Initial designs with all internal stiffening led to final designs with all internal stiffening. The same observation can be made with regard to all external stiffening and mixed internal external stiffening. It should be noted that the options provided in the program, to preassign any subset of design variables and to fix side constraint limits can be used as a creative control device. The user of the program, therefore, can force the optimization procedure to search for the best design within various subconcept regions in the design space. This situation illustrates the complementary relationship that exists between automated optimization procedures and man-machine communication. The experience reported in [9.1] suggests that the successful application of mathematical programming techniques to structural design optimization for complex special purpose applications requires tailoring the analysis and optimization procedures together.

9.2.5 Recent Further Developments

An extension to barrel shells by Stroud and Sykes [9.2] of the stiffened cylindrical shell optimization program reported in [9.1] should be noted. As an illustration of the important role structural optimization capabilities can play in evaluating design concepts the following quotation from [9.2] is cited: "For shells designed to support axial compressive loads, the results show that important weight savings can be provided by slight meridional curvature. For the particular shell examined herein, the maximum weight saving is about 30%. The large increases (factors of 5 to 9 in strength) recently attributed to barreling cannot be directly translated into weight savings when comparisons are made between minimum-weight designs. Yielding becomes an important failure constraint at lower loads for barreled shells than for cylindrical shells."

Kicher and Chao [9.3] have recently reported the development of a structural optimization capability for stiffened fiber composite cylinders. The overall length and radius of the cylinder are preassigned and both longitudinal and circumferential hat cross section stiffeners are considered. The design variables include the depth and width of the hat stiffeners, the stiffener spacings, the fiber volume content, and the ply orientation angles. Multiple load conditions are considered and each load condition is described in terms of a combination of axial, radial, and torsional load. In addition to constraints on the range of the design variables, geometric realizability constraints and behavior constraints are considered. The behavior constraints are formulated in terms of critical stresses and strains, and they guard against unsatisfactory behavior in each failure mode in each load condition. The following eight failure modes are considered in [9.3]: (1) gross buckling, (2) panel buckling, (3) skin buckling, (4) longitudinal stiffener buckling, (5) circumferential stiffener buckling, (6) material failure in the skin, (7) material failure in the longitudinal stiffeners, and (8) material failure in the circumferential stiffeners. The linear eigenvalue analysis for gross and panel buckling is based upon a method similar to that of Cheng and Ho [9.9]. The cylindrical shell is assumed to buckle into a torsional waveform. Eight sets of boundary conditions are provided, and the detailed development of the buckling analysis used is given in [9.10]. The weight of the fiber composite stiffened cylinder is taken to be the objective function.

The design optimization problem is formulated in design space using the Fiacco-McCormick interior penalty function and the sequence of unconstrained minimizations is carried out using the variable metric method. It is pointed out that the weight function is independent of the ply angles and hence the influence of changing the ply angles is present only in the penalty term of the Fiacco-McCormick function $\phi(\vec{D}, r_p)$. It is observed that the decreasing sensitivity of the $\phi(\vec{D}, r_p)$ function to changes in ply angle as r_p decreases leads to computational inefficiency. A device which artificially increases the influence of the ply angles on the penalty function is introduced. Numerical results for several example problems are presented in [9.3] and [9.10] and the effectiveness of the algorithmic modification is illustrated. These results also demonstrate the capabilities of the optimization procedure in the design of stiffened fiber composite cylinders. It is shown that alternative optima are common for the type of structure considered; i.e. the set of design variable values which yields the minimum weight is not unique. The research results reported in [9.3] and [9.10] extend the application of mathematical programming to include ply angles and fiber volume fraction as design variables in the minimum weight design of stiffened fiber composite shells.

Waddoups, McCullers, Olsen, and Ashton [9.4] have recently reported a minimum weight structural optimization capability for a class of anisotropic plate structures. This development includes capabilities to design: (1) a uniform plate with complex membrane load conditions, (2) a uniform plate with combined bending and membrane load conditions, and, (3) a simple multicell wing box with a refined design of the compression cover. A choice of thick plate, rigid core sandwich, or stiffened plate construction is available. In each case the skins are assumed to be of laminated fiber composite construction, and the design variables include the thickness and fiber orientation for each lamina. The most general problem formulated in [9.4] involves 21 design variables (12 for the cover plate and 9 for the wing box), 45 distinct failure modes, and a maximum of 3 independent load conditions. The program reported permits optional preassigning of a subset of design variables, and it provides for linking of fiber orientation and lamina thickness design variables. The Fiacco-McCormick interior penalty function formulation with a variable metric (Davidon-Fletcher-Powell) unconstrained minimization algorithm was employed. The use of various analysis approximations during major portions of the optimization procedure was the key to achieving the low machine running times reported. While the capability described in [9.4] is oriented toward a special class of structures (anisotropic fiber composite plates), it is viewed as an important practical application of mathematical programming techniques to structural design within the context of aerospace engineering practice.

9.3 Ablating Thermostructural Panel Example

This example, reported in [9.5], illustrates the application of mathematical programming techniques to the design optimization of a refurbishable composite type ablating heat shield. The design concept shown in Fig.9.6 is drawn from [9.11]. The functions of the major panel components in this concept are qualitatively describes as follows:

(1) the ablator protects the substructure from the severe thermal environment associated with re-entry,

(2) the substructure transfers the pressure loading through supporting structure to the primary structure (it must be stiff enough and thermally compatible with the ablator material so as to avoid cracking of the charred ablator),

(3) the insulation, which is assumed to be nonstructural, keeps the primary structure and the vehicle interior at an acceptably low temperature.

9.3.1 Problem Statement

The idealization on which the problem formulation rests is depicted in Fig.9.7. The non-linear transient thermal analysis is treated one-dimensionally, considering only temperature gradients through the thickness of the panel. The structural analysis assumes that the flat rectangular panel can be treated as a strip exhibiting curvature in the x direction only (see Fig.9.7).

The ablator, substructure, and insulator materials and their temperature dependent mechanical and thermal properties are preassigned parameters. The design variables are the various thicknesses x_1 through x_5 shown in Fig.9.7, and the planform dimensions of the panel, x_6 and x_7. The loading environment is described by the heat flux input as a function of time $q_c(t)$ and the pressure loading as a function of time $p(t)$. These depend upon the re-entry trajectory and the atmosphere.

Nine failure modes are guarded against by limiting:

(1) the temperature at the ablator substructure interface,

(2) the temperature at the back of the insulation,

(3) the panel midpoint deflection,

(4) ablator stress level,

(5) outer sandwich face stress level,

(6) intercell face buckling stress,

(7) inner sandwich face stress level,

(8) tensile strain in the ablator, and

(9) compressive strain in the ablator.

Two alternative objective functions are considered. Minimization of the weight per unit area of surface protected may be taken as the goal of the optimization procedure. In this case it may be desirable to impose a constraint on the maximum total depth of the shield. Alternatively, minimization of the total depth of the shield may be taken as the objective function subject to a constraint on the maximum weight per unit surface area protected.

9.3.2 Features of the Thermal Analysis

A simplified one-dimensional ablation analysis due to Swann and Pittman was used to predict the transient temperature distribution [see Appendix A of [9.5]]. This analysis takes into account the surface recession as well as the transient convective heating and reradiative effects. The charring ablator is treated as though it were a subliming ablator; however, the analysis considers the blocking effect of pyrolysis gases on convective heating rate and the oxidation of the char residue at the receding ablator surface. The material properties of all layers are taken to be temperature dependent.

Referring to Fig.9.7 the heat conduction equation for the ablator can be written as

$$\rho_1 c_{p1} \left(\frac{\partial T}{\partial t}\right)_x = \frac{\partial}{\partial x}\left[k_1 \left(\frac{\partial T}{\partial x}\right)_t\right] \quad ; \quad s(t) < x < x_1 \quad . \tag{9-52}$$

It proves convenient to introduce the following coordinate transformation

$$\lambda = \frac{x - s}{x_1 - s} \quad . \tag{9-53}$$

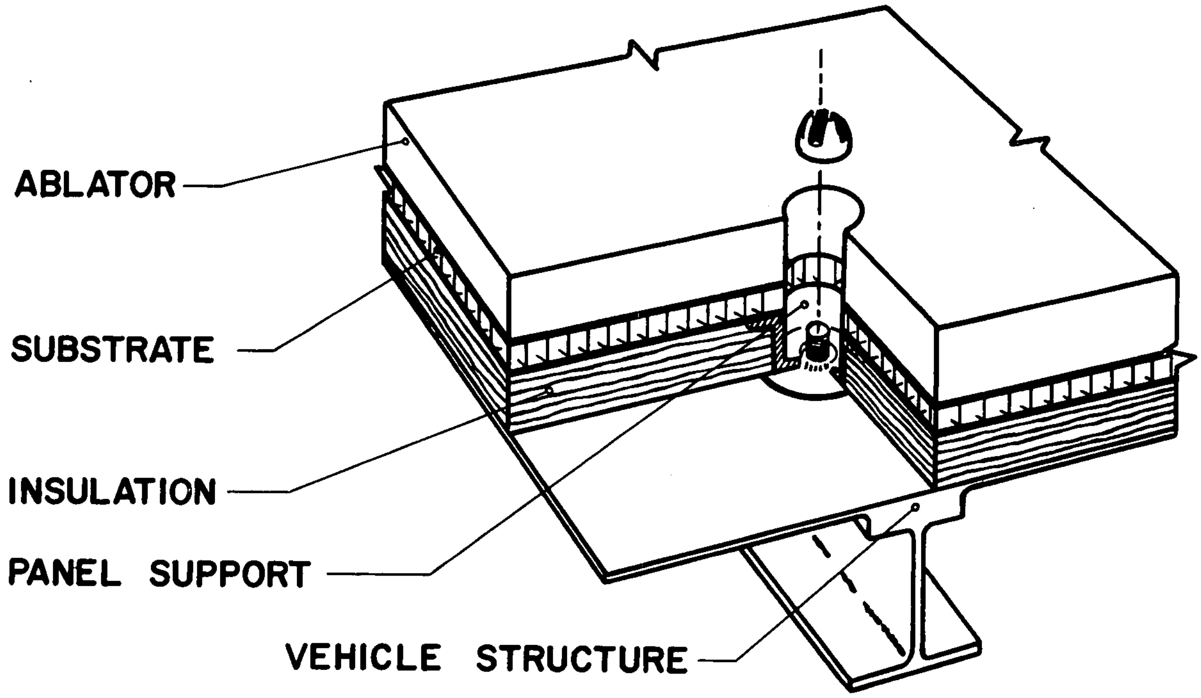


Fig.9.6 The Double Wall Ablative Heat Shield Concept

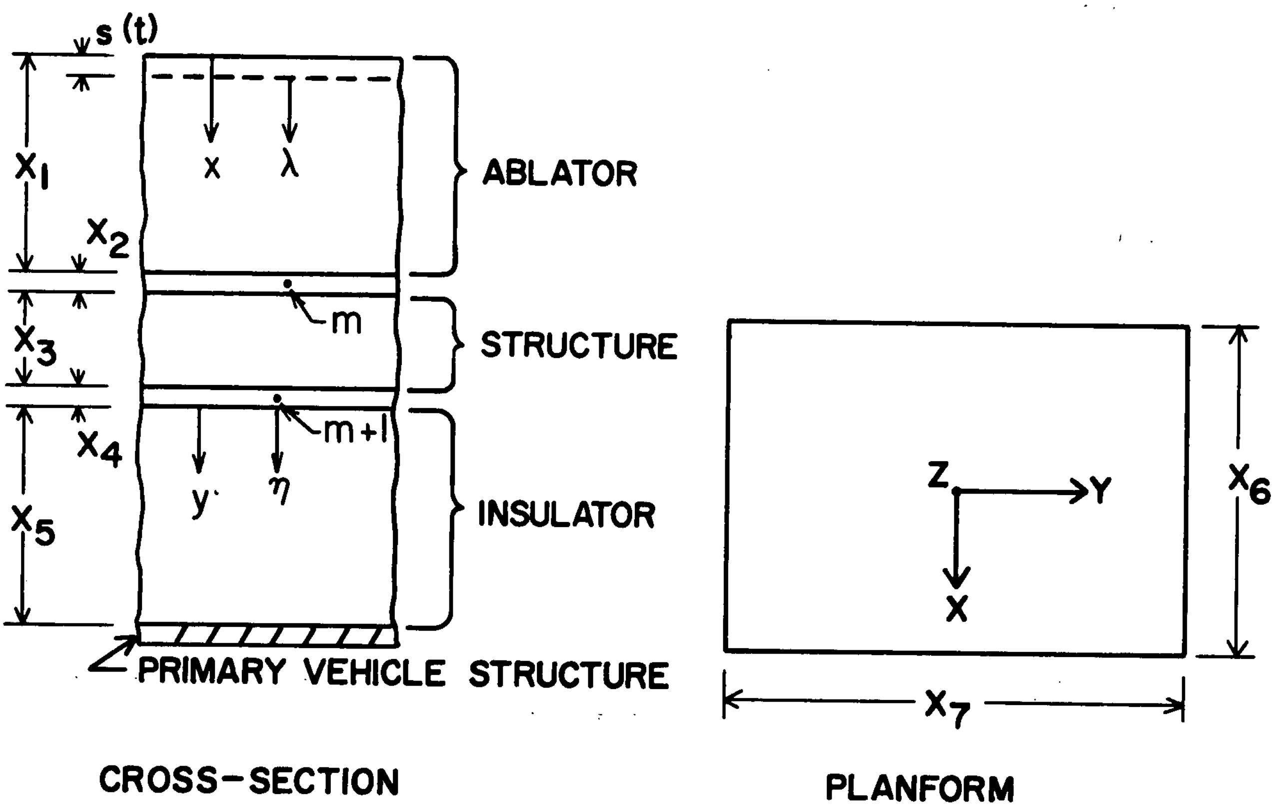


Fig.9.7 Design Variables

Making this change of variable in Eq. (9-52) yields

$$\rho_1 c_{p1} \left[\left(\frac{\partial T}{\partial t}\right)_\lambda + \frac{\lambda - 1}{x_1 - s}\frac{ds}{dt}\left(\frac{\partial T}{\partial \lambda}\right)_t\right] = \frac{1}{(x_1 - s)^2}\left[\frac{\partial k_1}{\partial \lambda}\left(\frac{\partial T}{\partial \lambda}\right)_t + k_1\left(\frac{\partial^2 T}{\partial \lambda^2}\right)_t\right] \quad ; \quad 0 < \lambda < 1 \quad . \tag{9-54}$$

The boundary condition at the receding surface is

$$q(t) = -k_1\left(\frac{\partial T}{\partial x}\right)_t \quad ; \quad x = s(t) \tag{9-55}$$

or making the change of variable indicated in Eq. (9-53)

$$q(t) = -\frac{k_1}{(x_1 - s)}\left(\frac{\partial T}{\partial \lambda}\right)_t \quad ; \quad \lambda = 0 \tag{9-56}$$

where

$$\begin{aligned} q(t) = {} & +(\text{convective heating}) + (\text{combustive heating}) \\ & -(\text{blocking}) - (\text{reradiation}) . \end{aligned} \tag{9-57}$$

In the transient temperature distribution analysis it is assumed that the face sheets of the sandwich substructure are thin, so that no temperature gradient exists through the thickness of a face sheet. However, the face sheets are assumed to have significant heat capacity. The core of the sandwich is assumed to have negligible heat capacity and a linear temperature gradient is assumed to exist through the core (between the two sandwich face sheets). On this basis, the heat balance relations for the sandwich faces (see Fig.9.7) are

$$x_2 \rho_2 c_{p2} \frac{\partial T_m}{\partial t} = -\frac{k_1}{(x_1 - s)}\left.\frac{\partial T}{\partial \lambda}\right|_{\lambda=1} - Q_{m,m+1} \tag{9-58}$$

and

$$x_4 \rho_4 c_{p4} \frac{\partial T_{m+1}}{\partial t} = \frac{k_5}{x_5}\left.\frac{\partial T}{\partial \eta}\right|_{\eta=0} + Q_{m,m+1} \tag{9-59}$$

where

$$Q_{m,m+1} = \frac{k_e}{x_3}(T_m - T_{m+1}) \quad , \tag{9-60}$$

$$\eta = \frac{y}{x_5} \tag{9-61}$$

and k_e denotes the effective thermal conductivity of the sandwich core. Referring to Fig.9.7 the heat conduction equation governing the transient temperature distribution in the insulator is

$$\rho_5 c_{p5}\left(\frac{\partial T}{\partial t}\right)_\eta = \frac{k_5}{x_5^2}\frac{\partial^2 T}{\partial \eta^2} + \left(\frac{1}{x_5^2}\frac{\partial k_5}{\partial \eta}\right)\frac{\partial T}{\partial \eta} \quad ; \quad 0 < \eta < 1 \quad . \tag{9-62}$$

It is assumed that no heat flows from the insulation into the primary structure and hence the appropriate boundary condition at the interface between the insulation and the primary structure is

$$\frac{\partial T}{\partial \eta} = 0 \quad ; \quad \eta = 1 \quad . \tag{9-63}$$

The thermal response is governed by the field equations (Eq. (9-54) and (9-62)), the heat balance relations (Eq. (9-58) and (9-59)), and the boundary conditions (Eq. (9-56) and (9-63)).

These governing relationship can be cast in implicit finite difference form (see Appendix C of [9.5]), so that

$$[C]\ \overline{T}_{t+k_j} = \overline{T}_t \tag{9-64}$$

where the matrix [C] is tridiagonal and its elements depend upon the temperature at time $(t + k_j)$. Given the time increments k_j and k_{j-1} as well as the temperature distribution $\overline{T}_t$ and $\overline{T}_{t-k_{j-1}}$ linear extrapolation is used to compute the estimated temperature distribution at time $(t+k_j)$, i.e. $\overline{T}'_{t+k_j}$. The elements of the matrix [C] in Eq. (9-64) are then evaluated using the estimate $\overline{T}'_{t+k_j}$ and Eq. (9-64) is then solved for $\overline{T}_{t+k_j}$. This result is compared with the estimated temperature distribution $\overline{T}'_{t+k_j}$. If the agreement is close enough the iterative process terminates, if not the temperature distribution obtained by solving Eq. (9-64) is used as an improved estimate (i.e. $\overline{T}'_{t+k_j} \leftarrow \overline{T}_{t+k_j}$) and the elements of the matrix [C] are re-evaluated. This iterative process is continued until the agreement between the estimated temperature distribution $\overline{T}'_{t+k_j}$ and the solution obtained from Eq. (9-64) (i.e. $\overline{T}_{t+k_j}$) agree within a preassigned tolerance. If five cycles of this iteration do not yield convergence the time increment k_j is reduced. The time increment to be used in each successive step is made to depend upon the number of iterations required to achieve convergence of the prior step. In particular, if convergence occurs in 3 or less iterations then the time increment is increased; if convergence occurs in 4 iterations, the time increment is not changed; and if convergence requires five iterations the time increment is decreased. The use of an implicit finite difference formulation makes it possible to assign the time increment size dynamically. This allows the use of large time increments when $q(t)$ is low and only requires the use of small time increments when $q(t)$ is high. When an explicit finite difference formulation of the equations governing the transient heat flow problem is employed, the stability criterion limits the size of the time increments rather severely. Explicit formulation run times for analyses of typical thermostructural panels were found to be about three times as long as the corresponding run times based on an implicit formulation. The use of an implicit formulation and dynamic assignment of time increment size led to analysis efficiency that was essential to successful development of the optimization procedure.

9.3.3 Features of the Structural Analysis

The structural analysis is a linear elastic analysis employing temperature dependent material properties. That portion of the ablator in which the temperature is less than 400°F is assumed to function structurally with the top face sheet of the sandwich. The substructure supporting the ablator is treated as a sandwich with unsymmetrical face sheets.

The bending stiffness of the face sheets is taken into account and transverse shear deformation of the core is considered. It is assumed that only antiplane stress is sustained by the core. It is further assumed that $x_7 \gg x_6$ (see Fig.9.7) and that since the aspect ratio $\frac{x_7}{x_6} \geq 3$ the flat rectangular panel can be treated as a strip with zero curvature in the y direction (i.e. $\frac{\partial^2 w}{\partial y^2} = 0$, see Fig.9.7). It should be noted that the face sheets are biaxially stressed under this assumption. The boundary conditions are assumed to be simple support in bending $\left(\text{i.e. } \frac{\partial^2 w}{\partial x^2} = 0 \text{ at } x = \pm \frac{x_6}{2}\right)$ and free to expand in plane membrane behavior. The structural analysis is described in detail in Appendix B of [9.5].

9.3.4 Features of the optimization procedure

The nine failure modes guarded against (see Section 9.3.1) are all parametric in time or in time and space (i.e. through some portion of the thickness of the shield). The first three failure modes are represented by inequality constraints that are parametric with time as follows:

(1) T at ablator-substructure interface

$$h_1(\vec{D}, t) = \frac{T_s(\vec{D}, t)}{(T_s)_{allow}} - 1 \leq 0 \quad ; \quad 0 \leq t \leq t_f \tag{9-65}$$

(2) T at back of insulation

$$h_2(\vec{D}, t) = \frac{T_B(\vec{D}, t)}{(T_B)_{allow}} - 1 \leq 0 \quad ; \quad 0 \leq t \leq t_f \tag{9-66}$$

(3) panel midpoint deflection

$$h_3(\vec{D}, t) = \frac{w_c(\vec{D}, t)}{(w_c)_{allow}} - 1 \leq 0 \quad ; \quad 0 \leq t \leq t_f \tag{9-67}$$

where $\vec{D}$ represents the vector of design variables and the reentry time period is denoted t_f. These three constraints are of the following form

$$h_j(\vec{D}, t) = \frac{Y_j(\vec{D}, t)}{\tilde{Y}_j(\vec{D}, t)} - 1 \leq 0 \quad ; \quad t_1 \leq t \leq t_2 \tag{9-68}$$

and it should be noted that in general the behavior variable Y_j and its allowable value $\tilde{Y}_j$ may both depend upon the design $\vec{D}$ and the parameter t. The remaining failure modes (4 through 9 in Section 9.3.1) are parametric with respect to both time and space. These six constraints are of the general form.

$$h_j(\vec{D}, t, z) = \frac{Y_j(\vec{D}, t, z)}{\tilde{Y}_j(\vec{D}, t, z)} - 1 \leq 0 \; ; \quad t_{1j} \leq t \leq t_{2j} \quad z_{1j} \leq z \leq z_{2j} \tag{9-69}$$

where $j = 4$ refers to stress in the ablator,
$j = 5$ refers to stress in the outer sandwich face,
$j = 6$ refers to intercell face buckling stress*
$j = 7$ refers to stress in the inner sandwich face,
$j = 8$ refers to tensile strain in the ablator,
and
$j = 9$ refers to compressive strain in the ablator.

It is noted that in general the behavior variable Y_j and its allowable value $\tilde{Y}_j$ may both depend upon the design $\vec{D}$ as well as the parameters t and z. It is also pointed out that the range of values over time $t_{1j} \leq t \leq t_{2j}$ and space $(z_{1j} \leq z \leq z_{2j})$ to which constraint is applied may in general differ for each failure mode (j). In the thermostructural panel example, the time period of interest was the same for all failure modes, namely the reentry time period from $t = 0$ to $t = t_f$. However, the various constraints $(j = 4 \rightarrow 9)$ were parametrically applicable to different regions through the thickness of the shield.

The thermostructural panel optimization problem was formulated using the integrated penalty function scheme previously mentioned in Section 2.6.2. This extension of the Fiacco-McCormick interior penalty function formulation to parametric inequality constraints has the following form as applied to the thermostructural panel problem in [9.5]:

$$\phi(\vec{D}, r_p) = M(\vec{D}) - r_p \left[\frac{1}{t_f} \sum_{j=1}^{3} \int_0^{t_f} \frac{dt}{h_j(\vec{D}, t)} + \frac{1}{t_f} \sum_{j=4}^{9} \frac{1}{(z_{2j} - z_{1j})} \int_{z_{1j}}^{z_{2j}} \int_0^{t_f} \frac{dt\, dz}{h_j(\vec{D}, t, z)} \right] . \tag{9-70}$$

The basic idea of this formulation is that the penalty function is influenced by the behavior constraints at all times $(0 \leq t \leq t_f)$ and at all locations of interest $(z_{1j} \leq z \leq z_{2j})$. Thus, the parameters t and z are accounted for in a natural way, and the entire response, rather than just the critical response, influences the sequence of designs generated. It should be noted that the parametric inequality constraints

$$h_j(\vec{D}, t) < 0 \quad \text{for} \quad 0 \leq t \leq t_f \quad j = 1,2,3 \tag{9-71}$$

and

*not strictly parametric in z.

$$h_j(\vec{D}, t, z) < 0 \quad \text{for} \quad 0 \leq t \leq t_f \quad j = 4,5....9$$

$$z_{1j} \leq z \leq z_{2j} \tag{9-72}$$

must be strictly satisfied at all times and locations of interest if the integrals in Eq. (9-70) are to be proper integrals.

The integrals in Eq. (9-70) are evaluated numerically using the information available from the thermal and structural analyses of a particular design $\vec{D}$. The unconstrained minimizations of $\phi(\vec{D}, r_p)$ are carried out using the variable metric method of Davidon-Fletcher-Powell and finite difference approximations are used to evaluate the gradient $\nabla\phi(\vec{D}, r_p)$ as needed. Care is taken to minimize $\phi(\vec{D}, r_p)$ over the acceptable region in the design space, since $\phi(\vec{D}, r_p)$ is not defined for unacceptable designs. The idea of using approximate or abbreviated analyses is also employed. For example changing the support spacing (x_6 see Fig.9.7) does not require that the thermal analysis be repeated. Also, if the ablator is thick ($x_1 > 2.25$ in) then small changes in the sandwich face sheet thicknesses (x_2 and x_4) do not require repetition of the thermal analysis.

9.3.5 Sample Result

A sample result taken from [9.5] is briefly described in this Section. The trajectory considered in this example is of the ballistic entry type, the thermal input q_c used is for a stagnation point location, and the time period of interest is $t_f = 900$ seconds. The altitude, velocity and cold wall convective heating rate are plotted versus time in Fig.9.8. Note that the maximum q_c is 500 BTU/ft^2 sec at $t = 100$ sec while the maximum dynamic pressure is found to be 1700 lb/ft^2 at $t = 850$ seconds. The materials employed in this example problem are:

(1) ablator - low density phenolic nylon,

(2) sandwich - fiberglass,

(3) insulation - microquartz.

The initial design and the final result obtained are shown schematically in Fig.9.9. The weight per unit surface area protected (the objective function in this example) is reduced from 18.2 lb/ft^2 to 8.56 lb/ft^2 and the total thickness of the shield is reduced from 7.92 in to 3.41 in. The near critical constraints for the terminal design are:

(1) temperature at back face of insulation (limit 660°R)

$$\min_t \; h_2(\vec{D}_{opt}, t) = -0.043 \text{ at } t = 900 \text{ seconds,}$$

(2) panel midpoint deflection (limit 0.24 in)

$$\min_t \; h_3(\vec{D}_{opt}, t) = -0.116 \text{ at } t = 851 \text{ seconds,}$$

(3) temperature at the ablator-sandwich interface (limit 1200°R)

$$\min_t \; h_1(\vec{D}_{opt}, t) = -0.131 \text{ at } t = 900 \text{ seconds,}$$

and (4) ablator stress level

$$\min_t \min_z \; (\vec{D}_{opt}, t, z) = -0.368 \text{ at } t = 370 \text{ seconds.}$$

The design improvement depicted in Fig.9.9 was achieved in 6 unconstrained minimization stages using a FORTRAN IV program on a Univac 1107 machine. The run time was approximately 120 minutes. It is interesting to note that a typical thermal analysis of a trial design required approximately 30 seconds while a structural analysis given the temperature distribution required approximately 5 seconds.

The capability reported in [9.5] is thought to be the first application of the integrated penalty function approach to a structural design problem involving complex parametric failure modes representative of practical application. The capability makes it possible to carry out trade-off studies between weight minimization subject to maximum depth constraints and total depth minimization subject to maximum weight constraints. It is also possible to use this capability to evaluate the relative merits of various combinations of candidate materials, based upon a comparison of optima. This special purpose application [9.5] also illustrates the importance of tailoring the analysis and the design optimization procedure together.

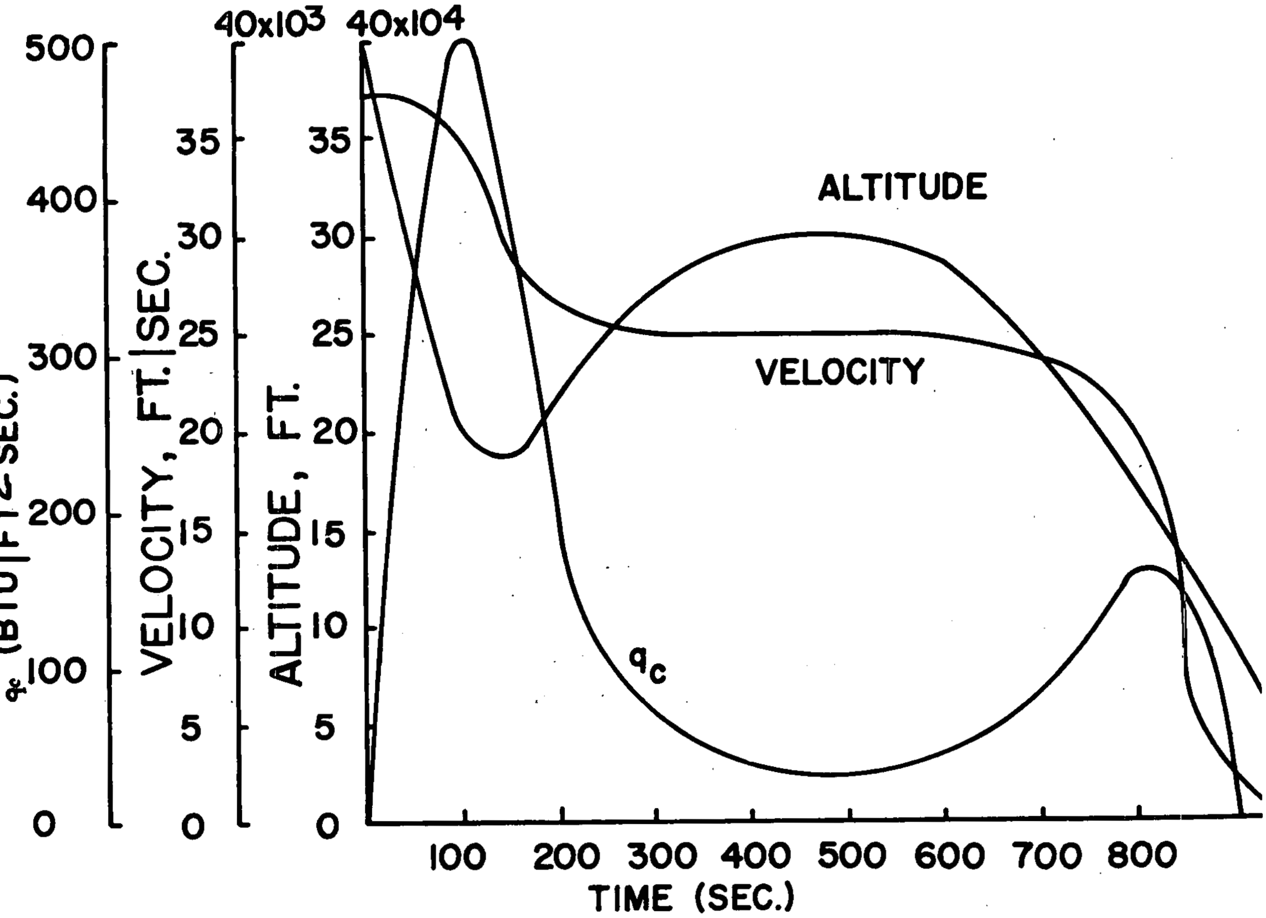


Fig.9.8 Typical Reentry Trajectory and Heating Rate

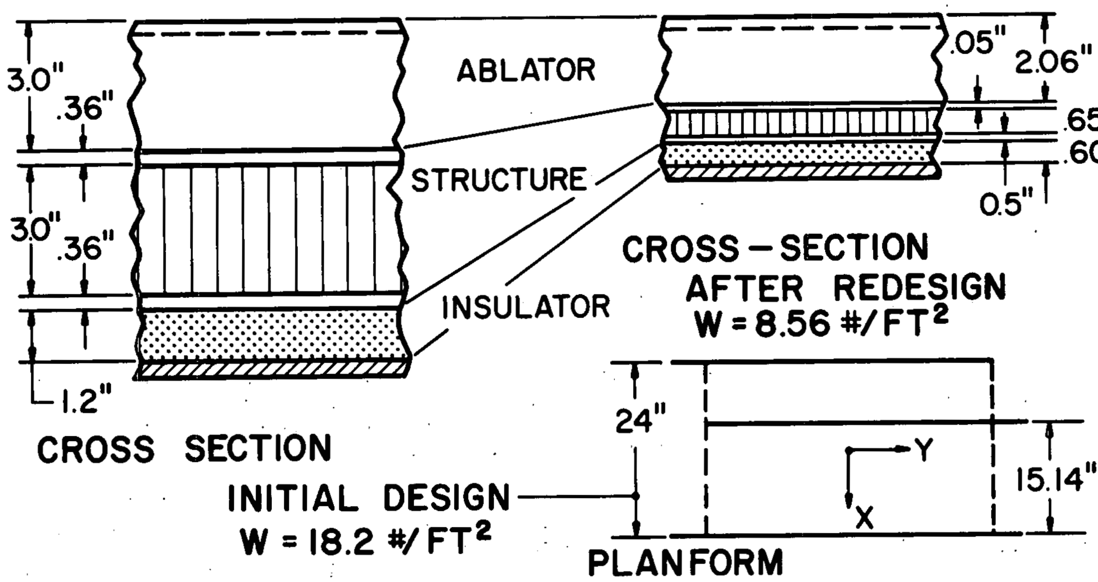


Fig.9.9 Composite Type Heat Shield

List of References

Ref.

9.1 Morrow, W. M. and Schmit, L. A., "Structural Synthesis of a Stiffened Cylinder", NASA CR-1217, December 1968

9.2 Stroud, W. J. and Sykes, N. P., "Minimum Weight Stiffened Shells with Slight Meridional Curvature Designed to Support Axial Compressive Loads", *AIAA Journal*, Vol.7, No.8, August 1969, pp.1599-1601

9.3 Kicher, T. P. and Chao, T.-L., "Minimum Weight Design of Stiffened Fiber Composite Cylinders", AIAA/ASME 11th Structures, Structural Dynamics, and Materials Conference, Denver, Colorado, April 1970, pp.129-145

9.4 Waddoups, M. E., McCullers, L. A., Olsen, F. O. and Ashton, J. E., "Structural Synthesis of Anisotropic Plates", AIAA/ASME 11th Structures, Structural Dynamics, and Materials Conference, Denver, Colorado, April 1970

9.5 Thornton, W. A. and Schmit, L. A., "The Structural Synthesis of an Ablating Thermostructural Panel", NASA CR-1215, December 1968

9.6 Flügge, W., *Stresses in Shells*, Springer Verlag, Berlin, 1966, pp.212-213

9.7 Hedgepeth, J. M. and Hall, D. B., "Stability of Stiffened Cylinder", *AIAA Journal*, Vol.3, No.12, December 1965, pp.2275-2286

9.8 Fiacco, A. V. and McCormick, G. P., "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Non-linear Programming", Management Science, Vol.10, No.4, 1964, pp.601-617

9.9 Cheng, S. and Ho, B. P. C., "Stability of Heterogeneous Aeolotropic Cylindrical Shells under Combined Loading", *AIAA Journal*, Vol.1, No.4, April 1963, pp.892-898

9.10 Chao, T.-L., "Minimum Weight Design of Stiffened Fiber Composite Cylinders", AFML-TR-69-251, September 1969

9.11 Laporte, A. H., "Research on Refurbishable Thermostructural Panels for Manned Lifting Entry Vehicles", NASA CR-638, November 1966

SECTION IV

FUTURE TRENDS AND RESEARCH NEEDS

Chapter 10

OPTIMIZATION OF STRUCTURES WITH RELIABILITY CONSTRAINTS

by

F. Moses

10.1 Introduction

The aim of this work is to explore the relationship between optimum design of structures as it is now formulated in almost 'Classical' terms and reliability or safety of structures. The discussion will focus on the kinds of structures for which reliability or failure probability can reasonably be analyzed and have been presented particularly in a redesign or optimization procedure. As the topic concerns safety in a probabilistic framework some attention must be given to relevant questions of probability sensitivity, failure costs, limited empirical information, analysis errors, and safety philosophy. Several examples of optimization with reliability or failure probability constraints will be presented.

By this time it has become classical on the part of researchers to formulate a structural optimization problem in the following format [10.1]:

$$\text{Minimize} \quad M(\vec{D}) \tag{10-1}$$

$$\text{such that} \quad h_j(\vec{D}) \leqslant 0 \quad ; \quad j = 1,2,\ldots J \quad . \tag{10-2}$$

The $\vec{D}$ are design variables that must be determined. $M(\vec{D})$ is an objective function usually weight or cost although some performance criterion may be introduced. The $h_j(\vec{D})$ are constraints which should also insure the safety of the structure as well as impose fabrication or construction requirements, or any other design rules which the engineer wishes to maintain. In most optimization studies reported in the literature the $h_j(\vec{D})$ constraints include fixed and predetermined safety factors which limit the stresses, deflections and stability coefficients to allowable values. In the best of situations the safety factors have been arrived at in a manner consistent with probabilistic and statistical analyses. This would be done by accumulating data on loads and strength. A load value P_{MAX} could be chosen such that it is not exceeded by any of the measured loads except say once in a hundred times. In a similar way a strength R_{MIN} is chosen such that it is exceeded by say 99.9% of all strength data. Then a safety factor or ignorance factor is introduced which in ultimate strength design is multiplied by P_{MAX} to give P_{ULT} or in working stress analysis is divided into R_{MIN} to give R_{design}. The safety factor expresses the ignorance or uncertainty regarding the stress analysis, fabrication details and other factors. Bouton has pointed out the difficulties in choosing the proper safety factor which has varied for missile and spacecraft from 1.25 to 1.35 to 1.5 as judgement dictated [10.2]. It should be noted that the safety factor values may have more of an effect on structural cost or weight than accurate analysis and optimization procedures. The trend to more rational choice of safety factors is seen in some recent American and European design codes [10.3]. In many cases, however, the safety factors have developed in an evolutionary way giving values which work for existing structures. An important factor, however, is introduced by an optimization approach. This is illustrated in Fig.10.1 which shows a design space with linear constraints and a linear objective function. It can be proven for such a problem as in Fig.10.1 that:

$$\begin{bmatrix}\text{number of active constraints}\\ \text{at the optima}\end{bmatrix} \geqslant \begin{bmatrix}\text{number of}\\ \text{design variables}\end{bmatrix} \quad . \tag{10-3}$$

A similar conclusion results for fully stressed elastic designs in which the number of active stress constraints at the termination of the design iteration equals the number of design variables. From a safety viewpoint the optimization technique has introduced a factor which may be detrimental. It has been pointed out that optimization methods for aircraft and aerospace structures push the design so that 'structural systems' are just barely on the high side of the minimum [10.4]. In the present approach safety will be viewed in a probabilistic sense such that the criterion for safety is the probability of failure, [10.5], [10.6], [10.7], [10.8]. This must recognize that the load environment and strength are random phenomena defined by frequency distributions. For a structural system, failure occurs when loads exceed strengths so that the overall safety or failure probability can be expressed as:

$$\text{Probability of Failure} = \text{Probability}\begin{bmatrix}\text{Any critical member or mode}\\ \text{reaches its capacity}\end{bmatrix} \quad . \tag{10-4}$$

By using conventional non-probability based optimization procedures more members or failure modes will be designed against the limit than if redesign were not done. In the absence of any other constraints this procedure from a probabilistic viewpoint reduces the safety of the structure below that of an unoptimized design. Furthermore the safety factors used to protect various elements against failure in the optimization process have been based mostly on previous experience and practice usually with nonfully-stressed and non-optimized designs. Also, the safety factors may be based on a single element and single load condition combination without regard to system interaction. A severe case of such system interaction would be brittle composite elements of a wing all subjected to the same aerodynamic environmental loading. The greater the number of elements the more likely failure is, unless the safety levels of all members are increased. The optimum way to apportion the increased safety levels is an example of the problems to be considered. Because a clear relationship exists between the safety of structures and a design process incorporating optimization this requires the development of both methods of mathematical programming to do optimum design [10.9] and mathematical methods to compute the expected safety or probability of survival [10.10], [10.11].

There are several problems that must be considered in the context of reliability or probability of failure based design. The first problem is the reliability analysis of structures with derived or assumed probability distributions for the various random variables including load and strength [10.12] but which also may include expansion coefficients and moduli of elasticity. This involves developing and evaluating computational models which account for factors such as indeterminacy, types of failure modes including elastic, brittle, and collapse modes and the numbers of load conditions and failure modes and the system interaction. A second problem is given a reliability or failure probability analysis to design or proportion the members of the structure within the reliability context. This could be the minimum cost or weight design for a specified allowable failure probability [10.13], [10.14], [10.15] or the fixed cost design which minimizes the probability of failure. In a more elaborate framework, it has been proposed to include the cost of failure directly and to find a design which minimizes total overall cost [10.16]. Some of the examples to be presented include multi-member elastic designs (weakest-link structures) and systems designed according to limit design theory ('fail-safe' structures) [10.17]. The approach generally presented herein is to design for a specified allowable overall probability of failure in which the failure probability constraint is evaluated from a sequence of numerical integrations.

In view of the computational and philosophical questions raised by a probability of failure analysis and reliability based design some further attention should be given to the reasons for considering its use. This includes some of the disadvantages of current deterministic approaches and some of the benefits to be realized by incorporating some features of a probabilistic approach to safety and design. It is, of course, recognized that a total attitude and approach to design cannot be put completely on a probabilistic basis since some factors such as expected analysis, manufacturing and fabrication errors are not fully described by probabilistic distribution [10.4]. Nevertheless, in an optimization application probability constraints rather than deterministic constraints will help insure a more balanced and rational design. Other aspects of the problem will now be considered.

1. In order to reach more significant levels of structural optimization it is necessary to compare optimized structures of different configuration, material and geometry. Within this decision context a rational comparison is possible only if the structures have the same level of safety as expressed in terms of probability of failure. This, of course, presumes that the same level of knowledge or data exists for each proposed configuration or system regarding mean levels and variability of loadings and element strengths. Otherwise a Bayesian or subjective approach to be discussed subsequently must be applied.

2. Reliability based optimum design may actually facilitate the mathematical optimization problem by replacing the numerous limitations (on member stress and deflections) in a deterministic design by a single constraint on overall structural failure. The mathematical and computational complexity, however, has been transformed from the design optimization aspect to the analysis of failure probability.

3. The application of new aerospace oriented materials such as ceramic composites, carbon composites, beryllium and molybdenum and the use of thin shell structures leads to improved strength and stiffness characteristics in the mean; however, these materials and structures often exhibit increased strength variability compared with conventional structures [10.4], [10.18]. Failure modes are also more complex often involving fatigue, creep and thermal considerations. This greater strength variability may necessitate such high safety factors that the benefits of the improved material properties will be unrealized unless a direct probabilistic approach is taken. Some current structural applications have also increased the complexity and the extremes of the structural loading environment. Nuclear reactors, deep submergence vehicles, space vehicles and high speed aircraft, for example, are often subject to such broad load spectrums that the picking of a 'worst' possible load condition P_{MAX} is economically meaningless.

4. Another factor is the need to balance the economy of a structure which is only one component of an overall system, which can include electrical, fire control and navigational systems. The allotment of additional costs or weights to various components including the structure to improve overall safety including trade-off between systems can be made economical when reliability including structural reliability is directly expressed as a function of design parameters [10-16], [10-19].

5. In considering a probabilistic approach it should be clarified that this approach to safety can only be applied to those phenomena that can be quantified; namely the treatment of high load and understrength values as random variables. Design, calculation and erection errors or in particular the failure to consider a particular load condition which turns out to be critical cannot be covered by any design format code - deterministic or probabilistic. This should emphasize the continuing need for full scale evaluation of structural behavior both with regard to verifying the structural analysis and also determining if the failure mode phenomena were properly identified. Quality control standards are also needed to insure that additional modes of failure are not introduced during the fabrication and assembly process. A reliability approach, further, does not eliminate the possibility of limitations on the operation of the structure such as maximum wind velocity during launch of a space vehicle or maneuver operations of an aircraft. In such cases the frequency distribution of the loads must be based on proper compliance with the operational limitations. The establishment of an acceptable allowable failure probability should also not be an obstacle to the rational use of the probabilistic approach. A study of existing structures can be undertaken to determine the percentage of failures or accidents in structures which have been due either to overload or understrength factors occurring. An acceptable allowable failure probability due to these factors under control of the structural engineering design code might be established as being of the order 1-10% of the total number of failures including those of construction, fire, blast, etc. beyond the

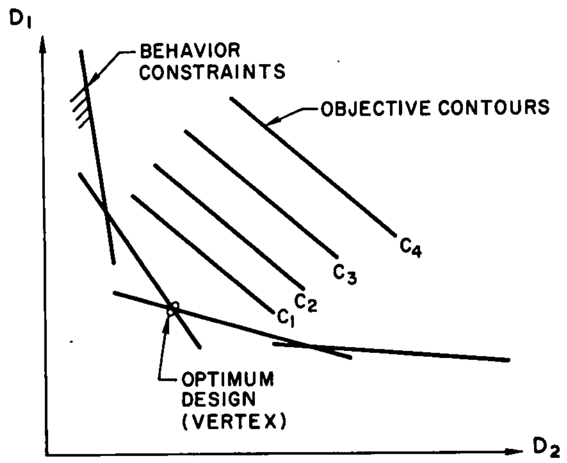


Fig. 10.1 Design Space with Linear Constraints and Objective Function. Optimum lies at a vertex.

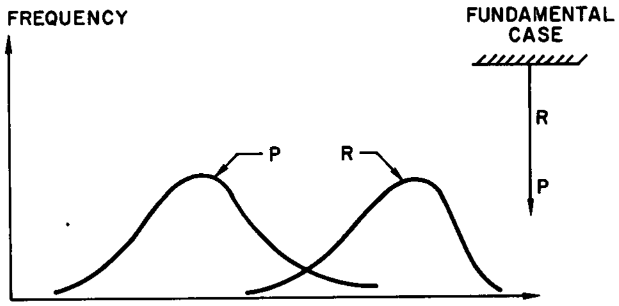


Fig. 10.2 Fundamental Case of Structural Reliability: One Member-One Load.

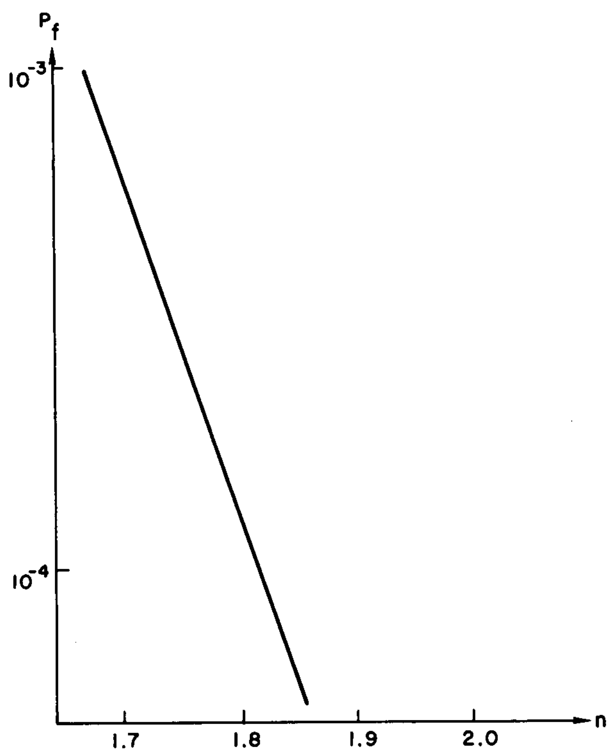


Fig. 10.3 $P_{f,\text{allowable}}$ vs. Safety Factor (n). Fundamental One Member Case: Coef. of Variation Load 20%, Strength 5%, Normal Distribution.

control of structural designers. A similar approach for ships has proposed that structural failure probability should be based on about 10% of the total number of failures expected, the remainder of failures being due to fire, navigation and human errors.

10.2 Reliability Analysis

In formulating a reliability analysis for a structure the first consideration is the structural analysis or failure modes applicable to the design. This means identifying the failure modes and levels of failure to be guarded against. There can be reliability values against yielding, excessive deflections and ultimate collapse. In each case, an appropriate type of structural analysis and failure criterion would be used whether linear or non-linear. For example, with a linear elastic structural analysis the failure criterion would bedefined on the basis of the yielding of any member under any load condition (Weakest-Link Design). This criterion in the case of indeterminate structures ignores the reserve strength that may exist after the yielding of a member. A total reliability analysis of a structural system would include all levels of failure and their associated probabilities of occurrence.

The development of reliability analysis usually begins with what is sometimes called the fundamental case. It consists of a single member of strength R subjected to a load P as shown in Fig.10.2 along with the frequency distributions of R and P. This problem has many of the elements that distinguish structural reliability from other reliability problems in electrical networks and systems. Namely, that both the strength inherent in the design and the load environment are random variables.

Several mathematical and statistical techniques have been used to evaluate failure probability including Monte Carlo, perturbation, and evaluation of integral equations. The Monte Carlo or simulation method involves constructing on a computer trial structures according to generated random numbers and determining the percentage of structures which fail. A large number of trial structures is needed if high confidence is wanted at small failure probability levels. Many investigations have used these methods for such problems as the reliability of rocket engines [10.16] and random vibration [10.20]. The Monte Carlo approach requires considerable calculation but it is useful for complex interelated structural systems or for verification of approximate reliability analyses [10.21].

The perturbation method linearizes the reliability expression and then usually uses a normal distribution approximation. It is especially applicable for problems in which the modulii of elasticity or thermal expansion are also random variables. Linear perturbation has been applied extensively by Diederich, et al [10.6] as in the following example of the reliability of a flat plate buckling under compressive load. Letting P be the applied load, f the critical stress, and n the safety factor, then

$$n = \frac{f\,bt}{P} = K\frac{E}{1-\nu^2}\frac{t^3}{bP} \quad . \tag{10-5}$$

Linearizing about the arbitrary values, n^*, E^*, t^*, P^* gives [10-6],

$$n = n^* + \frac{3}{t^*}n^*(t-t^*) + \frac{1}{E^*}(E - E^*) - \frac{n^*}{P^*}(P - P^*) \tag{10-6}$$

Thus the distribution of n can be constructed from the linear combination of distributions of t, E, and P. Assuming normal distributions greatly simplifies the problem although the Pearson distribution discussed subsequently could also be used [10.21]. The linear perturbation method is best used to find the distribution of strength phenomena which can then be incorporated into finding system reliability.

The third technique of reliability analysis developed extensively by Freudenthal and others [10.10] attacks the reliability evaluation directly by constructing integral equations which must then be evaluated numerically. For example, the probability of failure for the fundamental case is the probability that the load variable exceeds the strength and may be computed from either of the two integrals:

$$P_f = P_r\,(R < P) = \int_0^\infty [F_R(t)]\,f_P(t)\,dt = 1 - \int_0^\infty [F_P(t)]\,f_R(t)\,dt \tag{10-7}$$

where F(t) denotes the probability distribution and f(t) the density or frequency distribution. The reliability R_o is always determined from the failure probability as $1-P_f$.

A plot of P_f vs. n is shown in Fig.10.3 for a typical case where P and R follow the normal distribution with 20% coefficient of variation of load and a 10% on strength. Analysis by Freudenthal and others has shown the effects on P_f of changes in coefficient of variation, central safety factor and the form of the frequency distributions including normal, log normal and extremal functions [10.7]. The results are usually plotted in terms of the safety factor needed to achieve a specific failure probability [10.10], [10.17].

The fundamental case is useful in clarifying the numerical aspects of reliability by indicating the sensitivity of failure probability to input statistical parameters. The fundamental case, however, is only a single element of a complex structure with multi-member multiple load conditions and, therefore, numerous potential failure modes. Some examples of structures more complex than the fundamental one member one load case will now be considered using the integral equation approach.

'Weakest-Link Structures'. These structures fail if any single critical member fails. Such a model is useful for truss or framework like structures in which many elements or members are subjected to a loading of a single origin such as aerodynamic gusts. The model has also been proposed for a heat shield problem in which aerodynamic heating load causes thermal stresses in a vehicle which can cause failure at n points sufficiently separated so material strengths are independent [10.6]. A statistical correlation exists between failure modes because different members may simultaneously fail under the same load condition and the same member may fail under different load conditions. Fig.10.4a shows a single member subject to several load conditions or independent repetitions of a single load. It is easy to verify in this case that the failure probability is:

$$P_f = 1 - \int_o^\infty \left[\prod_{j=1}^{m} F_{P_j}(t)\right] f_R(t)\, dt \quad . \tag{10-8}$$

If there is only one loading but n members as in Fig.10.4b, P_f can be determined from the following equation:

$$P_f = 1 - \int_o^\infty \left[\prod_{i=1}^{n} [1 - F_{R_i}(a_i t)]\right] f_P(t)\, dt \quad . \tag{10-9}$$

This result is often approximated in the form [10.13], [10.15], [10.22], [10.23], [10.24], [10.25]:

$$P_f = 1 - \prod_{i=1}^{n} (1 - P_{fi}) \simeq \sum_{i=1}^{n} P_{fi} \quad , \tag{10-10}$$

where P_{fi} is the failure probability of the ith element. The objection to Eq. (10-10) is not with regard to the fact that P_{fi} is usually small which permits replacing the product term by a sum term but rather the assumption in the product term that the failure modes are independent. Bouton has pointed out that this approximation may have arisen by analogy with certain electrical components in which failure modes are independent [10.4]. Failure modes are not statistically independent for structural systems because the element stresses are completely correlated if they arise from the same load condition. This factor has been shown by several investigations and some results to be presented will show its effect on the optimization process and the minimum weight value.

Eq. (10-9) must be used to give the correct value of the reliability. The constant a_i relates the force or stress level, whichever is appropriate, in member i to the load value P, where t is used as a variable of integration. The a_i can be found from structural analysis methods such as the finite element methods. For indeterminate structures, as in Fig.10.4c, Eq. (10-9) would still be applicable if the 'Weakest Link' criterion of first member yielding is taken to be overall failure. If this criterion is deemed too conservative then a reliability analysis must include the 'fail safe' probability that the structure survives even if some members have failed or yielded. The computational model for an indeterminate structure is complicated because of the numerous alternate load paths and yielding of combinations of members to produce failure [10.26]. One factor, however, is that if the variability or coefficient of variation of the load is greater than that of the strength, there is little fail safe reserve probability. That is,the probability of yielding of any single member is only slightly less than the probability of collapse. This is because proportioning of members is based on a linear relationship between mean load and mean strength. If one member yields then it means a high load value has been reached and if there is small strength variability there is a high probability that other members yield and collapse ensues. Fail safe reserve strength is only expected when the strength variability is relatively large compared to the variability of the load. In addition to the ease of computing 'weakest-link' failure probabilities as compared to 'fail safe' values there is an added factor that most statically indeterminate trusses have many determinate members in addition to indeterminate members so that overall failure occurs if any of the determinate members yield. Thus it is concluded that the reliability for most indeterminate elastic structures can be analyzed by finding the overall probability of any member failing under any load condition. This greatly simplifies the analysis and is also a conservature approach.

If all loads are not independent repetitions of the same load but rather independent load conditions, then P_f could only be determined from multiple integrals [10.11]. Some work has been presented with approximations using only single integrals that include most of the statistical dependence between failure modes due to a single load on many members or a single member under several load conditions [10.11]. Some useful bounds on P_f, however, have been presented [10.27].

These works indicate the feasibility of obtaining exact values or when necessary reasonable bounds on the reliability for 'weakest link' structures which fail if any member fails. The probability of failure is computed from a sequence of single integrals and any form of frequency distribution for load and strength can be used. While 'weakest link' analysis is reasonable for most structures the introduction of ultimate collapse criteria makes it necessary to extend to looking at 'fail-safe' methods of reliability analysis.

<u>'Fail-safe' or Redundant Structures - (Ductile Materials).</u> In structures designed by limit or ultimate design methods or in statically indeterminate structures several members or elements must simultaneously reach their capacity before failure is reached. Some examples are shown in Fig.10.5. A failure mode corresponds to the sum of the independent load contributions exceeding the strength terms. If all the terms are linear this leads to an equation for the random variable of reserve strength, Z_j, in a mode j of:

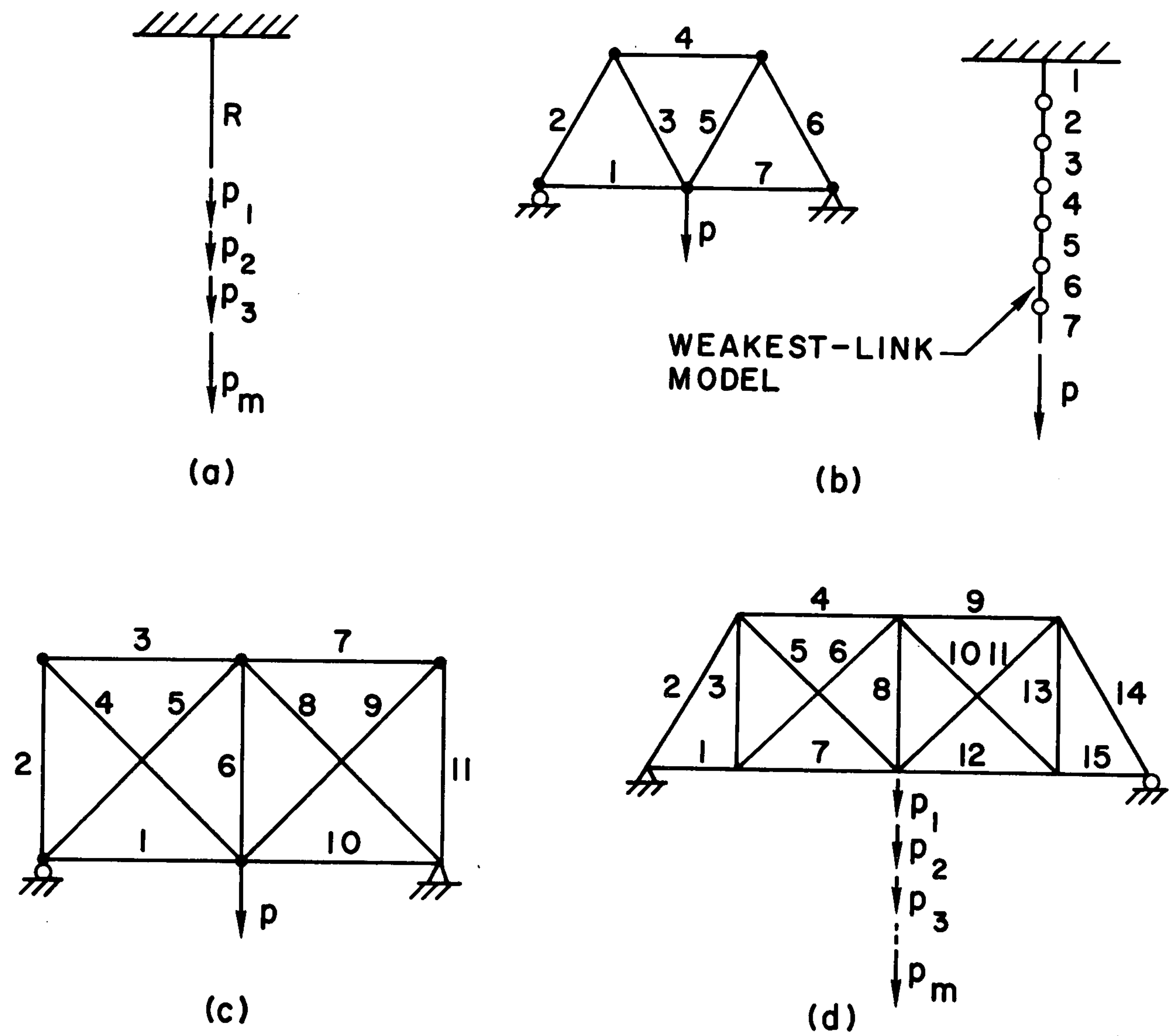


Fig. 10.4 Examples of System Reliability Problems

(a) One Member, m Loads

(b) One Load, n Members

(c) Indeterminate System, One Load

(d) Indeterminate System, m Loads

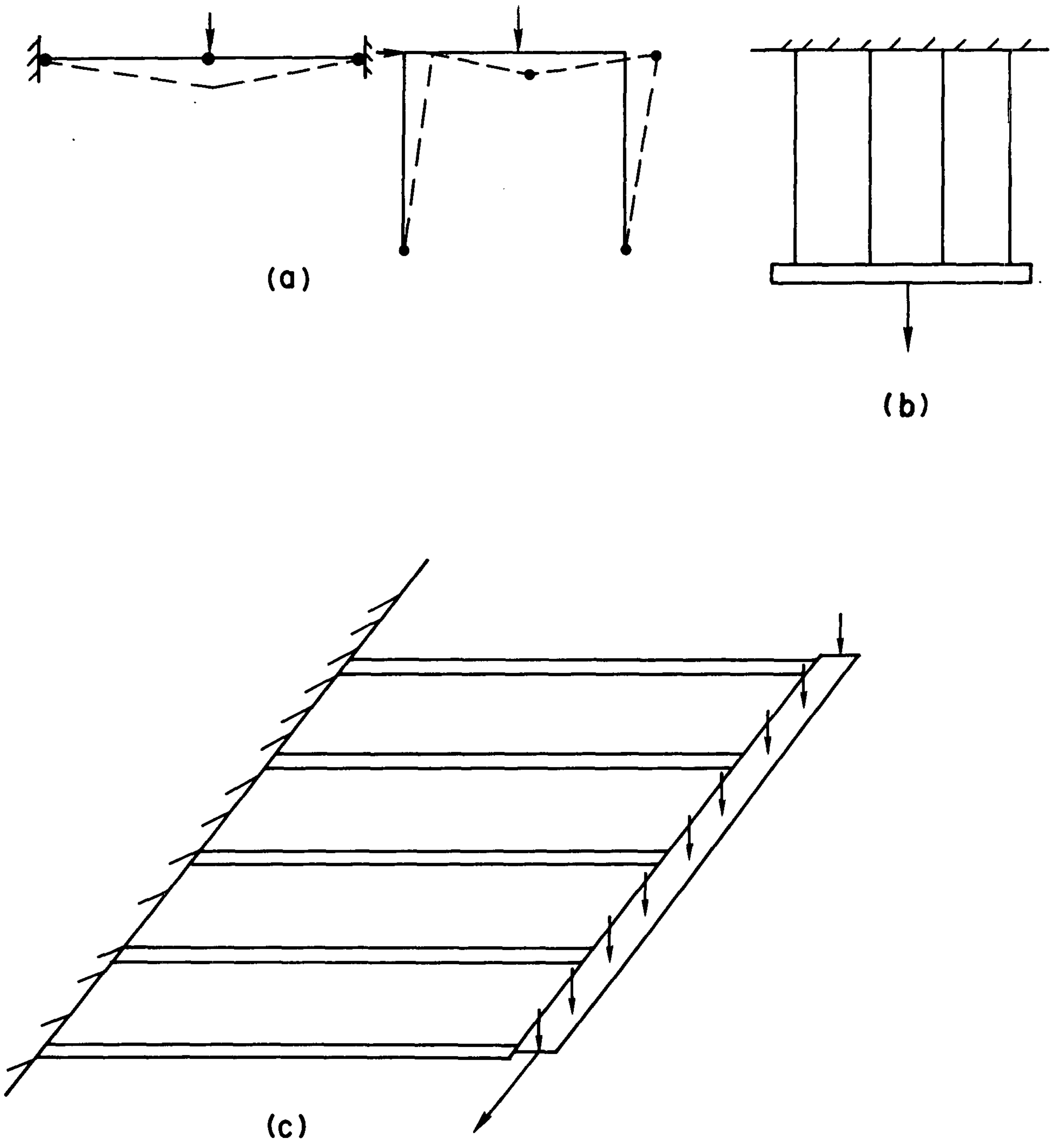


Fig. 10.5 Redundant (Fail-safe Systems)

(a) Beam and Frame Limit Design

(b) Parallel Tension Members

(c) Indeterminate Parallel Beams

$$Z_j = \sum_{i=1}^{n} a_{ji} R_i - \sum_{i=1}^{L} b_{jk} P_k \quad \begin{array}{l} i=1\ldots n \text{ - critical elements} \\ j=1\ldots m \text{ - collapse modes} \\ k=1\ldots L \text{ - loads} \end{array} \tag{10-11}$$

where R_j represents the strength contributions and P_k the load terms. The overall failure probability is the probability that any collapse mode has been reached. Thus:

$$P_f = \Pr[Z_1 \leq 0] + \Pr[Z_2 \leq 0, Z_1 > 0] + \Pr[Z_3 \leq 0, Z_2 > 0, Z_1 > 0] + \ldots \quad . \tag{10-12}$$

Two methods of numerical analysis other than Monte Carlo simulation have been used to compute the probability of a single failure mode occurring [10.21], [10.28]. The first approach was to evaluate directly using recursive integrations the frequency distribution of the reserve strength variable Z_j in Eq. (10-11). Since the terms in this equation are assumed, the distribution of Z_j can be found by successively evaluating the convolution integral numerically. In the specific case where all the R's and P's are normal then Z is also normal. For non-normal distributions a second method for finding the distribution for Z_j uses the Pearson family of distribution functions. This requires the first four statistical moments of the random variables. The results with the Pearson frequency distributions were conclusive and showed good agreement with the recursive integration procedure and Monte Carlo simulation. Normal, log normal and Weibull frequency distribution were studied. A further advantage of the Pearson distribution system is that it can incorporate both the correlation between load terms and strength of element combinations. The load correlation would arise, for example, when an entire structural system is subject to pressure or thermal variation with a known correlation function. Strength correlation could reflect the fact that elements may come from the same manufacturer or be subject to the same fabrication tolerances.

It should also be noted that the computation of collapse mode failure probability could also be done in the case where the terms in Eq. (10-11) are nonlinear as long as they are separable. Another factor to be noted about the collapse mode failure probability is the applicability of the central limit theorem. The sum of independent random variables approaches in the limit a normal distribution. This fact tends to make the choice of frequency distribution for load and strength less important than in the 'weakest-link' analysis. A further point is that the coefficient of variation of the sum of the load and strength terms decreases as compared to the value of an individual term as the number of terms increases.

Fail-safe - (Brittle Material). In many aerospace structures it has been found that increasing economy could be achieved by using brittle materials such as ceramics or carbon composites. In such redundant designs with brittle materials another factor enters the reliability analysis which makes Eq. (10-11) inapplicable. That is the fact that when a brittle member reaches its capacity it ceases to take any load at all. This is also the case with elements that fail through fatigue cracks or exhibit unstable buckling modes. Thus the evaluation of failure probability must consider the order and various combinations in which elements fail. Shinozuka has given for the case of m brittle members and one load condition an expression for P_f which requires evaluating an (m-1)th order integral by numerical integration [10.26]. This is also based on the assumption that all elements have the same strength distribution R. It is apparent from the multiple integrals that an exact reliability analysis for brittle members is limited to at most several members especially when it must be incorporated into an optimization routine. Several factors, however, suggest that statically indeterminate structures with brittle members or unstable elements which cannot maintain their load after reaching a critical value could be incorporated into the weakest-link analysis. One factor is that unless the strength coefficient of variation is relatively large the failure of one member and the redistribution of its load into adjacent members will almost certainly 'trigger' consecutive failures. This has been borne out in some of the computations by Shinozuka and others [10.26]. Another factor is if the load variability exceeds that of strength as is often the case, the failure of a member implies that a high load value has been reached. Since there is a linear relationship between load and stresses the high load reached, indicated by the failure of a member, will cause other members to be highly stressed and fail. In general significant fail safe reserve strength can only be expected when the strength variability is relatively large compared to the variability of the load. The approximation of the reliability analysis of a redundant system by a weakest-link model has also been used for elements which exhibit fatigue failure. This is particularly true if there is not constant inspection to check crack growth. Furthermore many statically indeterminate structures also have some important critical members which are determinate and thus belong in a 'weakest-link' analysis. If there are a large number of redundant brittle elements in an ultimate failure mode then the methods developed for fiber glass and other yarn materials may be applicable [10.29].

Time Dependent Problems and Random Vibration. Most structural reliability analyses have been based on a static approach to loads and strength. An overall viewpoint cannot neglect, however, such factors as stresses or fatigue strength which may be stochastic or time dependent. Numerous cases arise in structural mechanics as in phenomena such as wind, earthquake, vehicle loads, aerodynamic gusts and turbulences and ocean waves, in which the loads and stresses vary with time. When the time variation of loads is significant with respect to the natural period of the structure under investigation this gives rise to random vibration. Also the magnitude of the underlying load carrying phenomena may change over the life of the structure. As an example a structure may be subject to dynamic stresses and vibration due to wind gusting and also during the life of the structure the mean wind may be changing, causing variation in mean response.

Studies of random vibration and stochastic processes involve problems which are directly applicable to the safety and reliability question [10.12], [10.20], [10.30]. Among the results needed are two in particular:

(a) The frequency of occurrence of stress levels. These rates are used to compute the expected fatigue life based on experimental analysis of fatigue specimens along with extrapolation to a load spectrum [10.31].

(b) The probability of reaching a critical response level at any time during the life of a structure. A solution to this first passage problem is needed to predict the failure due to yielding or collapse [10.20], [10.30].

The typical recent results of these investigations have produced curves showing P_f vs. safety factor required. These results can be used to construct frequency distributions for loads to be used in an optimum design procedure with element strength distributions. The subject of optimum design with random loading represents an important area of future investigation.

10.3 Reliability Based Optimization

Optimizing element sizes in a design raises questions as to the meaning of an optimum in the context of a probabilistic model. One alternative minimizes the total cost of the structures, where

$$C(\text{total cost}) = C.I.(\text{initial cost}) + P_f \times C.F.(\text{cost of failure}) \quad . \qquad (10\text{-}13)$$

Letting the failure cost consist of two parts including the cost of reconstruction assumed to be the same as the initial cost and another factor C' which expresses the consequence of failure, leads to the result that the minimum cost P_f allowable should be [10.15]:

$$P_{f,\text{allowable}} = \frac{C}{(C.I.) + C'} = \frac{C}{C'} \frac{1}{1 + \frac{C.I.}{C'}} \simeq \frac{C}{C'} \quad . \qquad (10\text{-}14)$$

Eq. (10-14) shows the approximation if the initial cost is small compared to the consequence of failure as is sometimes the case in aircraft transport and certain other structural vehicle systems.

An alternative approach minimizes P_f subject to an allowable structural weight, so that given an allowable weight the optimum design distributes it to the various elements to minimize P_f of the structure. If the optimum P_f is too large, then either the structure's feasibility or assigned weight must be re-evaluated. The approach generally adopted, however, is to minimize the total structural weight subject to an allowable value of P_f. The present state-of-the-art in estimating failure costs suggests that $P_{f,\text{allowable}}$ be assigned and not determined by the designer as part of the optimization problem. Curves of minimum weight vs. $P_{f,\text{allowable}}$ are, of course, useful and should be considered in trade off studies between different parts of the entire system [10.16], [10.19].

In mathematical programming terminology, the optimization problem is a constrained minimization of the following form:

$$\text{minimize the weight,} \quad W = W(\vec{D}) \qquad (10\text{-}15)$$

subject to the inequality constraint

$$P_f(\vec{D}) \leqslant P_{f,\text{allowable}} \qquad (10\text{-}16)$$

where $\vec{D}$ are the design variables, and $P_f(\vec{D})$ is the failure probability as a function of the design variables.

If there are other constraints based on deterministic factors such as fabrication or construction rules, these may be written as:

$$h_j(\vec{D}) \geqslant 0 \quad . \qquad (10\text{-}17)$$

Eq. (10-15) to (10-17) are similar to the class of structural synthesis problems formulated by Schmit and others but differ in that a single constraint on P_f replaces the numerous constraints on stresses, deflections, and buckling in the usual structural optimization problem [10.1]. The use of the constraint in Eq. (10-16) without regard to optimum weight would apportion $P_{f,\text{allowable}}$ equally to all critical failure modes. If there are n failure modes, then each mode i would be designed for an individual failure probability of:

$$P_{fi} \leqslant \frac{1}{n} P_{f,\text{allowable}} \qquad (10\text{-}18)$$

An optimization approach will achieve greater economy and also provide a better basis of comparing alternative design schemes. A design space concept suggests itself consisting of a multi-dimensional space with a design point $(\vec{D})$ in the space corresponding to the values of the design variables that must be determined. These design variables may be, for example, element areas or beam sizes. A two-dimensional illustration is shown in Fig.10.6 with a constraint curve such that all points on the curve have a reliability value equal to the allowable value and designs lying to one side of the curve have unacceptable failure probabilities. There is also shown an objective function to be minimized which may be weight, cost or some other criterion, and it is also a function of $\vec{D}$. Eq. (10-15) to (10-17) should be a simple mathematical programming problem involving only one behavior constraint which is failure probability. A good minimization procedure is needed, however, because no explicit function for $P_f(\vec{D})$ exists without evaluating integral expressions such as Eq. (10-9), so the total number of redesign points at which P_f is evaluated should be kept to a minimum.

A review of the available inequality constrained minimization methods suggests either a gradient method based on usable feasible directions or a technique which successively linearizes the reliability and weight functions and minimizes using linear programming [10.32]. The problems considered thus far by this author have shown no examples of relative minima.

Examples

'Weakest-Link' Structures. Most investigations of reliability based optimization have used Eq. (10-10) to form the design constraint [10.13], [10.15], [10.22], [10.24], [10.25]. As discussed above, this equation neglects the statistical correlation between failure modes due to their being acted on by the same load conditions. Hilton used a Lagrange multiplier technique to minimize the weight subject to the P_f constraint based on Eq. (10-10), (10-13). Significant weight saving over an equal failure probability for each mode rule, as in Eq. (10-18), resulted because higher failure probabilities P_{fi} were allotted to heavier members than lighter members. Kalaba showed that a dynamic programming formulation could give the optimum member proportions more efficiently than the Lagrange multiplier technique [10.23]. A necessary condition for the dynamic programming method is that the contributions of member failure probability to the overall P_f are independent as in Eq. (10-10). Switzky in an important elaboration of Hilton's approach showed that at the optimum a linear relationship exists [10.15],

$$\frac{W_i \text{ (weight (member i))}}{\Sigma W_i \text{ (total weight)}} = \frac{P_f \text{ (member i)}}{P_{f,\text{allowable}}} \quad . \tag{10-19}$$

The development of Eq. (10-19) was based on several assumptions including static determinacy and linear dependence of the weight function on the design variables, namely,

$$W = \Sigma W_i \quad . \tag{10-20}$$

Using a Lagrange multiplier, λ, on the constraint equation and taking the partial derivatives, gives

$$\frac{\partial}{\partial W_i}[W + \lambda (P_{f,\text{allowable}} - \Sigma P_{fi})] = 1 - \lambda \frac{\partial P_{fi}}{\partial W_i} = 0 \quad . \tag{10-21}$$

Thus at the optimum design point,

$$\frac{\partial P_{fi}}{\partial W_i} = \frac{1}{\lambda}\text{ , constant for all } i \quad . \tag{10-22}$$

If it is also further assumed that a small change in the allowable failure probability does not affect the ratio of member sizes or weight, namely;

$$\frac{W_i}{\Sigma W_i} = \text{constant , independent of } P_{f,\text{allowable}} \quad , \tag{10-23}$$

then Eq. (10-19) follows directly from Eq. (10-22). Using a different P_f weight relationship of the form

$$P_{fi} = a_i e^{b_i D_i} \tag{10-24}$$

instead of the assumption in Eq. (10-22), Murthy gave the following relationship at the optimum [10.33]:

$$\frac{W_i}{W_j} = \frac{P_{fi} \ln (P_{fi}/a_i)}{P_{fj} \ln (P_{fj}/a_j)} \quad . \tag{10-25}$$

A similar result to Eq. (10-19) was found recently by Shinozuka for cases where proof-loading is incorporated in the design process [10.33]

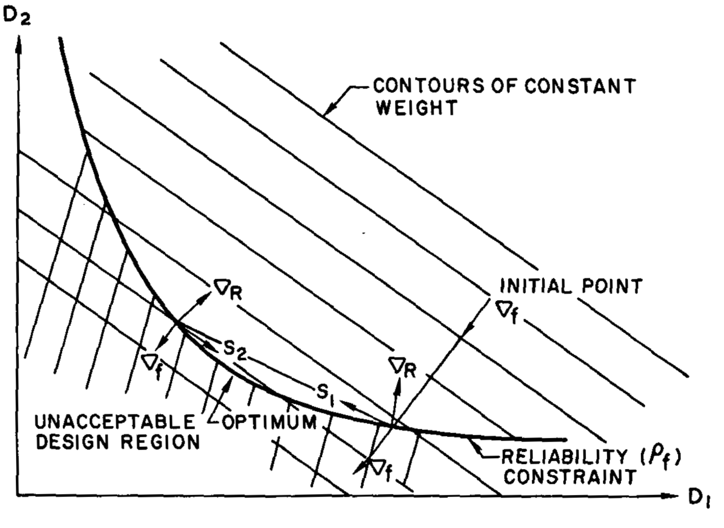


Fig.10.6 Design Space with Reliability Constraint: Arrows Show Minimum Weight Design Method using Usable Feasible Gradient Procedure

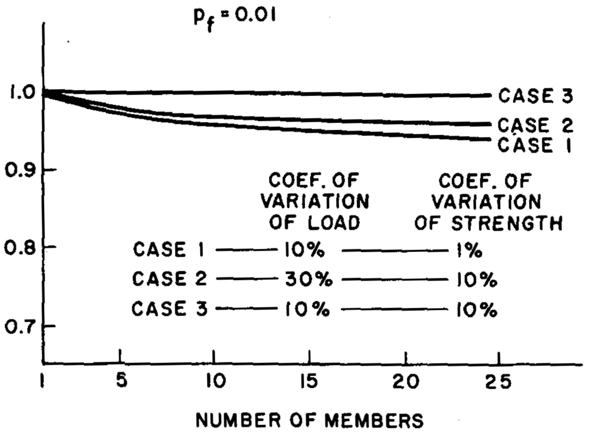


Fig.10.7 Optimum Weight using Exact P_f (Eq. (10–9)) Divided by Design Weight Neglecting Failure Mode Correlation (Eq. (10–19)) vs. Number of Members. Weakest-Link Structure

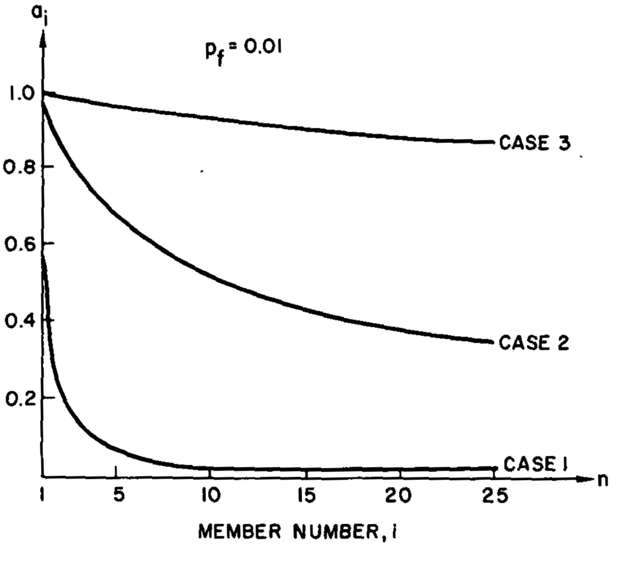


Fig.10.8 Member Influence Coefficient a_i vs. Member Number, Illustrating Failure Mode Correlation (see Eq. (10–28))

$$\frac{W_i}{\Sigma W_i} = \frac{C_i}{C_{allowable}} \tag{10-26}$$

where C_i includes the failure probability term and the probability of failing under the proof-load times a factor which reflects the ratio of cost of element to cost of failure. It should be emphasized that Eq. (10-21) assumes that the structure is statically determinate in which case:

$$\frac{\partial P_{fi}}{\partial W_j} = 0 \text{ , } i \neq j \quad . \tag{10-27}$$

This is not true for statically indeterminate structures since a change in one design element changes the force distribution and therefore the mean load levels and failure probability in other members.

To apply Eq. (10-19) for finding a minimum weight design or either of the other results in Eq. (10-25) and (10-26), a trial and error procedure can be used. For example, assumed ratios of member sizes can be made so that Eq. (10-19) gives P_{fi} for each member. Using this probability value and a relationship between element safety factor and failure probability such as determined in the fundamental case of structural reliability, a member weight W_i is found from the safety factor. The new computed ratios of member sizes are compared to the assumed values. When these ratios converge for all members the process is terminated. Table 10.1 shows the procedure for a 10 member truss example with load and strength frequency distributions being the same as in Fig.10.3. It converges in 2 cycles starting from an initial design based on the values of mean load and equal safety factor in each member. The design process for this example can be done quickly with slide rule calculations for this case since Fig.10.3 is available. For other distribution functions the curves such as those prepared by Freudenthal for the fundamental reliability cases are needed. This example illustrates the feasibility of doing reliability analysis and design for ordinary design practice at least within the assumptions of Eq. (10-19). The weight saved in the design over a uniform safety factor for all members as illustrated in Table 10.2 is obtained by arriving at a design such that heavier members have higher failure probabilities than lighter members.

An investigation is needed, however, of the assumptions in Eq. (10-19) and the possibility of further weight reductions. Recent applications of dynamic programming to several examples indicate that the assumption of Eq. (10-23) is reasonable and that the ratio of member weights is independent of allowable failure probability [10.24]. However, both this latter study [10.24] and the constraint used, namely in Eq. (10-10), neglect the correlation between failure modes which invalidates Eq. (10-19). Some results will be subsequently shown which allow further weight reductions by including this correlation in the computation of the overall failure probability.

The effect of failure mode correlation on the weight has been studied independently by considering a special example [10.14]. All members have equal mean loads and, therefore, the same optimum area. The consideration of the correlation -this is done by using Eq. (10-9) to compute the system failure probability rather than Eq. (10-10) -allows each member to be designed for a higher individual failure probability than if the correlation were ignored. The higher individual failure probability, of course, means a lower weight and the ratio of the optimum weight including the correlation factor (O.W.) to the weight assuming independence of failure modes (I.W.) is plotted in Fig.10.7. For the frequency distribution of load and strength shown the maximum weight reduction reaches 7.3% for case 1 in a 50 element structure. Fig.10.8 shows the effect of correlation when the overall failure probability is written as:

$$P_f = a_1 P_{f1} + a_2 P_{f2} + \ldots + a_i P_{fi} + \ldots + a_n P_{fn} \tag{10-28}$$

where P_{fi} is the probability of failure of the ith member under the single loading. If there were no correlation, and we would have Eq. (10-10), all a's would equal 1.0. If there was complete correlation and the elements were numbered so that the element with the highest individual failure probability were first, then a_1 equals 1.0 and all other a's equal zero. The shapes of the curves in Fig.10.8 depend primarily on the ratio of the load's coefficient of variation to that of the strength and secondarily on the value of the allowable overall failure probability [10.11]. Similar results were shown in [10.11] for $P_f = 0.0001$. The decrease of a_i vs. n shown indicates the important general conclusion that failure probability allotted per member need not be reduced proportionately for an increase in the number of members or failure modes in a structure as in Eq. (10-18). This would only be correct if the load had negligible variability compared to member strength. To consider an extreme case which may in fact be applicable to some aircraft under extreme gust or impact conditions, the conclusion is that if each member be designed for $P_{f,allowable}$ of the structure then the overall P_f will still be $P_{f,allowable}$. For a given shape of the curve of a_i vs. n the amount of total weight saved by incorporating the correlation factor and using Eq. (10-9) as the constraint depends on the number of members and the member's failure probability as a function of its weight. This is affected by frequency distributions and variance as discussed above in the fundamental one member one load case illustrated in Fig.10.3.

An example showing the effect of correlation for structures with unequal mean loads is given in Table 10.2. It is the same example discussed above presented in Table 10.1 based on Eq. (10-10) neglecting correlation. Table 10.2 shows the optimum design including the correlation effect (Eq. (10-9)) compared to designs in which a constant safety factor is used for each member and the design based on Eq. (10-10). The difference between the weights of the optimum design and the equal safety factor design, 2.8%,shown in Table 10.2 is due to both the correlation factor and the unequal proportioning of failure

Table 10.1

OPTIMUM DESIGN USING FAILURE PROBABILITY APPROXIMATION - 'WEAKEST-LINK' STRUCTURE

MEMBER	MEMBER LOAD	TRIAL 1				TRIAL 2			OPTIMUM(f) AREA
		WEIGHT(a)	P_{fi}(b) $\times 10^{-9}$	n(c)	AREA(d) IN^2	P_{fi} $\times 10^{-4}$	n	AREA	
1	0.1P	0.1	0.182	1.92	0.280	0.193	1.915	0.287	0.287
2	0.2P	0.2	0.364	1.875	0.562	0.377	1.87	0.561	0.562
3	0.3P	0.3	0.516	1.855	0.835	0.561	1.85	0.832	0.833
4	0.4P	0.4	0.728	1.835	1.100	0.735	1.835	1.100	1.101
5	0.5P	0.5	0.910	1.82	1.365	0.917	1.82	1.365	1.367
6	0.6P	0.6	1.09	1.81	1.630	1.095	1.81	1.630	1.630
7	0.7P	0.7	1.27	1.80	1.890	1.27	1.80	1.89	1.893
8	0.8P	0.8	1.45	1.79	2.150	1.45	1.79	2.158	2.153
9	0.9P	0.9	1.63	1.785	2.410	1.63	1.785	2.410	2.413
10	1.0P	1.0	1.82	1.78	2.670	1.79	1.78	2.670	2.672
WEIGHT(e):								253.2	253.2

(a) The weight for Trial 1 is assumed proportional to mean load.

(b) $P_{fi} = \dfrac{\text{Weight}_i}{\text{Total Weight}} \times P_{f,\text{allowable}}$ (Eq. (10-19)): $P_{f,\text{allowable}} = 0.001 = \sum P_{fi}$ (Eq. (10-10)).

(c) Safety factor based on fundamental one member-one load case. See Fig.10.3 for these values.

(d) $\text{Area}_i = n_i(\text{mean load } i)/\bar{\sigma}_y$; Mean P = 60000 lb; $\bar{\sigma}_y$ = 40000 psi; Length = 60 in.

(e) Weight: $W = \sum_{i=1}^{10} 0.283\, D_i \times 60.$

(f) See [10.33].

Table 10.2

OPTIMUM DESIGN USING EXACT FAILURE PROBABILITY EXPRESSION INCLUDING CORRELATION - 'WEAKEST-LINK' STRUCTURES

MEMBER	MEMBER LOAD	EQUAL SAFETY FACTORS		OPTIMUM DESIGN(b) NEGLECTING CORRELATION		OPTIMUM DESIGN(c) INCLUDING CORRELATION	
		AREA IN^2	$P_{fi} \times 10^{-4}$	AREA IN^2	$P_{fi} \times 10^{-4}$	AREA IN^2	$P_{fi} \times 10^{-4}$
1	0.1P	0.274	1.0	0.287	0.193	0.297	0.0519
2	0.2P	0.547	1.0	0.562	0.377	0.554	0.604
3	0.3P	0.817	1.0	0.833	0.561	0.818	0.958
4	0.4P	1.09	1.0	1.101	0.735	1.09	0.991
5	0.5P	1.37	1.0	1.367	0.917	1.35	1.23
6	0.6P	1.64	1.0	1.630	1.095	1.61	1.61
7	0.7P	1.92	1.0	1.893	1.271	1.86	2.08
8	0.8P	2.19	1.0	2.153	1.45	2.11	2.65
9	0.9P	2.46	1.0	2.413	1.63	2.35	3.25
10	1.0P	2.74	1.0	2.672	1.79	2.59	3.91
WEIGHT:		255.6		253.2		248.6	

(a) See Table 10.1 for Parameters of Example; $P_{f,\text{allowable}} = 0.001$.

(b) Results from Table 10.1.

(c) P_f computed from Eq. (10-9). Optimum proportioning found from mathematical programming solution of Eq. (10-15) and (10-16).

Table 10.3

OPTIMUM DESIGN - WEAKEST-LINK STRUCTURE INCLUDING PROOF-LOADING

MEMBER	MEMBER LOAD	OPTIMUM DESIGN NO PROOF-LOADING,[b] AREA	OPTIMUM DESIGN[c]	
			$\gamma^{(d)} = 10^{-6}$ AREA	$\gamma = 10^{-4}$ AREA
1	0.1P	0.287	0.257	0.283
2	0.2P	0.562	0.498	0.550
3	0.3P	0.833	0.734	0.812
4	0.4P	1.101	0.966	1.060
5	0.5P	1.367	1.196	1.322
6	0.6P	1.630	1.424	1.573
7	0.7P	1.893	1.65	1.821
8	0.8P	2.153	1.875	2.068
9	0.9P	2.413	2.098	2.313
10	1.0P	2.672	2.320	2.556
WEIGHT:		253.2	221.0	243.9
$P_f^{(a)}$		10^{-3}	0.613×10^{-3}	0.625×10^{-3}

(a) P_f completely based on neglecting correlation in all cases.

(b) See Table 10.2.

(c) See [10.33].

(d) γ is the ratio of cost of element to cost of failure.
Ref. [10.33] also shows optimum levels of proof-load testing.

Table 10.4

OPTIMUM DESIGN RESULTS OF TWO STORY TWO BAY FRAME SHOWN IN FIG.10.10

Example No.	OPTIMUM MOMENT CAPACITIES K-FT						COEF. OF VARIATION		$P_{f,allowable}$	WEIGHT FUNCTION	FREQUENCY DISTRIBUTION
	M_1	M_2	M_3	M_4	M_5	M_6	MOMENT CAPACITY	LOAD			
1	29.2	95.8	84.4	175.0	73.2	74.4	0.10	0.20	7.78(-2)[a]	312.47	NORMAL
2	27.8	96.3	84.4	173.8	72.0	77.9	0.10	0.20	7.80(-2)	312.89	LOG NORMAL
	MONTE CARLO VALUE OF P_f (9500 TRIALS)								7.59(-2)		
3	28.0	78.7	71.0	170.9	69.4	74.9	0.20	0.10	7.72(-2)	297.26	NORMAL
4	27.3	78.3	71.3	166.4	65.1	74.9	0.20	0.10	7.16(-2)	293.53	LOG NORMAL
5	29.1	87.8	72.3	170.3	68.0	74.1	0.15	0.15	7.52(-2)	300.56	NORMAL
	MONTE CARLO VALUE OF P_f (7000 TRIALS)								7.50(-2)		

(a) Exponents of failure probability are shown in parenthesis (m) and should be read as 10^{-m}.

probabilities to the elements as a result of the optimization and mathematical programming solution of the design problem. The additional weight saved by including the correlation factor depends on the ratio of load to strength variance, the number of members and independent failure modes, the allowable failure probability and the frequency distributions used.

Another factor has been introduced into the 'weakest-link' design recently by Shinozuka who included a proof-load testing to sort out weak members [10.33]. This means that a truncated frequency distribution must be used with a lower bound value corresponding to the level reached by the proof load stress. A cost is also introduced to cover the proof-testing which depends on maximum proof-load stress. Optimum design results for the same example discussed in Table 10.2 are shown in Table 10.3 for various ratios of the cost of an element to the cost of failure. It should be noted, however, that the results in Table 10.3 neglect the correlation factor discussed above and express the P_f constraint using Eq. (10-10). Further weight reduction would be shown if the correlation were included in the constraint expression.

A recent study considered reliability based optimum design for redundant structures using Eq. (10-11) as the basis of the reliability analysis [10.21], [10.28]. The specific application was for limit design of frames although the methods are applicable to any redundant structure such as in Fig.10.5 for which the collapse mode equation can be written as a linear combination of load and strength random variables. This includes redundant trusses, grillages and perhaps even plates using yield line analysis, or effective width concepts. It should be noted that any frequency distribution for independent load and strength variables can be handled. A similar study showed optimum design results for frames using loads and strength following the normal distribution laws [10.34]. Some examples of the results are shown in Fig.10.9 for a single story portal frame. There are two design variables corresponding to the plastic moment capacities of the columns M_C and beam M_B which give their respective mean strength values. The examples show optimum material cost or weight with reliability constraints for the single story frame of a unique geometry and loading.

Fig.10.9 also illustrates sensitivity studies which show the effect on the optimum cost due to changes in frequency distributions and their parameters and in the overall specified probability of failure. Cost increases with both allowable overall failure probability and increase in coefficients of variation. To illustrate the application of the reliability based design method for larger structures, the two story two bay frame shown in Fig.10.10 was optimized with a failure probability constraint. There are six design variables including 3 beams and 3 columns. A deterministic optimum design must have at least 6 active collapse modes. Table 10.4 shows some reliability based optima for this case. An interesting observation on some of these results is that the safety factor against collapse in a particular mode is often not a good indication of its probability of occurrence [10.15]. That is, collapse modes compared in the same structure might have higher safety factors based on their mean values but also have higher failure probabilities. This is due to the combination and interaction between random loads and element strengths in a specific collapse mode.

The results further show that the optimum proportioning of structural elements in a large system, with many potential collapse modes, for an allowable failure probability involves a complex interplay of members participating in different collapse modes. The computer is needed for both the reliability analysis of failure probability and the mathematical programming optimization methods for finding the minimum weight design [10.21], [10.28]. The 'fail-safe' design problem contrasts with some aspects of 'weakest-link' design for which in some particular cases a solution near to the optimum design can sometimes be found witn slide rule calculations as in Table 10.1.

10.4 Future of Reliability Based Optimum Design

In the light of these discussions and the results obtained and other studies underway it may be possible to consider the future of reliability based design. Although the designs thus far studied are mostly illustrative they do indicate the problems expected in both analysis of failure probabilities and design based on an allowable probability value. Some specific comments on reliability based design with particular reference to optimization may be based on the theory and results presented in this paper.

1. The results presented indicate the feasibility of using reliability or probability of failure constraints in solving for optimum multi-member structural designs. By using mathematical programming methods to proportion member sizes a design is obtained which has an overall failure probability equal to an allowable value. Examples presented include 'weakest-link' structures for which any member failure constitutes failure of the structure and 'redundant' (fail-safe) structures which fail by forming collapse mechanisms after several members have simultaneously yielded.

2. It is seen from the examples presented that a reliability based optimum design does not end up with equal safety factors for all elements. In a 'weakest-link' structure the heavier members have higher failure probability values than lighter members. This factor is influenced by the degree of statistical correlation between member failures which depends on the ratio of the variability or coefficient of variation of the load to the strength. In optimum 'redundant' structures the same safety factor was not found for each collapse mode at the optimum design. Rather the mathematical programming method proportions each member to achieve minimum weight within the constraint of overall failure probability.

3. An important factor influencing the magnitude of the optimum design as well as its member sizes will be the choice of load and strength frequency distributions and their parameters particularly the coefficients of variation. This includes the effect of proof-loading which has the effect of truncating the strength frequency distribution. Proof-loading actually occurs in all structures since very low strength levels will be detected by inspections or failure under dead weight. Another important factor is the choice of an allowable failure probability. This should depend on the function of the structure as well as the failure consequences in social and economic terms. The fact that many, if not most, structural failures occur because of designer judgements, analysis errors or fabrication oversights

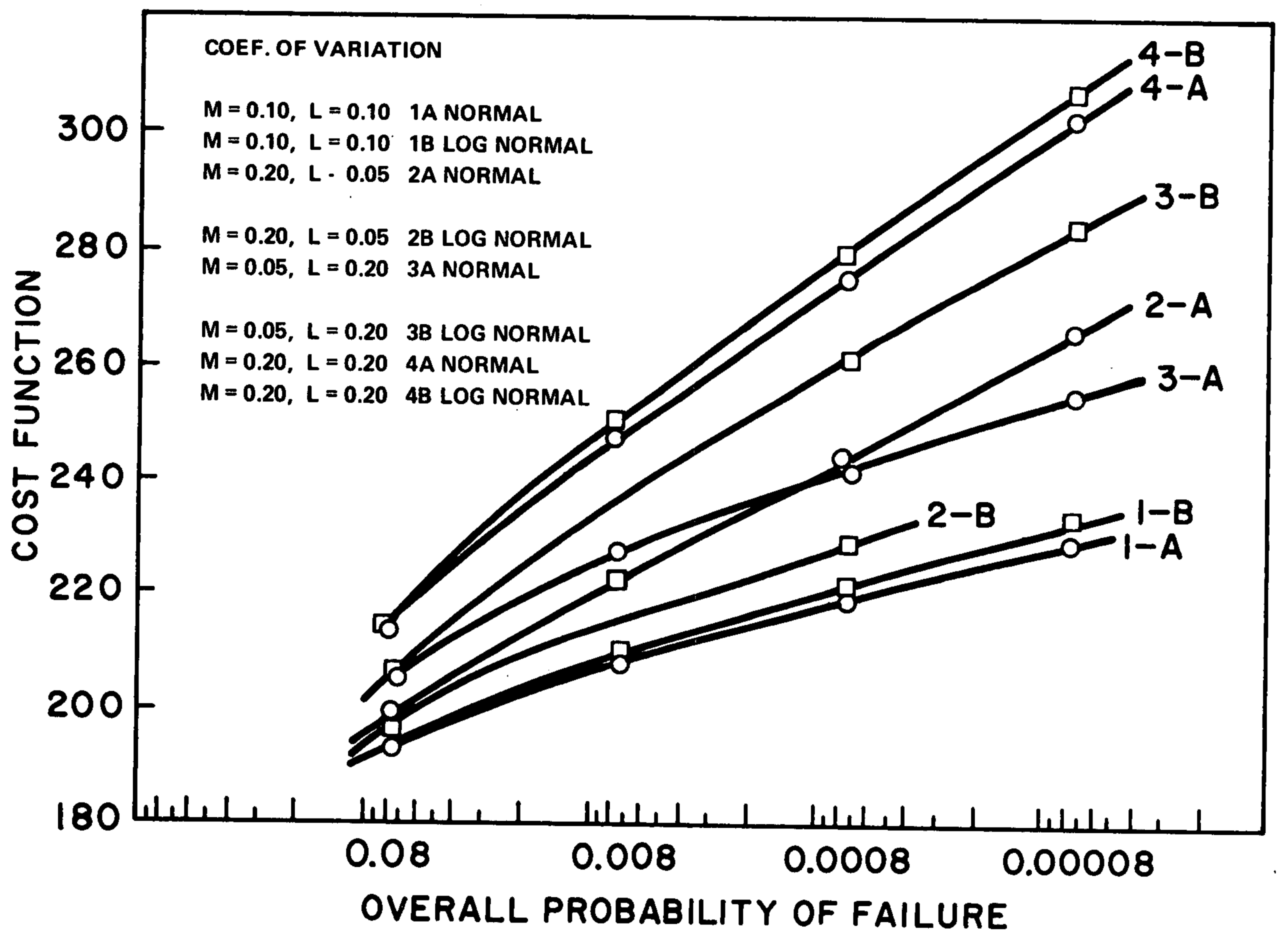


Fig. 10.9 Optimum Structural Costs vs. $P_{f,\mathrm{allowable}}$ for Limit Design of Single Story Portal Frame (Ref. [10.28])

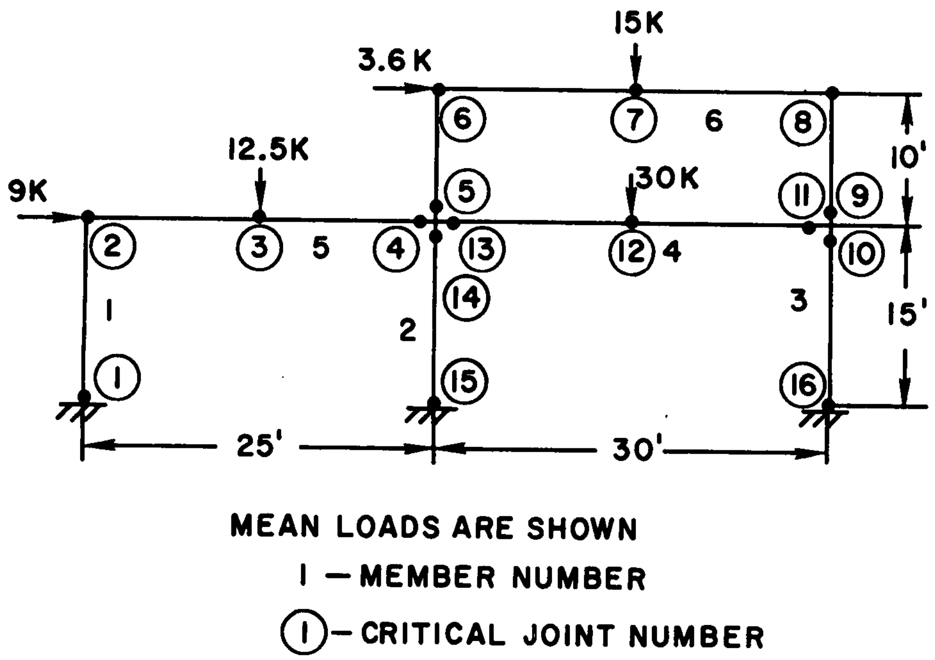


Fig. 10.10 Two Story Bay Frame Example. Optimum Design Results in Table 10.4

introduces some artificiality into a reliability based optimization. Nevertheless the reliability optimization approach is a rational way of distributing the unreliability of members consistent with the information available.

4. A truly optimum design should consider the behavior of the structure over various types of loading operations as well as possible strength deteriorations. In a more extensive approach an optimum design should be found which considers all levels of failure including yielding, formation of cracks, large deflections, fatigue, instability and collapse. Although for some 'weakest-link' structures yielding and collapse occur simultaneously, this is not true for all structures. One approach to this problem would be to assign allowable failure probabilities for each failure type and to seek an optimum design which satisfies all such constraints. Another approach is to combine the constraints into one reliability constraint which would contain the probability of a level of failure occurring multiplied by a factor which includes the associated damage.

5. Despite any forseeable advances in reliability analysis and frequency descriptions of loads and element strengths there will still remain design uncertainties. This arises because of imprecise knowledge or alternatively low statistical confidence in the frequency parameters used in computing P_f. In most cases, there are only estimates of statistical parameters and the final design may require intuitive, subjective, or Bayesian averaging of values by taking groups of applicable data from different sources and noting their coefficients of variation and their associated probability of occurrence. For example, data on buckling coefficients of axially loaded cylinders may show approximately 30% of investigations have 5% coefficient of variation, 40% of investigations, 10% C.V., and 30% of investigations, 15% C.V.. An optimum design can be found for each C.V. value and the weight of the structure determined from weighted averages according to the probability of a C.V. value being encountered. A reliability design procedure described herein, can indicate the savings if more effort and cost is spent to accumulate data to either reduce the uncertainty about the actual variability to be encountered or by controlling the fabrication, and perhaps the operating limitations of the structure so as to reduce the variability itself. Several other approaches have been made to the problem posed by lack of sufficient data. Confidence levels similar to classical statistical analysis have been proposed for aeronautical structures [10.6], [10.8] while the effect of full-scale tests [10.4], [10.35] and proof-load tests have also been considered [10.33].

ACKNOWLEDGEMENT

The author wishes to thank the National Science Foundation for supporting his research which has been carried out under NSF Grants GK-74 and GK-1871 on "Optimum Design of Structures Within a Reliability Philosophy" at Case Western Reserve University.

List of References

Ref.

10.1 Schmit, L. A., "Automated Design," *International Science and Technology*, June 1966, pp.63-78

10.2 Bouton, I., "Implementation of Reliability Concepts in Structural Design Criteria," Fourth Congress, International Council of the Aeronautical Sciences, Paris, France, 1964, (AIAA Paper 64-571)

10.3 Pugsley, A. G., *The Safety of Structures*, E. Arnold (Publishers), London, 1966

10.4 Bouton, I. and Trend, D. J., "Quantitative Structural Design Criteria by Statistical Methods," AFFDL-TR-67-107, Vol.1, June 1968

10.5 Ang, A. H. -S, and Amin, M., "Safety Factors and Probability in Structural Design," J. of the Structural Division, *ASCE*, Vol.95, No.ST7, July 1969, pp.1389-1405

10.6 Diederich, F. W., Broding, W. C. Hanawalt, A. J., and Sirull, R., "Reliability as a Thermo-structural Design Criterion," 6th Symposium on Ballistic Missiles and Space Technology, Vol.1, August, 1962

10.7 Freudenthal, A. M., Garrelts, J. M. and Shinozuka, M., "The Analysis of Structural Safety," J. of the Structural Division, *ASCE*, Vol.92, No.ST1, February 1966, pp.267-325

10.8 Serbin, H., "Reliability and Confidence Criteria in Structural Design," *Aerospace Engineering*, December, 1962, pp.37-40

10.9 Kowalik, J., "Nonlinear Programming Procedures and Design Optimization," *ACTA Polytechnica Scandinavica*, Mathematics and Computing Machinery Series No.13, Trondheim, 1966

10.10 Freudenthal, A. M., "Safety and the Probability of Structural Failure," *Transactions ASCE*, Vol.121, 1956

10.11 Moses, F., and Kinser, D. E., "Analysis of Structural Reliability," J. of the Structural Division, *ASCE*, Vol.93, No.ST5, October 1967, pp.147-164

10.12 Bolotin, V. V., *Statistical Methods in Structural Mechanics*, Holden-Day, San Francisco, 1969, translated by S. Aroni

10.13 Hilton, H. H. and Feigen, M., "Minimum Weight Analysis Based on Structural Reliability," *J. of the Aerospace Sciences*, Vol.27, No.9, September 1960, pp.641-642

List of References (Contd.)

Ref.

10.14 Moses, F. and Kinser, D. E., "Optimum Structural Design with Failure Probability Constraints," *AIAA Journal*, Vol.5, No.6, June 1967, pp.1152-1158

10.15 Switzky, H., "Minimum Weight Design with Structural Reliability," AIAA, 5th Annual Structures and Materials Conference, 1964, pp.315-322

10.16 Blake, R. E., "On Predicting Structural Reliability," 4th Aerospace Sciences Meeting, Los Angeles, California, June 1966, (AIAA Paper 66-503)

10.17 Freudenthal, A. M., "Critical Appraisal of Safety Criteria and their Basic Concepts," IABSE, New York, September 1968, pp.13-25

10.18 Duckworth, W. H., "Designing with Brittle Materials," *Materials in Design Engineering*, Vol.82

10.19 Kluger, P., "Prediction of Design Reliability of Very Large Solid-Rocket Motors," *J. of Spacecraft and Rockets*, Vol.1, No.2, March-April 1964, pp.139-142

10.20 Crandall, S. H., Chandiraman, K. L., and Cook, R. G., "Some First-Passage Problems in Random Vibration," *J. of Appl. Mech.*, ASME, September 1966, pp.532-538

10.21 Stevenson, J. D., "Reliability Analysis and Optimum Design of Structural Systems with Applications to Rigid Frames," Division of Solid Mech., Structures and Mechanical Design Report No.14, CWRU, Cleveland, Ohio, November 1967

10.22 Ghista, D. N., "Structural Optimization with Probability of Failure Constraint," NASA TN D-3777, December, 1966

10.23 Kalaba, R., *Introduction to Dynamic Programming*, Academic Press Inc., New York, 1962

10.24 Khachaturian, N. and Halder, G. S., "Probabilistic Design of Determinate Structures," Proceedings of the Joint Specialty Conference on Optimization and Nonlinear Problems, Engr. Mechanics and Structural Division, ASCE, October 1966, pp.623-647

10.25 Murthy, P. N. and Subramanian, G., "Minimum Weight Analyses Based on Structural Reliability," *AIAA Journal*, Vol.6, No.10, October 1968, pp.2037-2039

10.26 Shinozuka, M. and Itagaki, H., "On the Reliability of Redundant Structures," presented at the 5th Annual Reliability and Maintainability Conference, New York, July 1966

10.27 Cornell, C. Allin, "Bounds on the Reliability of Structural Systems," J. of the Structural Division, *ASCE*, Vol.93, No.ST1, February 1967, pp.171-200

10.28 Moses, F. and Stevenson, J. D., "Reliability Based Structural Design," J. of the Structural Division, *ASCE*, Vol.96, No.ST2, February 1970, pp.221-244

10.29 Gucer, D. W., and Gurland, J., "Comparison of the Statistics of Two Fracture Modes," J. Mech, Phys. Solids, Vol.10, 1962, pp.365-373

10.30 Racicot, R. L., "Random Vibration Analysis Application to Wind Loaded Structures," Division of Solid Mech., Structures and Mechanical Design, Report No.30, CWRU, Cleveland, Ohio, February 1969

10.31 Eugene, J., "The Design Department and the Problem of Fatigue Reliability," Fourth Congress, Intl. Council of the Aeronautical Sciences, Paris, France, 1964

10.32 Zoutendijk, G., *Methods of Feasible Directions*, Elsevier Publishing Co., Amsterdam, 1960

10.33 Shinozuka, M. and Yang, J. N., "Optimum Structural Design Based on Reliability and Proof Load Test," Annals of Assurance Science, Proceedings of the Eighth Reliability and Maintainability Conference, Vol.8, 1969, pp.375-391

10.34 Shinozuka, M. and Hanai, M., "Structural Reliability of a Simple Rigid Frame," Annals of Reliability and Maintainability, Vol.6, 1967

10.35 Bouton, I., Trent, D. J. and Chenoweth, H. B., "Deterministic Structural Design Criteria Based on Reliability Concepts," 4th Aerospace Sciences Meeting, Los Angeles, California, June 1966, (AIAA Paper 66-504)

Chapter 11

OPTIMIZATION UNDER AEROELASTIC CONSTRAINTS

by

H. Ashley, S. C. McIntosh, Jr., and W. H. Weatherill

11.1 Introduction

In the structural design of large, high-performance aircraft, considerations of stiffness and aeroelasticity often play nearly as prominent a role as does static strength. Important examples of phenomena which may influence the sizes of both lifting-surface and fuselage members are the following: primary wing, empennage or control-surface flutter; effectiveness of controls on a flexible wing; influence of elastic deformations on static stability and trim; loads or response in turbulent air; and ride qualities at locations near the extremities of an elongated body. Indeed, cases can be cited where the required margins on flutter speed could be met only through the addition of several thousand pounds of material to a wing which had already met all static-loading design conditions.

In these circumstances, any monograph dealing with the search for optimal airframe configurations should address the question of aeroelastic and structural dynamic constraints. At least this must be done to the degree that such constraints impose uniquely different features on the optimization process. Although techniques for the analytical prediction of aeroelastic properties of a given structure are highly developed [11.1], [11.2], the introduction of such features into formal structural optimization has lagged by several years the use of more conventional conditions of strength, stiffness and stability. Hence the literature is smaller by a substantial factor. To date this literature has tended to remain rather distinct and to focus on simplified one-dimensional problems aimed at revealing what potential improvements _might_ accrue to more realistic structure if practical methods of aeroelastic optimization could be developed. The future will see these constraints appearing more routinely in the 'mainstream' of structural design, but, as of the time of writing, only rather modest published progress in this direction can be reported.

At the outset of this Chapter, one point must be emphasized. It is that, given suitable computational routines for performing the required analysis, the imposition of such a condition as*

$$V_F \geqslant V_o \text{ , a prescribed minimum allowable maximum speed,} \tag{11-1}$$

during the optimal selection of a finite vector of design variables, should be a routine matter. For instance, in an application like that described by Morrow and Schmit [11.3], the inverse of $[V_F - V_o]$ would be added to the penalty portion of their composite function F. When finding the gradient of F, needed for the unconstrained minimization process employed in [11.3], the flutter contribution would be calculated by forward differencing as with their other 'behavior constraints'. The only difficulties one can imagine, beyond those already overcome in [11.3], might arise either while seeking an initial design that meets (11-1) or from the sheer volume of computation inherent in three-dimensional flutter prediction. When well away from the flutter constraint boundary, simplifying approximations might be permissible as with some of the buckling conditions in [11.3].

Despite these observations, the nearest thing to such an application so far published appears to be the wing design described by Schmit and Thornton [11.4]. In their paper, the 'criterion function' chosen for minimization consists of the total propulsive work required to be done against the drag of a rectangular wing, while the wing supports a given payload and flies a series of mission segments at fixed speeds and altitudes. The design is constrained through bounds on airfoil thickness and chord, as well as by limiting values, over each mission segment, for four 'behavior functions': angle of attack at the wing centerline; elastic deflection at the leading edge of the wingtip; principal stress in the skin at the wing root; and Mach number of bending-torsion flutter. Adopting thickness and chord as their two design variables, the authors employ a method called the gradient-steep descent, alternate step method to calculate the optimum. The variables are adjusted during each step in such a way as to move antiparallel to the (numerically differentiated) gradient of the criterion function. This process is continued until a constraining boundary is encountered, whereupon the procedure moves parallel to this boundary until no further reductions in the criterion can be achieved. In the examples presented [11.4], quite reasonable double-wedge airfoil shapes are produced. The propulsive work is also found to depend strongly on the maximum allowable structural weight of the wing. Although these examples tend to be rather elementary, it is clear that the method is capable of considerable generalization.

With regard to the history of the subject, probably the earliest published account† of anything approaching aeroelastic optimization is to be found in a 1953 note of MacDonough [11.5]. Later

---

*Important symbols are defined in Appendix 11A. Here V represents flight speed at an assigned altitude, and V_F is the critical speed of flutter.

†See, however, the remark about rib structures of a fighter airplane in the first paragraph, page 2, of Turner [11.13]. It is believed that the developments reported by MacDonough and Turner received their initial impetus from the work of S. J. Loring. In 1942, an internal company report [11.28] was prepared in which a condition of minimum deflection, for given weight, is found to involve uniform strain energy density for axially loaded members, shear panels and bending elements composed of similar material.

Head [11.6] explained how the ideas of [11.5] had been used for some years at intermediate stages of the design of high-performance aircraft. The problem addressed in [11.5] is to minimize the structural mass of a single-box shell wing while holding constant the fundamental frequency of torsional vibration. The point is made that the critical speed of primary lifting-surface flutter is rather closely determined by this torsional frequency, so that making it the object of optimization tends simultaneously to optimize the structure for flutter. MacDonough states, [11.5], "it can also be shown that the minimum weight of structure to attain a given frequency is approached when the energy per unit volume is constant under the loading associated with the primary mode". Although no proof is given, this is an observation of great insight and agrees closely with certain static and dynamic energy conditions discussed in [11.7] and other references cited therein.

A series of internal reports from North American Aviation, [11.8], [11.9], [11.10] and [11.11], treat the utilization and extension of MacDonough's ideas. They also invoke a condition of uniform shear strain energy density at the torsional divergence speed as another criterion of optimal aeroelastic performance. It is surprising that more complicated and realistic applications have not been undertaken since 1964.

In the evolution of more recent literature on aeroelastic optimization, two fairly distinct points of view are emerging:

(1) The structure is idealized and its degrees of freedom limited by the use of assumed-mode or finite-element techniques. One is thus led to the minimization of an algebraic or transcendental function, by the choice of a finite vector of design variables under algebraic constraints. Refs. [11.3], [11.4] and [11.8] through [11.11] are representative of this approach, as are the more practical examples of Turner, [11.12], [11.13]. Their motivation is to achieve the capability of treating complex built-up structures of the sort encountered in actual flight-vehicle design. Some current work in this area is described in Section 11.3 below.

(2) Simplified, and therefore less realistic, structures are optimized, so that solutions can be found by exact or numerical integration of sets of differential equations. Results published to date have been limited to one-dimensional configurations such as rods, beams and bars. This search for solutions in function space will hopefully make it possible to explore, to the fullest, the potential savings to be gained from formal optimization. There are, as yet, several important theoretical questions (e.g., uniqueness in problems with dynamic aeroelastic constraints) that remain unanswered. They deserve further study in connection with cases whose mathematical description is not too complicated. This approach is discussed first in the following paragraphs and in Section 11.2.

The lead-in to research under the second category may be said to have occurred through analyses of minimum-mass structures with constraints of fixed fundamental natural frequency of vibration. Niordson's paper [11.14] on the simple beam was apparently the first published on such a problem. This approach was continued in the work of Taylor [11.15] and Prager and Taylor [11.16]. These latter articles concern a variety of both static and dynamic problems and contain important proofs of uniqueness and optimality in certain cases. Taylor also suggested [11.15] that in some instances it may be profitable to interchange the roles of the constrained eigenvalue and the merit function. For example, the bar of minimum mass for a fixed fundamental frequency of axial vibration can be found in two ways: one can directly minimize mass for fixed frequency [11.12], or one can maximize the frequency for fixed total mass. Solutions in these cases can be proved to be equivalent [11.15].

Section 11.2 begins with a general discussion of how such problems in a single independent space variable can be identified with the variational problems of Bolza or Lagrange and thus reduced to systems of first-order ordinary differential equations. The merit function in these, as well as more complicated situations is usually chosen to be total system mass or structurally-effective mass. Other criteria, such as minimum mass moment of inertia, may be more suitable in some instances, but little of value will be accomplished here by including such generalizations. One observation worth noting is that all of these optimal designs are subject to unstated constraints, which are really a matter of common sense. They normally have to do with a limitation on the total volume or cross-sectional area that can be occupied by the structure. To illustrate, if one seeks the circular cylindrical column of fixed length and minimum mass to sustain a given Euler buckling load, a zero-mass solution is possible through the application of internal pressure or by allowing the radius to become infinite (i.e., the mass is proportional to the product tR whereas the area moment of inertia grows with tR^3 for a thin shell). Obviously the outer radius must be bounded before the design becomes physically meaningful.

A final observation to be made in this introduction is that energy considerations can often be used for the direct construction of an equation which is, in actuality, the Euler-Lagrange minimizing equation associated with a control or design variable. Relative to the subject of aeroelastic optimization, Prager and Taylor [11.16] gave the first theorems of this type. They studied such extremal problems as the bar with maximum buckling load or maximum fundamental vibration frequency, wherein the control variable enters linearly both the integrand of the merit function and the differential equation of equilibrium. Their theorems are based on the variational principle underlying the latter equation and result in non-linear control equations expressed entirely in terms of the displacement function.

In Table 1 of [11.7] some of these control equations are listed, and it is remarked that these are theorems of 'constant specific Lagrangian density'. For instance, in the case of torsional divergence of an optimal single-box wing, the control equation reads

$$(\theta')^2 = \text{const.} \tag{11-2}$$

in terms of the spanwise derivative of the elastic twist $\theta(x)$. Eq. (11-2) is precisely the aforementioned condition of uniform specific torsional strain energy.

As an illustration of these energy theorems [11.7], consider a three-dimensional elastic solid occupying a volume U and acted upon statically by a system of external forces which are not necessarily conservative. The density of structurally-effective material is ρ, the elastic displacement vector from the unstrained state is $\vec{q}$, and the externally applied force per unit volume is $\gamma \vec{R}$. Here γ is some parameter, such as the dynamic pressure of an airstream impinging on a diverging wing, which is held constant during optimization. $\vec{R}$ may involve contributions both dependent on and independent of the state of (small) deformation. Surface forces like aerodynamic pressure can be included in $\vec{R}$ through terms containing a Dirac delta function of distance from the bounding surface S. All integrals are taken over the unstrained positions of mass elements, in the customary manner of the theory of elasticity.

Hamilton's principle of static equilibrium is

$$\delta \int_U \rho \, e(\vec{q}) \, dU = \int_U \gamma \, \vec{R} . \delta\vec{q} \, dU \tag{11-3}$$

for any small variation $\delta\vec{q}$ satisfying the displacement boundary conditions. $e(\vec{q})$ is <u>independent</u> of ρ and is the quantity called 'specific elastic strain energy' by Prager and Taylor [11.16]. Because of this independence, there is a limitation to structures whose stiffness is directly proportional to structurally effective mass. The reduction to one- and two-dimensional situations is self-evident, and examples of such structures would then be (1) the thin shell in torsion and (2) the sandwich beam or plate, with thin face sheets relative to core depth, in flexure.

Let subscript zero identify a solution optimal in the sense that, for all neighboring density distributions corresponding to the same γ,

$$\int_U [\rho - \rho_o] \, dU \geqslant 0 \quad . \tag{11-4}$$

Hamilton's principle, for the optimal structure under the load system $\vec{R}_o$, reads

$$\delta \int_U \rho_o \, e(\vec{q}_o) \, dU = \int_U \gamma \, \vec{R}_o . \delta\vec{q}_o \, dU \quad . \tag{11-5}$$

It is also a well-known consequence of this principle that, if the structure is strained into the kinematically-admissible deformation shape $\vec{q}_o$, the energy variation will have the right-hand side of Eq. (11-5) as a lower bound:

$$\delta \int_U \rho \, e(\vec{q}_o) \, dU \geqslant \int_U \gamma \, \vec{R}_o . \delta\vec{q}_o \, dU \quad . \tag{11-6}$$

Subtracting Eq. (11-5) from (11-6), one obtains

$$\delta \int_U [\rho - \rho_o] \, e(\vec{q}_o) \, dU = \int_U [\rho - \rho_o] \frac{\partial e}{\partial \vec{q}_o} . \delta\vec{q}_o \, dU \geqslant 0 \quad . \tag{11-7}$$

The meaning of $\partial e/\partial\vec{q}_o$ will become evident from what follows.

For general forms of the energy function $e(\vec{q})$, no obvious conclusion can apparently be drawn from Eq. (11-7). If e is a symmetrical homogeneous quadratic form, however, a useful result is deducible. The quadratic form is the general rule for linear elastic structures. It then follows that one can choose a particular variation $\delta\vec{q}_o \sim \vec{q}_o$ in Eq. (11-7) and employ the familiar relation

$$\frac{\partial e}{\partial \vec{q}_o} . \vec{q}_o = 2e(\vec{q}_o) \quad . \tag{11-8}$$

Eq. (11-8) and (11-7) yield

$$\int_U [\rho - \rho_o] \, e(\vec{q}_o) \, dU \geqslant 0 \quad . \tag{11-9}$$

The only way that Eq. (11-9) and (11-4) can be made consistent for all ρ neighboring the optimum is to require

$$e(\vec{q}_o) = \text{const.} \quad . \tag{11-10}$$

This result encompasses the torsional divergence problem, Eq. (11-2), and a variety of other static aeroelastic cases.

An even more general theorem relating to optimally stiff structures was recently suggested by Taylor [11.17]. The forms of control equations like those appearing in many examples of Section 11.2 below suggest the probable existence of generalizations covering cases of simple harmonic motion, e.g., free vibration and flutter.

11.2 Cases Governed by Ordinary Differential Equations

When the system constraints can all be written as ordinary differential equations, the optimization can be identified as a variational problem of Bolza or Lagrange (e.g., Halfman [11.18]). This is not, however, the only possible formulation. In some instances, it may be more fruitful to pose a variational statement in isoperimetric form (cf. the approach of Taylor [11.15]). As mentioned above, it is assumed that mass will always serve as a suitable figure of merit. It is further assumed that a relation is known between the distributions of structural thickness and stiffness, so that the thickness appears explicitly in the constraint equations.

Reference quantities for the corresponding uniform-thickness system with the same aeroelastic eigenvalue will be used where convenient to render all variables dimensionless. Thus if $T(X)$ is the optimum thickness or running mass distribution and T_o the thickness of the reference system, then the ratio of optimized mass to the reference mass is given simply by

$$m = \int_o^L [T(X)/LT_o]\, dX = \int_o^1 t(x)\, dx \quad . \tag{11-11}$$

Here $X = Lx$, and L is the length of linear structure under study.

Only one-dimensional configurations will be considered in this section. The dependence of the equations of motion on time is eliminated, if appropriate, by the assumption of simple harmonic motion. The constraint equations are then obtained from the aeroelastic equations of equilibrium by rewriting the latter into an equivalent system of first-order ordinary differential equations:

$$q_i' - f_i(q_1,\ldots,q_N,t) = 0 \quad , \quad i = 1,2,\ldots,N \quad . \tag{11-12}$$

The $q_i(x)$ represent the N dependent variables, some of which may be artificially-introduced derivatives of system properties, along with the unknown thickness distribution $t(x)$. A functional is formed [11.18], consisting of the thickness distribution to be optimized, augmented by Lagrange multipliers $\lambda_i(x)$ factoring in the constraint equations:

$$F = t + \sum_{i=1}^{N} \lambda_i(f_i - q_i') \quad . \tag{11-13}$$

Conditions for a stationary value, or extremum, of this functional are given by the Euler-Lagrange equations [11.18]

$$\left.\begin{aligned} &\frac{d}{dx}\left(\frac{\partial F}{\partial q_i'}\right) - \frac{\partial F}{\partial q_i} = 0 \quad , \quad i = 1,2,\ldots,N \\ &\frac{d}{dx}\left(\frac{\partial F}{\partial t'}\right) - \frac{\partial F}{\partial t} = 0 \quad . \end{aligned}\right\} \tag{11-14}$$

This formulation results in $2N + 1$ unknowns - the N q_i, the N λ_i, and t - with $2N + 1$ equations - the $N + 1$ Euler-Lagrange equations plus the N constraint equations.

Boundary conditions are provided for the physical variables q_i by the restraint conditions at the extremities of the structure and for the 'adjoint variables' λ_i by the transversality conditions [11.18]. The equations are non-linear, involving products or quotients of $t(x)$ with certain of the q_i or λ_i. Typically, boundary conditions for the q_i (and the λ_i) are given at both ends of the structure, so the problem is a two-point boundary-value problem. It is usually too complicated to be solved analytically, except in certain simple cases, so a numerical iteration scheme must be employed. Furthermore, there is in general no *a priori* guarantee that a physically meaningful solution exists, nor is there any assurance that a stationary point, once found, represents an absolute optimum. For certain types of constraints, however, such as those on buckling load or on a frequency of free vibration, proofs of optimality can be stated [11.16].

One of the first problems to be solved analytically under what is essentially the formulation described above was that of determining the minimum-weight non-uniform bar with tip mass for fixed fundamental frequency of longitudinal vibration [11.12]. The arrangement is illustrated in Fig.11.1. When the motion is simple harmonic with frequency ω, the axial displacement $U(X)\, e^{i\omega\tau}$ must satisfy the differential equation

$$\frac{d}{dX}\left(M \frac{dU}{dX}\right) + \left(\frac{\omega^2 \rho}{E}\right) MU = 0 \quad . \tag{11-15}$$

Length variables are divided by L and mass per unit length $M(X) = \rho A(x)$ by the reference quantity ρL^2. Eq. (11-15) then becomes

$$(mu')' + \beta^2 mu = 0 \quad . \tag{11-16}$$

The boundary conditions for the restrained root and free end carrying mass $\bar{M}_1$ are

$$\left.\begin{aligned} u(0) &= 0 , \quad u(1) = 1 \quad , \\ (mu')\big|_{x=1} &= \beta^2 \frac{\bar{M}_1}{\rho L^3} \quad . \end{aligned}\right\} \tag{11-17}$$

(Although not required, the deflection amplitude has been normalized to unity at the tip.) The frequency parameter appearing in Eq. (11-16) and (11-17) is

$$\beta = \omega L(\rho/E)^{\frac{1}{2}} \quad . \tag{11-18}$$

Here the dimensionless mass per unit length plays the role of the thickness in the general formulation discussed above. The objective therefore is to minimize

$$\vartheta = \int_0^1 m\, dx \quad , \tag{11-19}$$

subject to fixed β and other physical conditions as stated. Note also that the reference system is obtained by setting m constant in Eq. (11-16) and (11-17).

The Euler-Lagrange differential equations (11-14) are applied to the functional

$$F = m(x) + \lambda_y(x)(-\beta^2 mu - y') + \lambda_u(x)\ (y/m - u') \quad , \tag{11-20}$$

where y is an auxiliary variable proportional to axial force in the bar. This gives rise to the following equations:

$$\left.\begin{aligned} \lambda_y' + \lambda_u/m &= 0 \\ \lambda_u' - \beta^2 m \lambda_y &= 0 \\ \beta^2 u \lambda_y + \lambda_u y/m^2 &= 1 \quad . \end{aligned}\right\} \tag{11-21}$$

The constraint equations bring the total to five, for the five unknowns u, y, m, λ_u, and λ_y:

$$\left.\begin{aligned} y' + \beta^2 mu &= 0 \\ u' - y/m &= 0 \quad . \end{aligned}\right\} \tag{11-22}$$

The physical boundary conditions are given by Eq. (11-17), with y replacing mu', and the transversality condition gives one boundary condition for an adjoint variable:

$$\lambda_y(0) = 0 \quad . \tag{11-23}$$

In accordance with the terminology of optimal-control theory, the third of Eq. (11-21), which relates m algebraically to the other variables, is called the control equation, since m here corresponds to the aforementioned design or control variable. The differential equations (11-21) and (11-22), with boundary conditions (11-17) and (11-23), are in a form amenable to numerical solution by one of several techniques developed in connection with optimal control [11.23]. However, a discussion of these techniques will be deferred, since an analytical solution can easily be found.

The number of unknowns is reduced to three by some elementary manipulations, so that Eq. (11-21) and (11-22) become

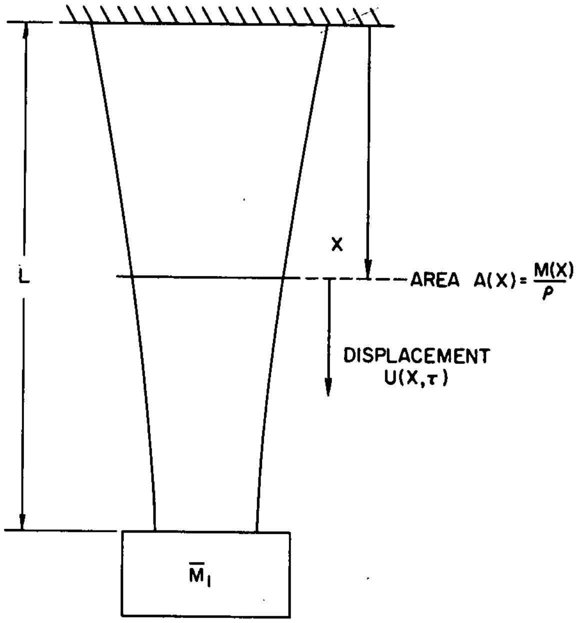


Fig. 11.1 Non-uniform Elastic Bar with Tip Mass

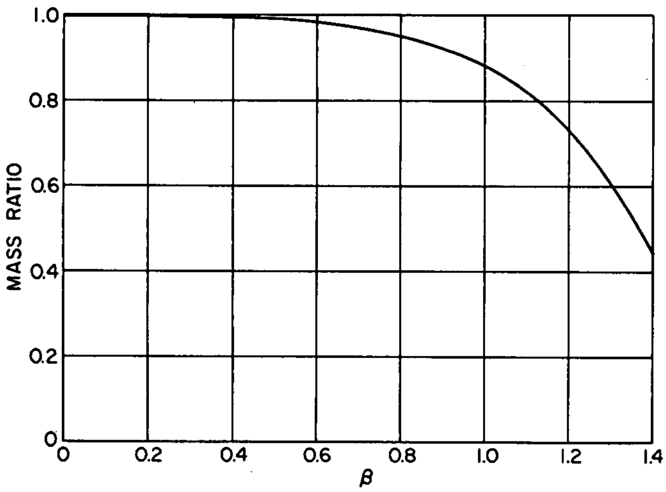


Fig. 11.2 Ratio of Structural Mass of Optimum Bar to that of a Uniform Bar with the same Value of β for the Fundamental Frequency of Longitudinal Vibration

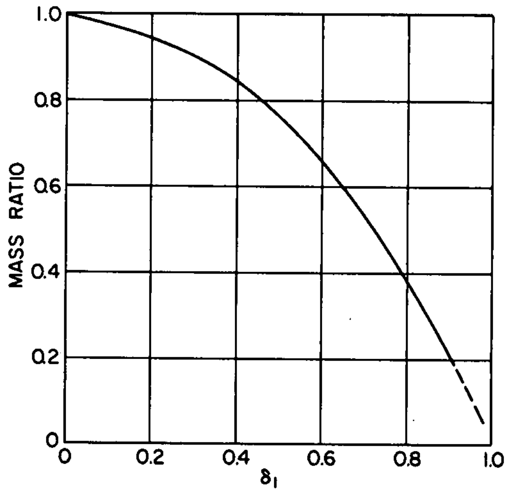


Fig. 11.3 Ratio of Mass of Optimum Cantilever Rectangular Wing to that of a Uniform Wing with the same Fundamental Torsional Frequency, Plotted Versus the Fraction δ_1 of Total Mass Devoted to Skin Material Effective in Torsion

$$\left.\begin{aligned} (m\lambda_y')' + \beta^2 m\lambda_y &= 0 \\ \beta^2 u\lambda_y - \lambda_y' u' &= 1 \\ (mu')' + \beta^2 mu &= 0 \quad . \end{aligned}\right\} \tag{11-24}$$

As observed by Turner [11.12], λ_y and u must satisfy the same differential equation and are governed by the same boundary condition at the origin $x = 0$. Hence they are related by a proportionality factor,

$$\lambda_y = -u/K \quad . \tag{11-25}$$

Therefore the control equation, the second of Eq. (11-24), becomes

$$(u')^2 - \beta^2 u^2 = K \quad . \tag{11-26}$$

Differentiating Eq. (11-26) and dividing by u' produces a linear differential equation whose solution is a linear combination of $\sinh \beta x$ and $\cosh \beta x$. Clearly a hyperbolic sine is indicated, and the form satisfying $u(1) = 1$ is

$$u(x) = \sinh \beta x/\sinh \beta \quad . \tag{11-27}$$

The third of Eq. (11-24) can be integrated to get the mass distribution

$$m(x) = A(x)/L^2 = m(1) \cosh^2 \beta/\cosh^2 \beta x \quad . \tag{11-28}$$

The final boundary condition contributes

$$m(1) \equiv M(L)/\rho L^2 = \beta \bar{M}_1 (\tanh \beta)/\rho L^3 \quad . \tag{11-29}$$

It should be noted that a condition $u'(x) \neq 0$ is assumed during the solution of (11-26), which restricts the fixed-frequency constraint to the fundamental vibration mode. Higher harmonics do not remain fixed as $m(x)$ is varied.

The total mass $\bar{M}_1 \sinh^2 \beta$ of the elastic structure can turn out substantially less than the mass of other bars that would ensure the same fundamental frequency. The ratio of bar mass to tip mass for the optimized structure is simply $\sinh^2\beta$, while the corresponding ratio relative to a uniform bar is $\beta \tan \beta$. The latter ratio can be interpreted as a measure of the weight saving in the optimized bar, as compared to a uniform bar with the same density, length, tip mass, and fundamental frequency. This quantity is plotted versus β in Fig.11.2. As β increases, the weight saving becomes quite significant, although the comparison is not strictly valid as β approaches $\pi/2$.

The value $\pi/2$ of the frequency parameter corresponds to a uniform bar with zero tip mass, for which case the optimum solution is the degenerate one $m \equiv 0$. This situation comes about because the frequency of a uniform bar without a tip mass does not really depend on m, but only on a ratio of two quantities that are both linearly proportional to m. This example is but one of many similar ones that could be cited to illustrate how seemingly well-posed optimization problems do not always yield meaningful results.

In the foregoing analysis it was assumed that all of the mass in the bar itself was available for optimization. From a practical standpoint this is not a very useful assumption, since certain portions of the total mass of a structure do not contribute to rigidity. To illustrate an approximate analytical way of allowing for nonstructural mass, consider a wing of rectangular planform and span L, whose torsionally effective material is concentrated in a single box of fixed cross-sectional shape and size [11.7]. The box thickness $T(X)$ is small compared with its depth; Bredt's formula then shows that the torsional rigidity $GJ(X)$ is proportional to T. Let the uniform reference wing have constant rigidity GJ_o, thickness T_o, and mass moment of inertia I_o (per unit span about the elastic axis). The dimensionless differential equation for the torsional vibration amplitude $\theta(x)$ is

$$\theta'' + \Omega^2 \theta = 0 \tag{11-30}$$

where

$$\Omega = \omega L(I_o/GJ_o)^{\frac{1}{2}} \quad . \tag{11-31}$$

With cantilever boundary conditions

$$\theta(0) = \theta'(1) = 0 \quad , \tag{11-32}$$

one determines for a uniform bar the familiar quarter-sine-wave fundamental mode corresponding to $\Omega = \pi/2$.

For the optimization problem, note that

$$GJ(x)/GJ_o = T(x)/T_o \equiv t(x) \quad . \tag{11-33}$$

Let a fraction δ_1 of the running moment of inertia $I_\theta(x)$ be contained in the skin; let the remaining inertia, which is assumed for convenience to have the same radius of gyration as the skin box, be equal to that of the reference wing. It follows that

$$I_\theta(x)/I_o = \delta_1 t(x) + \delta_2 \tag{11-34}$$

where $\delta_1 + \delta_2 = 1$. The dimensionless differential equation and boundary conditions read

$$\left.\begin{aligned} (t\theta')' + (\delta_1 t + \delta_2)\,\Omega^2\,\theta &= 0 \\ \theta(0) = (t\theta')\big|_{x=1} = 0 \quad , \quad \theta(1) &= 1 \quad , \end{aligned}\right\} \tag{11-35}$$

with $\Omega = \pi/2$ held constant. Note that if all of the mass were assumed concentrated in the skin and therefore torsionally effective, δ_2 would be zero, δ_1 would be unity, and Eq. (11-35) would be directly analogous to Eq. (11-16) and (11-17) with $\bar{M}_1$ equal to zero. It will be seen that the provision of some nonstructural mass is sufficient to ensure a nontrivial optimal solution even when there is no tip mass.

Solution for a minimum value of

$$\vartheta = \int_0^1 t\,dx \tag{11-36}$$

proceeds in the same manner as for the bar. Thus the control equation for the wing can be manipulated to give

$$(\theta')^2 - (\pi/2)^2\,\delta_1\,\theta^2 = K \quad . \tag{11-37}$$

The optimum vibration mode becomes [cf. Eq. (11-27)]

$$\theta = \sinh\left\{\frac{\pi}{2}(\delta_1)^{\frac{1}{2}}\,x\right\}\Big/\sinh\left\{\frac{\pi}{2}(\delta_1)^{\frac{1}{2}}\right\} \quad . \tag{11-38}$$

The thickness distribution is slightly different from that of the bar, because of the difference in boundary conditions:

$$t(x) = \frac{\delta_2}{2\delta_1}\left\{\left[\frac{\cosh\left\{\frac{\pi}{2}(\delta_1)^{\frac{1}{2}}\right\}}{\cosh\left\{\frac{\pi}{2}(\delta_1)^{\frac{1}{2}}x\right\}}\right]^2 - 1\right\} \quad . \tag{11-39}$$

Recalling that masses and moments of inertia have been arranged to be in proportion, one finds for the overall mass ratio

$$\delta_1\int_0^1 t\,dx + \delta_2 = [1 + \{\sinh \pi(\delta_1)^{\frac{1}{2}}\}/\pi(\delta_1)^{\frac{1}{2}}]\;(1 - \delta_1)/2 \quad . \tag{11-40}$$

This expression is plotted versus δ_1 in Fig.11.3. The uniform-wing limit of unity when δ_1 approaches zero is self-evident, whereas the limiting case of δ_1 approaching unity is the unrealistic solution $t \equiv 0$ discussed earlier.

There is yet another unrealistic aspect of the solution (11-39). This involves the fact that at the free end of the wing t goes to zero. The same behavior has been observed in a number of instances where one end of the structure is either free or simply supported and the thickness distribution is unbounded. An obvious means of avoiding this situation is to impose an inequality constraint that forces the thickness to be greater than some specified minimum value. To illustrate the application of this constraint, which could readily have been specified in either of the foregoing examples, consider instead the (relatively rare) occurrence of pure-torsional flutter [11.20]. The differential equation of motion for the torsional amplitude $\bar{\theta}(X)$ (with simple harmonic motion assumed) reads

$$\frac{d}{dX}\left(GJ\,\frac{d\bar{\theta}}{dX}\right) + I_\theta\,\omega^2\,\bar{\theta} = -\bar{M}_X(X) \quad . \tag{11-41}$$

Here $\bar{M}_X$ is the amplitude of the section aerodynamic pitching moment about the elastic axis. With GJ and I_θ both assumed to be proportional to the skin thickness $T(X)$, as in Eq. (11-33), and with incompressible unsteady strip theory used for $\bar{M}_X$, Eq. (11-41) can be written in dimensionless form as

$$(t\bar{\theta}')' + (\delta_1\,t + \bar{\delta}_2)\,\bar{\theta} = 0 \quad . \tag{11-42}$$

Here

$$\left.\begin{aligned} \delta_1 &= (\omega/\omega_\theta)^2 \\ \bar{\delta}_2 &= (\omega/\omega_\theta)^2\,[\bar{M}_\alpha - e(\bar{L}_\alpha + M_h) + e^2\,\bar{L}_h]/\mu r_\alpha^2 \quad . \end{aligned}\right\} \tag{11-43}$$

The terms $\bar{L}_\alpha$, $\bar{L}_h$, $\bar{M}_\alpha$, M_h are dimensionless functions of reduced frequency $k = \omega C/2V$, as tabulated, say, by Scanlan and Rosenbaum [11.19]. The other quantities are defined in any text on aeroelasticity (e.g., [11.19]). The cantilever boundary conditions are as given in the second of Eq. (11-35).

The reference solution for $t = 1$ has the normalized mode shape

$$\bar{\theta}(x) = \sin\{(\delta_1 + \bar{\delta}_2)^{\frac{1}{2}}x\}/\sin(\delta_1 + \bar{\delta}_2)^{\frac{1}{2}} \quad . \tag{11-44}$$

Moreover, the zero-torque condition at the tip requires that $\bar{\delta}_2$ be real and establishes the fundamental eigenvalue

$$\delta_1 + \delta_2 = \pi^2/4 \quad . \tag{11-45}$$

Smilg's solution [11.20] furnishes information on elastic-axis locations and other wing properties that can satisfy Eq. (11-43) and (11-45). In particular, the imaginary part of $\bar{\delta}_2$, which is the component of aerodynamic moment out of phase with respect to the torsional displacement, may vanish only when the elastic axis is ahead of the quarter-chord line.

First, the case of unconstrained thickness is examined. By immediate analogy with the torsional-vibration problem, the optimal solution for real $\bar{\delta}_2$ leads to a thickness distribution similar to that of Eq. (11-39):

$$t(x) = \frac{\delta_2}{2\delta_1}\left\{\left[\frac{\cosh(\delta_1)^{\frac{1}{2}}}{\cosh(\delta_1)^{\frac{1}{2}}x}\right]^2 - 1\right\} \quad . \tag{11-46}$$

In its present form, δ_2 is proportional to the aerodynamic moment in phase with θ; Smilg's calculations show this always to be negative. Thus one arrives at the meaningless result that the 'optimal' $t(x)$ is negative over the whole wing!

It is evident that to produce a viable result requires somehow changing the sign of δ_2. One way of doing this is to allocate a certain portion of the total mass to nonstructural purposes, as was described in the problem of free torsional vibration. If η is the fraction of total cross-sectional mass to be effective structurally, then Eq. (11-42) is altered simply by redefining δ_1 and δ_2:

$$\delta_1' = \eta\,\delta_1\,, \quad \delta_2' = (1-\eta)\,\delta_1 + \delta_2 \quad . \tag{11-47}$$

Radii of gyration are taken equal, as before. A number of parameter combinations can be found which produce a positive δ_2'. One case was studied from Smilg [11.20] in which the rotational axis was at the leading edge and the flutter k (defined in Appendix 11A) was approximately 0.04. With 50% of the mass in the skin of the reference wing ($\eta = 0.5$), δ_1' and δ_2' are calculated to be 2.04 and 0.43 respectively. A computation similar to that indicated in Eq. (11-40) then shows a 39% saving in total mass and a 78% saving in skin weight achieved by going from the uniform wing to the optimum wing with the same flutter speed.

The unrealism of t going to zero at the tip still remains. As suggested above, one can introduce a constraint to keep the thickness greater than or equal to some minimum value. There are a number of ways of accomplishing this; a convenient one that will be followed here was employed by Taylor [11.21]. The constraint is stated

$$t(x) - t_o - a^2(x) \geq 0 \tag{11-48}$$

where t_o is an arbitrary minimum thickness and $a(x)$ is a real function to be determined. Since the reference wing is given by $t \equiv 1$, it follows that t_o must lie between zero and unity.

The functional for this problem becomes (for real $\bar{\delta}_2$)

$$F = t + \lambda_\theta(s/t - \theta') + \lambda_s[-(\delta_1 t + \delta_2)\theta - s'] + \lambda_t(t - t_o - a^2) \quad . \tag{11-49}$$

The Euler-Lagrange differential equations are as before, except for the addition of new variables λ_t and a^2:

$$\left.\begin{aligned} \lambda_\theta' - (\delta_1 t + \delta_2) &= 0 \\ \lambda_s' + \lambda_\theta/t &= 0 \\ 2\lambda_t a &= 0 \\ \lambda_\theta s/t^2 + \delta_1 \lambda_s \theta - \lambda_t &= 1 \quad . \end{aligned}\right\} \tag{11-50}$$

The system of differential equations is completed by the constraint equations

$$\left.\begin{aligned} \theta' - s/t &= 0 \\ s + (\delta_1 t + \delta_2)\theta &= 0 \quad . \end{aligned}\right\} \tag{11-51}$$

From the third of Eq. (11-50), it is seen that either λ_t or a^2 must be zero. Choosing zero λ_t leaves the thickness unconstrained, whereas choosing zero a^2 requires the thickness to be a constant, t_o. One supposes that outboard of some station x_o, $0 \leq x_o \leq 1$, one can choose zero a^2, or $t = t_o$. Inboard of x_o one sets $\lambda_t = 0$ and allows the thickness to vary.

Physical boundary conditions plus transversality at the wing root are

$$\theta(0) = \lambda_s(0) = 0 \quad , \tag{11-52}$$

while at the tip

$$s(1) = 0 \quad , \quad \theta(1) = 1 \quad . \tag{11-53}$$

At x_o, the Weierstrass-Erdmann corner conditions [11.18] require continuity of all variables. In part, these requirements can be manipulated to give

$$\theta(x_o) \; , \quad \theta'(x_o) \; , \quad t(x_o) \text{ continuous} \quad . \tag{11-54}$$

The solution for the inboard section $x < x_o$ proceeds as before, giving for θ and t

$$\left.\begin{aligned} \theta &= A \sinh\{(\delta_1)^{\frac{1}{2}} x\} \\ t &= t_o + a^2 = B\,[\cosh\{(\delta_1)^{\frac{1}{2}} x\}]^{-2} - \delta_2/2\delta_1 \end{aligned}\right\} \; x \leq x_o \quad . \tag{11-55}$$

Here A and B are arbitrary constants yet to be determined. The solution for the outboard strip, $x \geq x_o$, is found from Eq. (11-50) and (11-51) with $t = t_o$ = const. In particular, the normalized result for θ reads

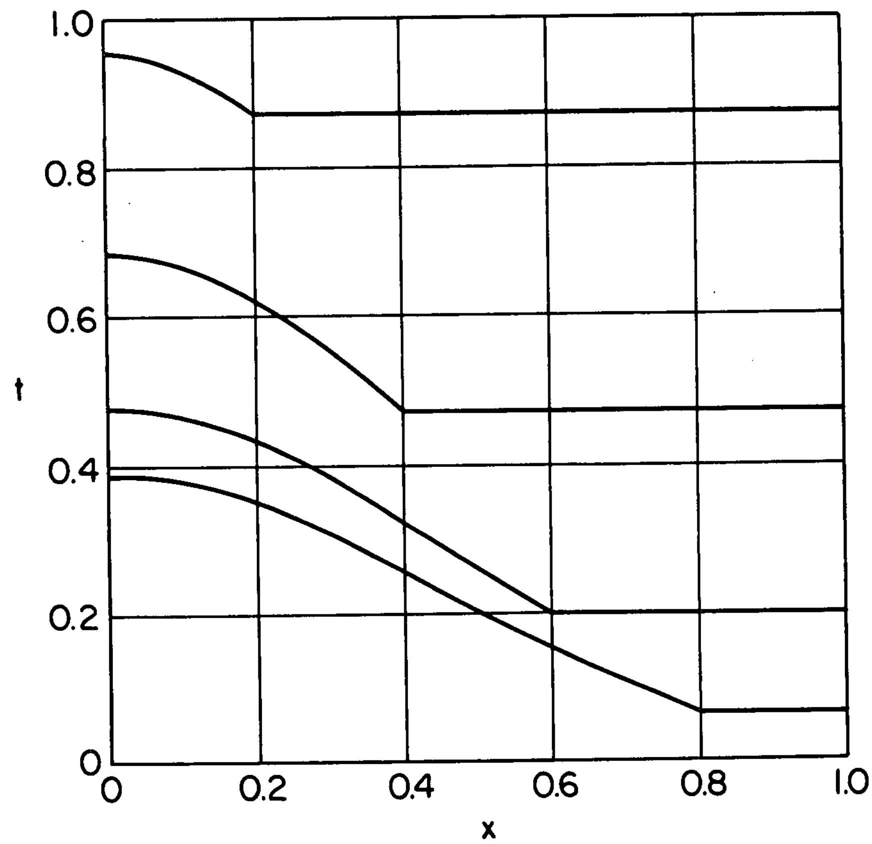


Fig. 11.4 Optimum Dimensionless Skin Thickness Distributions for Pure Torsional Flutter of a Rectangular Cantilever Wing, with δ_1 = 0.5 and Various Minimum Values of Skin Thickness

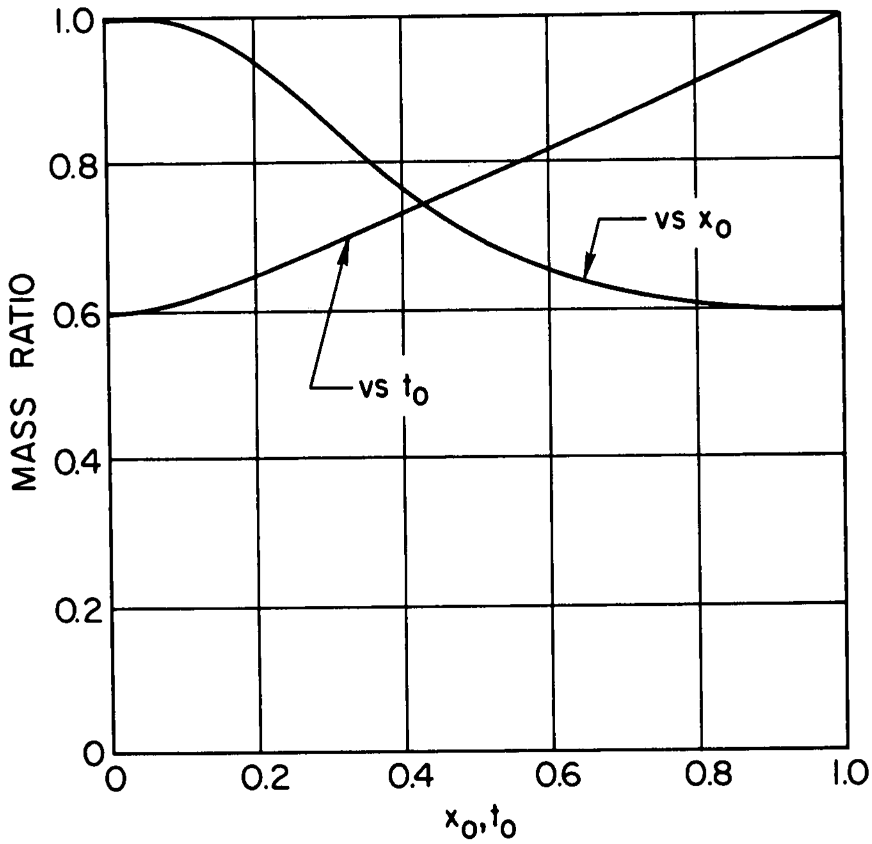


Fig. 11.5 Mass Ratio Versus Minimum Skin Gauge t_o for the Wings of Fig. 11.4. The same Ratio is also Plotted Versus x_o , the Point where Minimum Thickness begins

$$\theta = \sin\gamma \sin\gamma x + \cos\gamma \cos\gamma x\ , \quad x \geqslant x_o \tag{11-56}$$

with

$$\gamma^2 = \delta_1 + \delta_2/t_o > 0 \quad . \tag{11-57}$$

Note here that, if δ_2 is negative, the requirement for positive γ^2 puts a constraint on the minimum thickness:

$$t_o > -\delta_2/\delta_1 \quad . \tag{11-58}$$

There are now four unknown constants - t_o, x_o, A and B - and three continuity conditions (11-54) to relate them to each other. The simplest way to proceed is to eliminate A and B and arrive at a transcendental relationship between x_o and t_o, which is

$$\gamma \tan\{\gamma(1 - x_o)\} = (\delta_1)^{\frac{1}{2}} \coth\{(\delta_1)^{\frac{1}{2}} x_o\} \quad . \tag{11-59}$$

Once x_o or t_o is chosen, the other is found from Eq. (11-59). A simple integration of the optimal t over the span then yields the ratio of the mass of the optimized wing to that of the uniform reference wing. It turns out that the constraint on t_o, Eq. (11-58), is not enough to produce a reasonable answer. When $\delta_2 < 0$, the thickness is no longer negative, but the optimal mass is greater than the mass of the uniform reference wing. Only for positive δ_2 is a saving in mass realized by the optimal solution. It is still necessary, therefore, to allow some of the mass to be nonstructural.

Numerical results for the case discussed above, where 50% of the mass was assumed nonstructural, are shown in Figs.11.4 and 11.5, adapted from [11.7].

It should be evident from the foregoing that only rather simple optimization problems can be solved analytically. The introduction of more complicated aeroelastic constraints, even in the function-space framework to which this section is devoted, of necessity implies that numerical solution techniques have to be used. One of the first examples which served to validate a particular numerical procedure for integrating the optimizing equations was that of minimizing the skin weight of a constant-chord unswept wing of fixed torsional divergence speed (Fig.11.6; [11.22] and [11.7]). The constraint differential equations and boundary conditions for the problem are (with aerodynamic induction neglected)

$$\left.\begin{aligned} \theta' - s/t &= 0 \\ s' + \Omega^2 \theta &= 0 \\ \theta(0) = s(1) &= 0 \end{aligned}\right\} \tag{11-60}$$

$$\Omega^2 = \frac{\rho V_D^2 CEL^2}{2GJ_o} C_{L_\alpha} = \pi^2/4 \quad . \tag{11-61}$$

The Euler-Lagrange equations for the function

$$F = t + \lambda_\theta (s/t - \theta') + \lambda_s (-\Omega^2 \theta - s') \tag{11-62}$$

are found to read

$$\left.\begin{aligned} \lambda_\theta' - \Omega^2 \lambda_s &= 0 \\ \lambda_s' + \lambda_\theta/t &= 0 \\ \lambda_\theta s/t^2 &= 1 \quad . \end{aligned}\right\} \tag{11-63}$$

Transversality produces two further boundary conditions,

$$\lambda_s(0) = \lambda_\theta(1) = 0 \quad . \tag{11-64}$$

An exact solution to this problem is easily discovered by following the same reasoning that led to solutions of the previous examples. Furthermore, a minimum-thickness constraint can be introduced just as was done for the torsional flutter problem, leading to a transcendental relation between x_o and t_o as follows:

$$x_o = [(t_o)^{\frac{1}{2}}/\Omega] \cot [(1 - x_o)\, \Omega/(t_o)^{\frac{1}{2}}] \quad . \tag{11-65}$$

As an aside, it is remarked that an interesting aspect of Eq. (11-65) is the possibility of multiple solutions. That is, for a small enough value of t_o (or a value of x_o close enough to unity), a second branch of the cotangent curve, the branch for arguments between $\pi/2$ and $3\pi/2$, also yields an x_o, t_o combination. As t_o becomes still smaller, at some point a third branch, for arguments greater than $3\pi/2$, comes into play, and so on. It therefore appears that an infinite number of optimal solutions can be found. Furthermore, the corresponding thickness distributions can be made virtually as small as desired by selecting the proper branch. An eigenvalue analysis of these 'optimal' wings reveals, however, that solutions associated with the second cotangent branch have their fundamental characteristic speed of divergence below that corresponding to the eigenvalue of $\pi/2$ held fixed in the analysis. In fact, the number of eigenvalues below $\pi/2$ in any given solution turns out to equal the number of branches, of the cotangent curve in Eq. (11-65), taken beyond the fundamental. Hence the only truly minimum-mass solutions are those found with arguments of the cotangent less than $\pi/2$. There is an obvious conclusion that every solution of this sort should be carefully examined, before it is accepted, to ensure that all constraints on the optimization have been satisfied.

As mentioned above, a computational check on the solution of the system of equations and boundary conditions (11-60), (11-63) and (11-64) was carried out by Ashley and McIntosh [11.7]. A transition matrix algorithm was adapted from Bryson and Ho [11.23] for purposes of numerical integration. In this relatively simple case, direct numerical differentiation was successfully carried out for the purpose of determining the required elements of the transition matrix. Essentially exact agreement with the known solution was attained after about half-a-dozen iterations.

Unfortunately, the rather attractive transition-matrix scheme has proved too inaccurate, in the absence of special refinements, for more complicated problems. A procedure involving the determination of 'unit solutions' [11.23] turns out to yield much more precise transition matrices, although considerably more computer programming is needed. It operates as follows:

The above differential equations are all seen to be in the form [cf. Eq. (11-12)]

$$\left.\begin{aligned} \frac{dy_i}{dx} &= f_i(y_j, t) \qquad i = 1,N \\ t^2 &= g(y_i) \qquad\quad j = 1,N \end{aligned}\right\} \tag{11-66}$$

where the y_i's and y_j's are either the physical variables (such as θ) or their adjoints, λ. If these differential equations are perturbed by means of small changes δy_i to all dependent variables, a new set is created. These additional equations may be written as

$$\left.\begin{aligned} \frac{d}{dx}(\delta y_i) &= \sum_{j=1}^{N} \frac{\partial f_i}{\partial y_j}\delta y_j + \frac{\partial f_i}{\partial t}\delta t \\ \delta t &= \frac{1}{2t}\sum_{i=1}^{N} \frac{\partial g}{\partial y_i}\delta y_i \quad . \end{aligned}\right\} \tag{11-67}$$

The combined equations (11-66)-(11-67) can be solved simultaneously, using appropriate boundary conditions, to produce the transition matrix for the system. For instance, if the boundary conditions

$$\left.\begin{aligned} \delta y_i(0) &= 0 \qquad i \neq 1 \\ \delta y_1(0) &= 1 \end{aligned}\right\} \tag{11-68}$$

are chosen for the perturbation equations, while the usual specified and 'guessed' boundary conditions for $y_i(0)$ are used, the two sets of equations may be numerically integrated from $x = 0$ to 1. The values of $\delta y_i(1)$ are equal to the changes in the variables y_i at $x = 1$ caused by a unit change in the variable y_1 at zero, with all other changes at $x = 0$ held equal to zero. This is precisely the definition of a column of elements in a transition matrix. Thus, this procedure would be carried out N times to obtain successive columns of the transition matrix.

It should be noted that, although the original system of differential equations is non-linear, the perturbed equations (11-67) are linear in the perturbation variables. If the system (11-66)-(11-67) is not too large, the entire transition matrix may be generated in one cycle of a typical numerical-integration program. The drawback of this scheme is that there are N governing equations and N perturbation equations, so that the computer must handle 2N simultaneous differential equations. As pointed out, however, the perturbation equations are essentially linear with variable coefficients, and the computer can integrate them with little additional effort relative to the non-linear system of total order N.

A particularly valuable dividend obtained by using the foregoing method occurs when a minimum-thickness bound is included in the problem. Imposing this additional constraint on a numerical scheme requires the addition of a single decision statement. This statement determines whether or not the computed value of the thickness is greater than or less than the specified minimum t_o. If t_o exceeds the computed thickness, the computational scheme sets $t = t_o$ and $\delta t = 0$. These values are then assumed in the succeeding steps. This method of constraining the thickness has been employed successfully in a variety of optimization problems.

The method described above has proved to be quite accurate, and analytical solutions, when available, can be reproduced with great precision. Figs.11.7 and 11.8 present the resulting optimal thickness distributions and weight savings for various values of minimum thickness t_o.

As a final numerical example, minimizing the weight of a cantilever-free sandwich beam, of constant core height but variable face-sheet thickness (Fig.11.9), is considered. Again the fundamental bending frequency is held constant. This case will also serve to illustrate an alternative scheme for formulating the optimization problem, based on a functional called the Hamiltonian [11.23]. Once again it is desired to minimize

$$\vartheta = \int_o^1 t \, dx \quad , \tag{11-69}$$

where $t = T(x)/T_o(x)$ and $x = X/L$, subject to the constraints

$$\left.\begin{aligned} w' &= p & &\left(w = \frac{Y(x)}{L}\right) \\ p' &= q/t & &(q = tw'') \\ q' &= r & & \\ r' &= (\alpha t + \beta)\, w & &\begin{cases} \alpha = \delta_1\, \omega^2 \\ \beta = \delta_2\, \omega^2 \end{cases} \end{aligned}\right\} \tag{11-70}$$

These constraint equations are adjoined to the function to be minimized, t, to form the Hamiltonian

$$H = t + \lambda_w p + \lambda_p q/t + \lambda_q r + \lambda_r [(\alpha t + \beta) w] \quad . \tag{11-71}$$

A necessary condition for a minimum of ϑ is that

$$\frac{\partial H}{\partial t} = 0 = 1 - \frac{\lambda_p q}{t^2} + \alpha \lambda_r w \quad . \tag{11-72}$$

Other necessary conditions for the constrained minimum read

$$\left.\begin{aligned} \lambda'_w &= -\frac{\partial H}{\partial w} = -\lambda_r(\alpha t + \beta) \\ \lambda'_p &= -\frac{\partial H}{\partial p} = -\lambda_w \\ \lambda'_q &= -\frac{\partial H}{\partial q} = -\frac{\lambda_p}{t} \\ \lambda'_r &= -\frac{\partial H}{\partial r} = -\lambda_q \quad . \end{aligned}\right\} \tag{11-73}$$

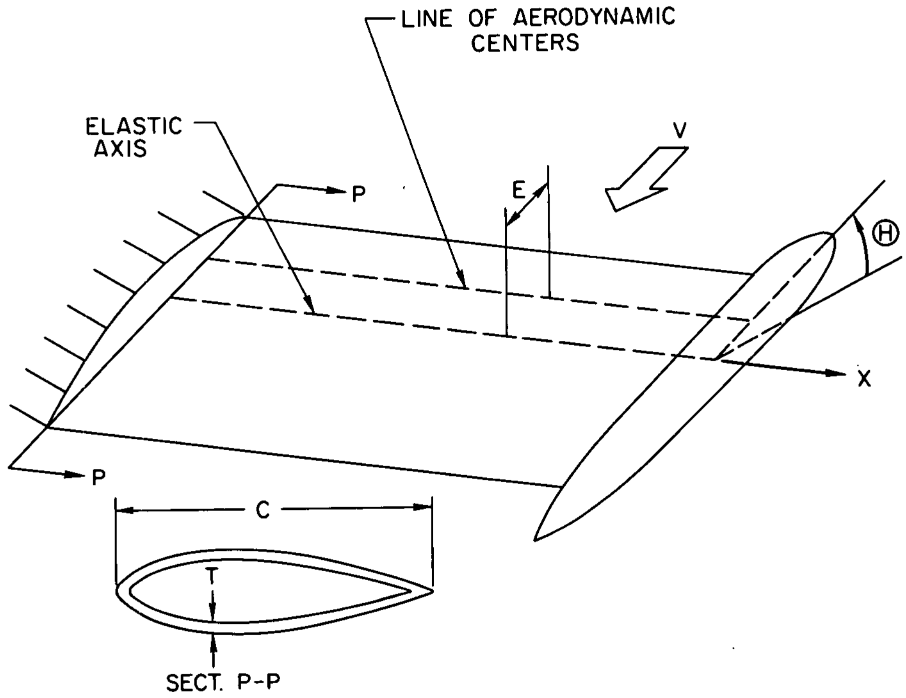


Fig. 11.6 Rectangular Cantilever Wing used for Torsional Divergence Calculations

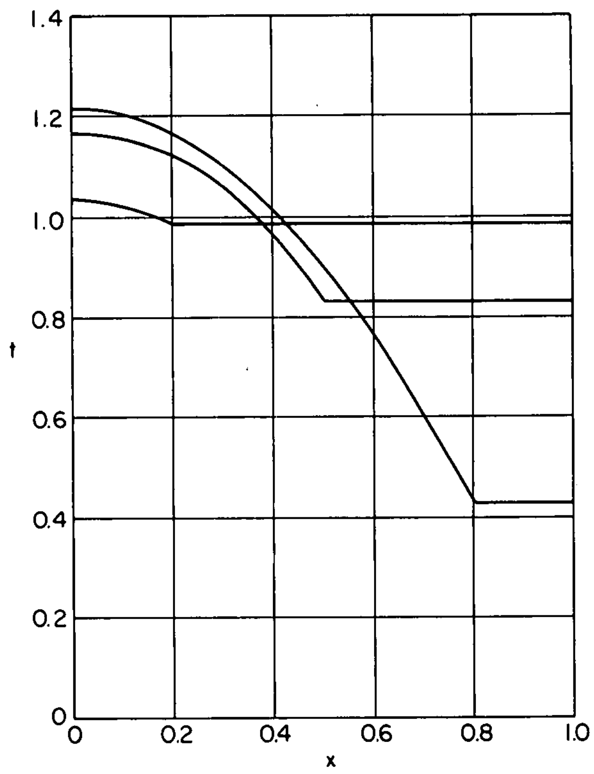


Fig. 11.7 Optimum Dimensionless Skin Thickness Distributions for Torsional Divergence of a Rectangular Cantilever Wing with Various Minimum Values of Skin Thickness

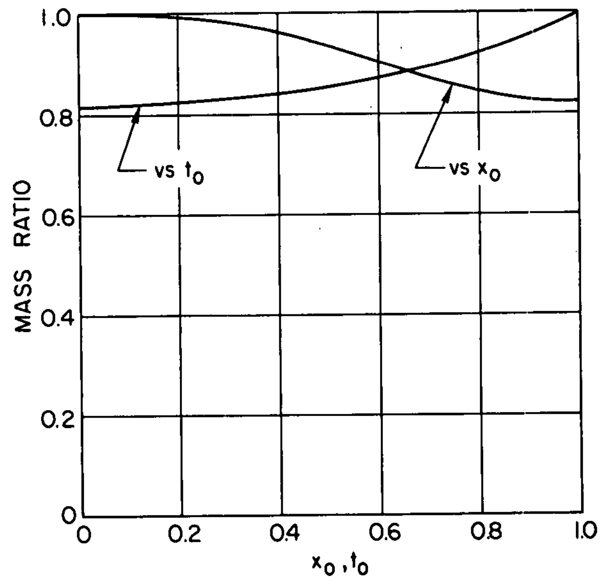


Fig. 11.8 Mass Ratio (Structural Mass Saving Relative to the Uniform Case) Plotted Versus x_0 and t_0 for the Family of Wings in Fig. 11.7

The system (11-70), (11-72) and (11-73) gives the differential equations which govern the problem.

The specified boundary conditions are

$$\left.\begin{aligned} w(0) &= p(0) = \lambda_q(0) = \lambda_r(0) = 0 \\ q(1) &= r(1) = \lambda_w(1) = \lambda_p(1) = 0 \quad . \end{aligned}\right\} \tag{11-74}$$

In addition to the above equations, one can impose a minimum thickness constraint, which may be expressed as

$$t_o - t \leqslant 0 \quad . \tag{11-75}$$

In this case, an augmented Hamiltonian would be used:

$$\left.\begin{aligned} & H^* = H + \mu\,(t_o - t) \\ \text{where} \quad & \begin{cases} \mu \geqslant 0 & \text{for} \quad t = t_o \\ \mu = 0 & \text{for} \quad t_o - t < 0 \quad . \end{cases} \end{aligned}\right\} \tag{11-76}$$

Note that the control equation (11-72) may then be expressed as

$$\mu = 1 - \frac{\lambda_p q}{t^2} + \alpha \lambda_r w \quad . \tag{11-77}$$

An analytical solution to this problem is not known. The transition-matrix algorithm, using determination of unit solutions to find the transition matrix, converged readily and produced thickness distributions like those shown in Fig.11.10, with the associated weight savings.

The limits of research in aeroelastic optimization, by direct integration of differential equations, may be said at the time of writing to be characterized by two problems - both currently under investigation but without numerical results ready for presentation. The first involves minimizing the mass, for fixed hypersonic flutter speed, of a thin homogeneous or sandwich plate in two dimensions. The second consists of optimizing for bending-torsion flutter a cantilever beam-rod wing, in which case the true complex nature of the aerodynamic forces is accounted for. Although both of these can be set up like other preceding examples, the airload expressions cause the functional F to be a complex function of the real argument x.

Special measures must be taken to ensure that the optimal thickness t remains a real quantity. For this purpose, Turner [11.13] has shown that it is sufficient to treat only the real part of F. Some of the details of setting up these problems can be seen in [11.22]. Following Turner, the manner of dealing with the complex behavior is reviewed, in Section 11.3, in connection with discrete-element systems.

11.3 Discretization by Assumed-Mode and Finite-Element Methods

It is self-evident that applications of aeroelastic optimization which are to have potential practical value in improved aircraft structural design must, in one way or another, involve the approximation of continuous systems by means of discrete elements. The design or control variables of Section 11.2 are then replaced with a finite vector of n adjustable element properties. Minimization of the chosen merit function amounts to a search of n-vector space rather than function space.

Schmit and Thornton's example [11.4] of the rectangular supersonic wing with minimum propulsive work, wherein $n = 2$, was described in Section 11.1 and cited as the only published instance to date of a flutter constraint applied in combination with more conventional constraints of structural optimization. Section 11.1 also observed that there is apparently no fundamental obstacle to placing bounds on aeroelastic properties during minimum-weight design of aeronautical structures. In current literature, however, the process of merging such constraints into the mainstream has not yet taken place. It therefore seems appropriate to review the present status of efforts in this direction, inasmuch as they are compatible with the hoped-for future progress*.

Whether used separately or jointly, there are two general ways of discretizing a structural-inertial system. The first consists of division into several compatible finite elements, for each of which the state of stress and deformation is specified by a set of scalars. When this method is employed in isolation, r independent quantities are chosen (e.g., the normal displacements at an array of 'panel

---

*For interesting examples of recent work on finite-element optimization of various structures with constraints on free vibration or dynamic-response amplitudes, the reader is referred to the second, third and fourth papers in the proceedings of the AIAA Structural Dynamics and Aeroelasticity Specialist Conference, April 1969. This volume is cited in connection with [11.13].

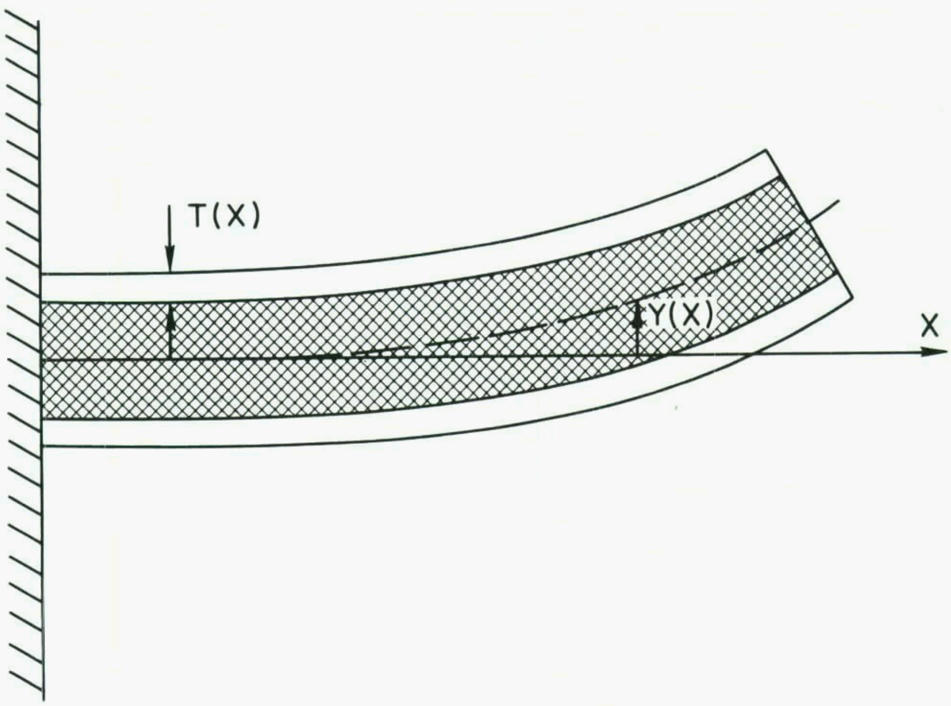


Fig.11.9 Cantilever Sandwich Beam or Panel with Constant Core Height but Variable Face-Sheet Thickness

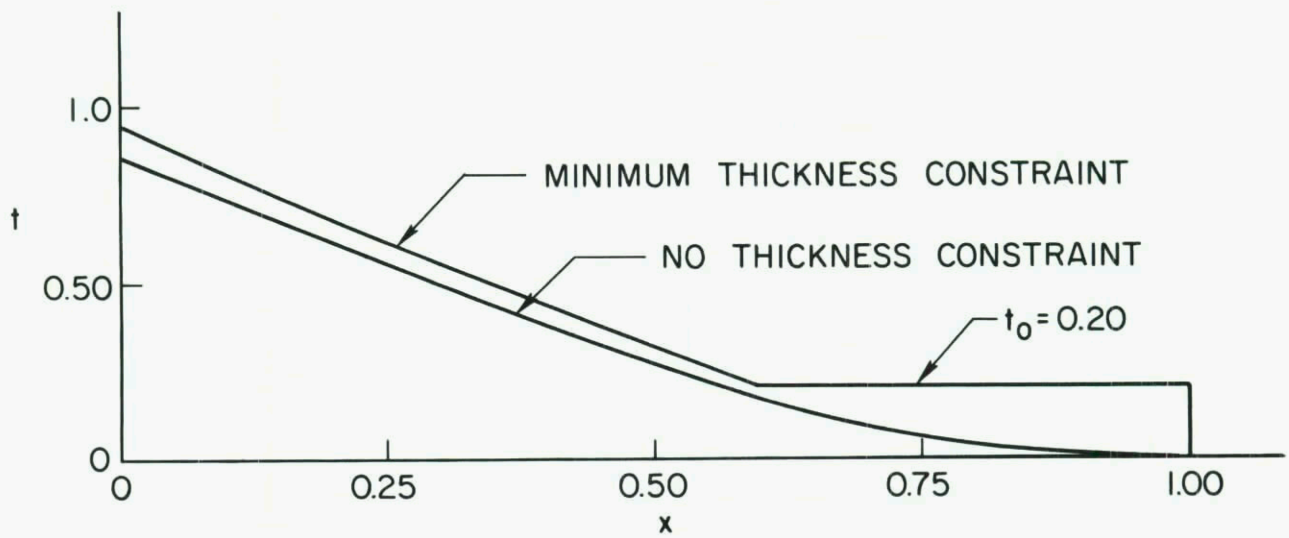


Fig.11.10 Face-Sheet Thickness Distributions for Optimum Cantilever Sandwich Beams, with Fixed Fundamental Bending Frequency and Two Different Thickness Constraints. Value of $\delta_1 = 0.5$. Overall Mass Ratios Relative to Uniform Case: 67% for $t_o = 0$ and 78% for $t_o = 0.20$

points' over the surface of a thin wing) which completely specify the state, and they become the degrees of freedom for construction of equations of motion.

Alternatively, all external forces can be removed from the system and the resulting homogeneous equations solved for up to r natural frequencies, and associated normal modes, of free vibration. This is one avenue leading into the second scheme for discretization, which is the superposition of finite numbers of normal or assumed modes of deformation. The (time-dependent or constant) modal amplitudes then serve as degrees of freedom. Since the numerous variants of this scheme are amply described in any advanced text (e.g., Chapters 3-4 of Bisplinghoff, Ashley and Halfman [11.25]) on structural dynamics, they require no elaboration here.

It should be mentioned, however, that one way of discretizing is to apply Galerkin's procedure to the sort of differential-equation systems discussed in Section 11.2. An attempt to optimize a rectangular wing for low-speed flexure-torsion flutter, wherein this procedure was applied to the spanwise distributions of skin thickness and bending and twisting amplitudes, is reported in [11.24]. Since it is clear that the approximation has not converged with the rather limited number of degrees of freedom assumed in [11.24], and since there is a question whether the constrained flutter condition actually constitutes the minimum critical speed of the 'optimal' designs, it would be premature to reproduce those results here. This is not to say that such an approach holds no promise for the future.

Turner's procedure for discrete systems [11.13] starts from the following form of the equations of motion, describing a state of neutrally stable oscillation (flutter, free vibration, etc.):

$$([K] - \omega^2[M] - \omega^2 \rho C^4 L[A])\{q\} = \{0\} \quad . \tag{11-78}$$

Here $[K]$, $[M]$ and $[A]$ are square matrices of stiffness elements, inertia elements, and dimensionless aerodynamic loads, respectively. The quantities in $[A]$ are normally complex numbers, representing generalized forces, aerodynamic coupling and the like; they depend on reduced frequency k, flight Mach number, and the dimensionless manner in which the motion is approximated (mode shapes, panel-point locations on a given wing planform, etc.). There are r dimensionless coordinates q_j in the column matrix. The system (11-78) has sufficient generality that virtually <u>any</u> discretized problem can be cast in this form.

For definiteness, let it be assumed that the flutter speed is held fixed at a single flight condition. Let the equations of motion first be set up for a reference structure, identified by a superscript (0), and let there be n adjustable mass elements m_j added to (or subtracted from, if negative) this structure. During this adjustment process, let the modal shapes on which Eq. (11-78) are based be held fixed. It will then be true, under rather broad conditions when the increments m_j are sufficiently small, that $[A]$ remains unchanged and the alterations to $[K]$ and $[M]$ are linear in the m_j. For instance, this would be the case if the m_j were associated with thickness modifications at a set of thin skin elements distributed over the area of a wing. It follows that, for any slightly altered structure,

$$\left.\begin{aligned} [K] &= [K^{(0)}] + \sum_{j=1}^{n} m_j[K^{(j)}] \\ [M] &= [M^{(0)}] + \sum_{j=1}^{n} m_j[M^{(j)}] \quad , \end{aligned}\right\} \tag{11-79}$$

the normalized correction matrices $[K^{(j)}]$ and $[M^{(j)}]$ being found by the same procedure that produced the original equations.

Turner introduces the shorthand definitions

$$-\omega^2 \rho C^4 L[A] \equiv [C] \quad , \text{ (known function of } k) \tag{11-80}$$

$$[K] - \omega^2[M] \equiv [B] = [B^{(0)}] + \sum_{j=1}^{n} m_j[B^{(j)}] \quad . \tag{11-81}$$

For an assigned value of airspeed $V = V_F$, but possibly allowing for variations in flutter frequency ω, one's objective is to minimize the structural mass, subject to the r algebraic constraint equations

$$([B] + [C])\{q\} = \{0\} \quad . \tag{11-82}$$

A normalizing condition, such as $q_r' = 1$, is introduced to make the solution of Eq. (11-82) unique for the prescribed eigenvalues. Turner also defines a similarly-normalized row matrix $\lfloor p \rfloor$ by means of

$$\lfloor p \rfloor([B] + [C]) = \lfloor 0 \rfloor \quad . \tag{11-83}$$

All the complex quantities p_i, q_j, as well as the complex elements of C, are separated into real and imaginary parts according to the generic notation

$$p_i = p_i' + i\, p_i'' \quad . \tag{11-84}$$

Since the m_j are obviously real, as are all other elements of [B], rationalization of (11-82) produces

$$\left.\begin{aligned} ([B] + [C'])\,\{q'\} - [C'']\,\{q''\} &= \{0\} \\ [C'']\,\{q'\} + ([B] + [C'])\,\{q''\} &= \{0\}. \end{aligned}\right\} \tag{11-85}$$

A set of r Lagrange multipliers $\lambda = \lambda' + i\,\lambda''$ is used to associate the constraints (11-85) with the merit function

$$\vartheta = \sum_{j=1}^{n} m_j \tag{11-86}$$

which is to be minimized. According to the algebraic theory of extremals, the desired result can be found by defining the Euler-Lagrange function

$$F = \sum_{j=1}^{n} m_j + \lfloor\lambda'\rfloor \left[([B] + [C'])\,\{q'\} - [C'']\,\{q''\}\right] + \lfloor\lambda''\rfloor \left[[C'']\,\{q'\} + ([B] + [C'])\,\{q''\}\right] . \tag{11-87}$$

F must be stationary for independent variations of all m_j and the non-normalized values of q_k', q_k''. The former condition yields the n 'control equations'

$$1 + \lfloor\lambda'\rfloor \left[\frac{\partial B}{\partial m_j}\right]\{q'\} + \lfloor\lambda''\rfloor\left[\frac{\partial B}{\partial m_j}\right]\{q''\} = 0 \ , \qquad j = 1,2,\ldots n \ , \tag{11-88}$$

where the forms of the derivative matrices are obvious from Eq. (11-81) and (11-79). The latter condition can easily be shown to be equivalent to the 2r real and imaginary parts of

$$\lfloor\lambda\rfloor\,([B] + [C^*]) = \lfloor 0 \rfloor \ , \tag{11-89}$$

where the asterisk denotes complex conjugate. It follows from Eq. (11-89) and the definition (11-83) of the row $\lfloor p \rfloor$ that λ and p^* are proportional, a relationship which Turner writes

$$\left.\begin{aligned} \lfloor\lambda'\rfloor &= \Lambda'\lfloor p'\rfloor - \Lambda''\lfloor p''\rfloor \\ \lfloor\lambda''\rfloor &= -\Lambda''\lfloor p'\rfloor - \Lambda'\lfloor p''\rfloor \ . \end{aligned}\right\} \tag{11-90}$$

Here Λ' and $(-\Lambda'')$ are the real and imaginary parts of a constant Λ, to be determined in the solution process. Eq. (11-90) permit the λ to be eliminated in favor of the p. The resulting final forms of the optimizing equations and constraints (11-88), (11-82) and (11-83) become

$$\mathrm{Re}\left(\Lambda\lfloor p\rfloor[B^{(j)}]\ \{q\}\right) = -1 \ , \quad j = 1,2,\ldots n \ , \tag{11-91}$$

$$\left([B^{(0)}] + \sum_{j=1}^{n} m_j[B^{(j)}] + [C]\right)\{q\} = \{0\} \tag{11-92}$$

$$\left([B^{(0)}] + \sum_{j=1}^{n} m_j[B^{(j)}] + [C]^T\right)\{p\} = \{0\} \ . \tag{11-93}$$

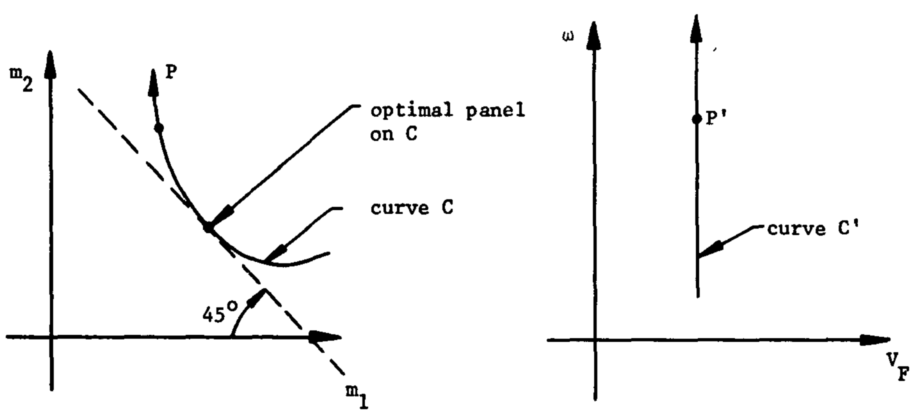
In principle, Eq. (11-91)-(11-93) constitute a system of 4r + n real, non-linear, algebraic equations in the optimal n masses, the (4r - 2) undetermined parts of {q} and {p}, and the two parts of Λ. Variations in frequency must be handled by trying a set of values of ω until an absolute minimum merit function is discovered. In Section IV of [11.13], Turner discusses the practical process of separating Eq. (11-91)-(11-93) into reals and imaginaries, some details of solution, and limitations on allowable values of r and n. The algebraic system is solved by an iterative scheme, described as a generalization of the Newton-Raphson method. The work of Freudenstein and Roth [11.26] is cited with respect to the importance of finding a starting approximation that is close enough to the desired solution to ensure convergence. In this connection, however, it should be mentioned that considerable research is currently in progress on improvement of algebraic optimization procedures; once an appropriate system of equations is established, the non-linearity may no longer constitute such an obstacle and many ways will soon be available for efficiently computing the desired solution.

In [11.13], Turner presents two examples. The first involves a three-segment finite-element approximation to a sandwich panel fluttering at Mach number 3 and standard sea level density. It is interesting that, for the system parameters selected, the panel of minimum face-sheet mass differs insignificantly from a uniform panel with the same critical speed. Because the latter was chosen as the reference condition for optimization, convergence turned out to be very rapid. Since no proofs of uniqueness are available, however, an intriguing question concerns whether other, more substantially improved designs might be discovered which are remote from the reference case. The author remarks, "It is not known whether these findings would be altered significantly if the panel were divided into a greater number of segments or if the effects of independent variation of core density and associated effects of shear deformation were included in the analysis."

Turner's second example deals with the cantilever wing, some of whose dimensions are illustrated in Fig.11.11. Each of the three segments shown was assumed to have constant properties. Bending displacement in each segment was represented by a cubic polynomial and twist by a linear polynomial. Thus the total of deflection, bending-slope and twist coordinates at Sections 1, 2 and 3 add to $r = 9$. Nonstructural mass was taken to be a uniform 0.0181 $\text{lb-sec}^2/\text{in}^2$ across the span; other details will be found in the paper.

The reference case was established by finding the combination of m_1, m_2 and m_3 which added to a minimum while holding the fundamental frequency of torsional vibration <u>in vacuo</u> at 15 Hz. A flutter analysis, based on incompressible aerodynamic strip theory, then gave a <u>sea level</u> $V_F = 675$ knots at a frequency of 8.99 Hz. Fig.11.12 shows the properties of this initial approximation. Also plotted, vs. flutter frequency, is the succession of optimal states arrived at by the foregoing method and leading to the indicated minimum-mass system. The reference state was expected to be close to the desired solution (cf. [11.5], [11.6]), so it is not surprising to obtain a final result only 2% lighter. Turner estimates, however, that the optimized structure is 18% lighter than a <u>uniform</u> wing having the same flutter performance.

In connection both with the cantilever wing example and with the foregoing quotation from Turner regarding his three-segment panel, it is possible to speculate about the effects of the number n of design variables. In view of the proof in [11.13] that the optimal panel has a thickness distribution symmetrical about its midchord, there are really only two independent variables: m_1 and m_2. The relationship between these variables and the flutter eigenvalues V_F and ω may be likened to a transformation from an $m_1 - m_2$ - plane to a $V_F - \omega$ - plane, as in the sketch below:



If one assumes that the transformation between points like P and P' (or curves like C and C') is unique and one-to-one, it is clear that with $m_1 = m_3$ there is only a single design corresponding to a given flutter speed and frequency. This explains, for instance, why the uniform panel is also the optimal panel for $\omega = 54.75$ Hz on Fig.2 of [11.13]. All other points on this figure would be encompassed by a pair of curves similar to C and C'. Should <u>three</u> independent variables be available, as with the cantilever, then a curve in three-space corresponds to a point like P'; there would seem to be considerably greater freedom available to the search for the best system. With the panel, this could be achieved by going to five or more segments. In general, an adequately large n would always seem to be necessary for the finite-element design to 'converge'.

Another investigation will next be discussed which aims, by means of somewhat less sophisticated mathematical analysis, at the ultimate flutter optimization of very complicated airframe structures with many hundreds of finite-element degrees of freedom and dozens of design variables. Based on a representation similar to Turner's, Eq. (11-78), this work was conducted at The Boeing Company's Commercial Airplane Group.

The procedure starts by generating a full stiffness matrix [K] and a corresponding lumped mass matrix [M], with which a set of normal mode shapes are calculated. These modes are then used as generalized distributed coordinates to formulate the flutter equations, also in the form of Eq. (11-78). In determining the sensitivity of flutter speeds to redistributions of structural stiffness, and the corresponding total structural mass, the most direct procedure is to recalculate new stiffness and lumped

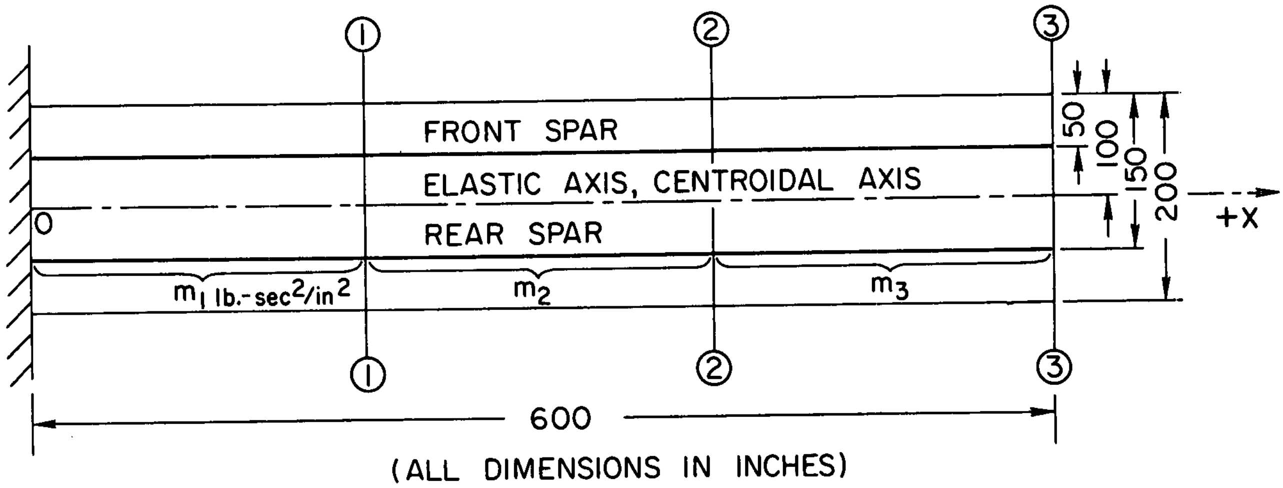


Fig.11.11 Rectangular Cantilever Wing used for Bending-Torsion Flutter Optimization by Turner (Ref. [11.13])

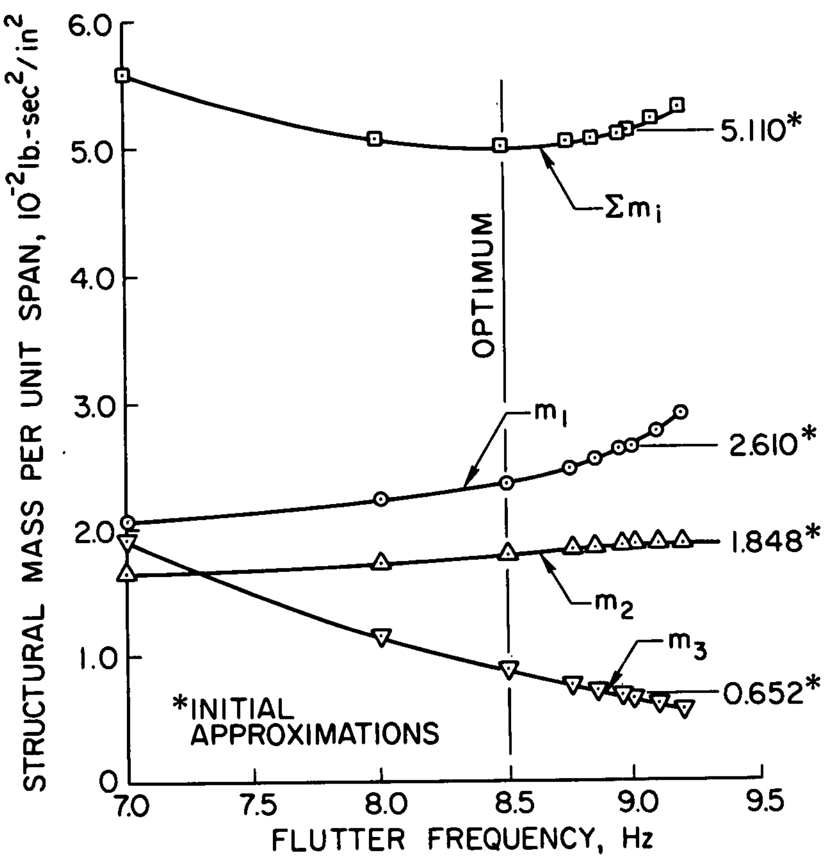


Fig.11.12 Evolution of Values of Mass per Unit Span during Optimization of Wing in Fig.11.11

mass matrices at each step. In the example to be described below, the need for reformulating the stiffness matrix from the beginning was avoided by dividing the configuration into substructures. Rearrangement of the stiffness distribution was achieved by simply rescaling the stiffness matrix for one or more of the substructures and then forming the required full stiffness matrix by a straightforward matrix merge procedure. New stiffness matrices could be obtained in less than a minute of computer time on the CDC 6600, or well less than 1/30 of what would have been required if the actual finite elements had been rescaled and the new stiffness matrix generated in one pass. The overall procedure is outlined in the following steps:

(1) Generate basic substructure structural mass and stiffness data.

(2) Scale substructure structural mass and stiffness matrices. (Enter here with new scale factors.)

(3) Merge scaled stiffness matrices, reduce out those degrees of freedom needed for merging but not required for vibration solution. Merge structural mass matrices, combine with fixed mass matrix.

(4) Solve vibration problem for normal mode shapes and frequencies of modified structure.

(5) Interpolate modal data from structural control points to aerodynamic control points. Generate chordwise slope data for aerodynamic program. (Enter here with new speed regime, new Mach number. Advanced three-dimensional oscillatory lifting-surface theory is used; see, e.g., Vol.II of [11.1].)

(6) Calculate generalized aerodynamic forces, generalized mass and stiffness matrices for the modes of the modified structure. (Enter here with new Mach number, if subsonic or piston theory being employed.)

(7) Set-up and solve the flutter problem using the so-called 'V-g method' (see Section 9.5. of [11.25]).

For purposes of a sample problem, a low-aspect-ratio configuration of supersonic-transport type was divided into 13 substructures. These regions, shown in Fig.11.13, were chosen to isolate logical structural regions such as the wheel well (panel 4) and the main wing spar (panels 5, 9, 11 and 12). This should be considered a fairly gross representation. Once general sensitivities have been established, however, it would be simple to go back and sub-divide panels of particular interest for further study. Also there is no reason why specific structural members could not be broken out and considered as separate regions by themselves.

The results presented here are for two Mach numbers (one subsonic and one supersonic) and two aircraft weight conditions (one light-weight and one heavy, corresponding nearly to maximum gross weight). The aerodynamic generalized forces were calculated using either supersonic Mach-box or subsonic kernel-function theory as appropriate*. The critical flutter modes included two low-frequency modes and one high-frequency mode, any one of which could prove the most critical at a given flight condition.

Table 11.1 shows the dependence of flutter speed on changes in panel stiffness and the corresponding (proportional) structural weight. 'Sensitivity' R relates to equivalent airspeeds and is defined below the table. Since weight is of primary interest, these sensitivities are given as the ratio of the change in flutter speed for 1000 lb change in structural weight to the flutter speed of the reference condition. Here the structural weight is assumed to vary in direct proportion to the stiffness. This may be justified by assuming that the increase in stiffness is achieved by increasing skin and spar thicknesses and spar cap widths†. Table 11.1was obtained by increasing the stiffness and corresponding structural weight of each region in succession while holding the remaining panels at their reference level. Thus, these numbers are first-order forward-difference approximations to the derivatives of flutter speed with respect to structural weight.

Changes in stiffness ranged from 10% to 20% of each panel. Changes in flutter speed were small, and care had to be taken in interpreting the V-g solutions to be sure of identifying the most critical condition. As the calculations progressed, larger changes were used. Some idea of the actual linearity of these derivatives with size of stiffness changes may be gained from Table 11.2. Generally, the derivatives were fairly linear for ±20% modifications in panel stiffness, and flutter speeds for distributions obtained by rescaling several panels within these limits could be adequately predicted.

The data displayed in Table 11.1 have been used to generate two sets of redistributions of structural stiffness. The first, based on column 3 of Table 11.1, was designed to raise the flutter speed for that condition with no net increase in structural weight. Here the amounts of weight added to or subtracted from the panels were made roughly proportional to the values of their derivatives**. This redistribution is shown in Fig.11.13and labeled A-1x in Table 11.3. It results in the rearrangement of some 3000 lb per side of structural weight, but no net weight change. The consequences of doubling this redistribution are shown as A-2x and of tripling it as A-3x. The numbers in Table 11.3 are given as flutter speed divided by the flutter speed for the reference structure for each flight condition.

---

*The reader is again referred to Vol.II of [11.1], and citations made therein, for information on these three-dimensional aerodynamic theories.

†Informal talks with weights engineers indicate that 0.75 for a value of the ratio of change in structural weight to change in stiffness would be more realistic.

**This procedure is obviously equivalent to one step in a gradient or steepest-ascent method [11.23] of optimal search.

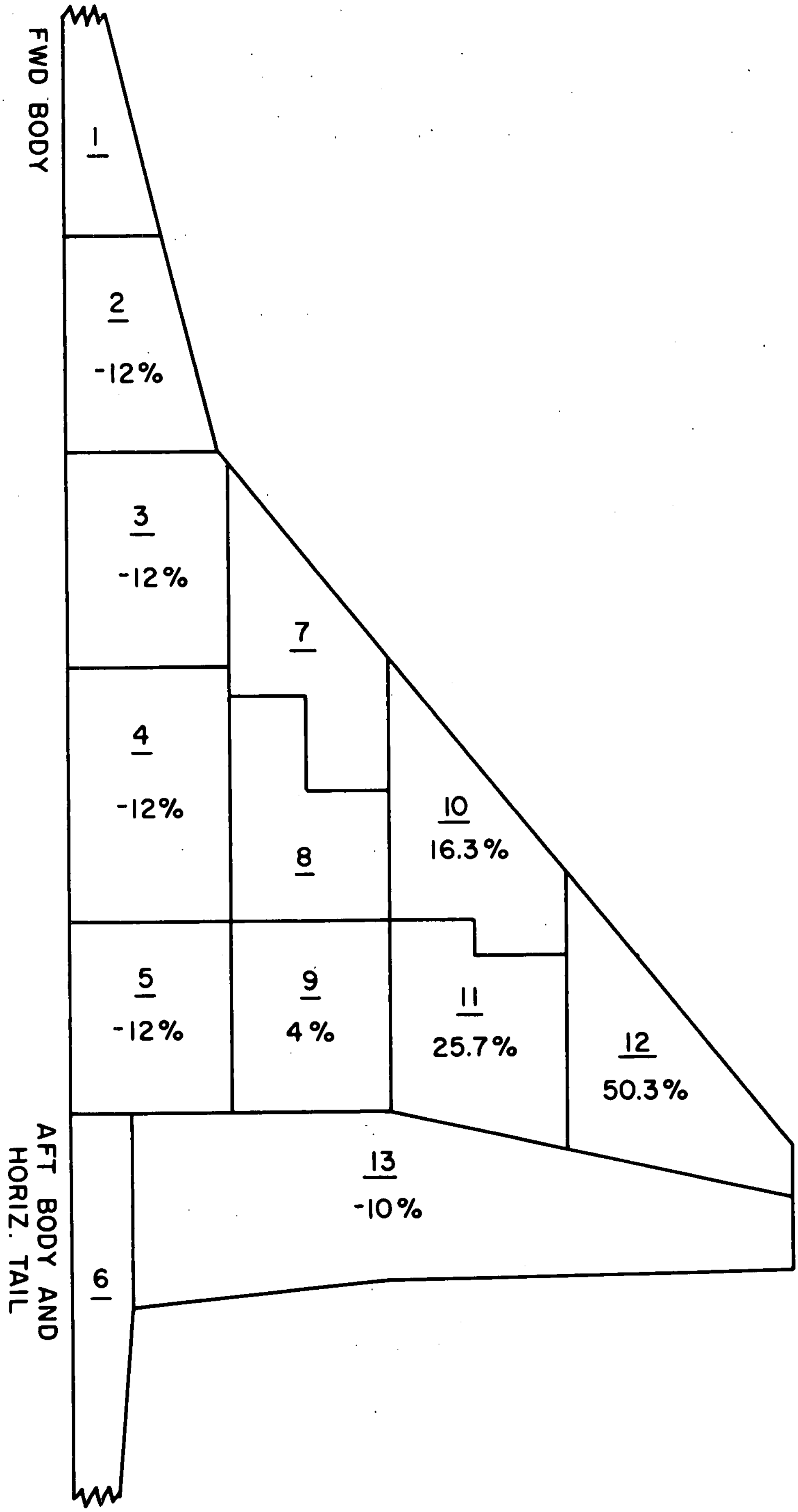


Fig.11.13 Plan View of Aircraft of Supersonic Transport Type, showing the Arrangement of Substructures used for Mass and Stiffness Redistribution. Also shown are the Mass Changes in per cent Associated with Redistribution A-1x in Table 11.3

For three of the four conditions this distribution is very beneficial. For the light-weight, subsonic condition, however, the flutter speed decreases. The reason for this may be readily seen in Table 11.1, where for the heavy-weight subsonic condition the flutter performance is improved most by stiffening the wingtip and softening (probably lightening) the trailing edge. The exact opposite is true for the light-weight subsonic condition.

Table 11.1

Linear Derivatives of Change in Flutter with Respect to Change in Panel Stiffness

	Supersonic light-weight	Subsonic light-weight	Subsonic heavy-weight
Panel	R	R	R
1	-0.030	-0.007	0
2	-0.027	-0.011	-0.001
3	0	0	0
4	0	0.003	0.002
5	0.037	0.022	0.003
6	-0.008	-0.040	0.003
7	0.033	0.011	0.003
8	0.024	0.022	0.009
9	0.052	0.044	0.015
10	-0.015	0.014	0.013
11	0.022	0.012	0.022
12	0.016	-0.027	0.028
13	-0.033	0.013	-0.008

$$R = \frac{\text{change in } V_F}{1000 \text{ lb}} \times \frac{1}{(V_F)_{REF.}}$$

Table 11.2

Typical Non-Linearities in Derivatives of Change in Flutter Speed with Respect to Change in Panel Stiffness

	Panel	% change in stiffness of panel	$\Delta V_F/(V_F)_{REF.}$
Subsonic light-weight	12	+50 +100	-0.032 -0.038
	13	+10 -10 -20	0.007 -0.008 -0.018
Supersonic light-weight	5	+10 +30 +50	0.024 0.075 0.191
	9	10 30 50	0.029 0.100 0.143

If the purpose of the distribution were solely to improve the one condition (subsonic heavy-weight), when the improvement in flutter speed has fallen off significantly (say, after the rearrangement of some 6000 lb), a new set of derivatives should be calculated and used to form a new redistribution. In more realistic circumstances where the purpose is to clear all flight conditions, then the sets of derivatives for all the critical conditions should be used in formulating the redistributions.

As a second example, a redistribution was designed to improve stability in the low frequency mode for the light-weight supersonic condition using the derivatives of column 1, Table 11.1, and Table 11.2. Here, the ground rule was to determine a distribution that would require a minimum additional amount of weight. No structural weight was to be removed. This distribution, identified as B-1x in Table 11.3, consisted of increasing the stiffness in panel 5 by 18% and panel 9 by 25%, resulting in the addition of 2600 lb structural weight per side. The gain in flutter speed is more than would have been predicted from the derivatives. Also, in doubling this additional stiffness distribution, this particular critical mode vanished and was replaced by a high frequency mode.

Table 11.3

Dimensionless Effects on Flutter of Modifying Structural Mass of an Aircraft of Supersonic Transport Type. Tabulated Quantities represent Flutter Speeds divided by the Reference Flutter Velocity at the Corresponding Flight Altitude

		Supersonic		Subsonic	
		Light-weight	Heavy-weight	Light-weight	Heavy-weight
no net weight change	BASE	1.0	1.0	1.0	1.0
	A-1x 3000 lb rearranged (See Fig.11.13)	1.04	1.08	0.98	1.11
	A-2x 6000 lb rearranged	1.07	1.17	0.98	1.17
	A-3x 9000 lb rearranged	1.11	1.23	0.99	1.19

(a)

		Supersonic, light-weight	
		Low frequency	High frequency
structural weight added	BASE	1.0	1.17
	B-1x 2600 lb added	1.17	1.17
	B-2x 5200 lb added	—— *	1.17
	B-3x 7800 lb added	—— *	1.17

(b)

Distribution A, No net increase in weight - each increment represents rearrangement of 3000 lb of structural weight per side.

Distribution B, Addition of 2600 lb of structural weight per side per increment.

* Low frequency instability no longer exists.

The foregoing obviously constitutes a very preliminary attempt at coping with many degrees of freedom and several flutter constraints on a given design. Following Taylor [11.15], however, one can hypothesize a relationship between a system of given mass and maximum flutter speed and one of minimum mass for given flutter speed. As with other methods aimed at the same objective, a key step in the computations is to estimate sensitivities - derivatives of flutter eigenvalues with respect to changes in the physical system. Many years ago, van de Vooren([11.27], Section 9) investigated the analytical determination of such derivatives from the properties of basic flutter equations like Eq. (11-78). His work is expected to have special significance in aeroelastic optimization, because of its potentialities for simplifying the calculations. Ref. [11.27] also presents formulaé for second- and third-order effects of system changes; although more involved, these hold out the possibility of accounting for the sort of non-linearity exposed in Table 11.2.

The essence of van de Vooren's approach can be explained starting from Eq. (11-78). It is rewritten, in a form closer to the notation of [11.27], as follows:

$$([K] - \mu[U])\{q\} = \{0\} \quad . \tag{11-94}$$

Here $[U]$ is a combined inertia-aerodynamic matrix - an array of complex numbers when a particular structure, reduced frequency, flight Mach number and altitude have been chosen. μ represents a complex eigenvalue, which can be equated to the combination $\omega^2/(\omega^2_{REF.}[1 + ig])$ when the V-g method is being used. The r eigenvalues, and associated complex eigenvectors, are computed for Eq. (11-94). The same is done for the transposed equation ($[K]$ is symmetrical):

$$([K] - \mu[U]^T)\{p\} = \{0\} \quad . \tag{11-95}$$

The r values of μ can be proved equal for these two equations.

Square matrices [Q] and [P] are constructed from the modal columns of Eq. (11-94) and (11-95), with the columns ordered on increasing frequency. By the bi-orthogonality relation of matrix algebra, one can show that the following matrices are diagonal:

$$\left.\begin{aligned} [\bar{K}] &= [P]^T [K][Q] \\ [\bar{U}] &= [P]^T [U][Q] \quad . \end{aligned}\right\} \tag{11-96}$$

Now suppose that incremental matrices $[\varepsilon]$ and $[u]$ are added to $[K]$ and $[U]$, respectively, by a small system alteration of the sort envisioned in this section. The following nondiagonal matrices are then calculated:

$$\left.\begin{aligned} [\bar{\varepsilon}] &= [R]^T [\varepsilon][Q] \\ [\bar{u}] &= [R]^T [u][Q] \quad . \end{aligned}\right\} \tag{11-97}$$

It is demonstrated quite straightforwardly in [11.27] that the resulting change in the ith eigenvalue μ_i of Eq. (11-94) is linearly estimated by

$$\Delta\mu_i = \frac{\bar{\varepsilon}_{ii} - \mu_i \bar{u}_{ii}}{\bar{U}_{ii}} \quad . \tag{11-98}$$

Here the double-i subscripts designate the ith elements of the principal diagonals of the corresponding matrices.

Each desired sensitivity is determined from Eq. (11-98), when $[\varepsilon]$ and/or $[u]$ are linearly related to the unit change in mass/stiffness, by factoring this change out of Eq. (11-97) and (11-98) and dividing Eq. (11-98) by it. Usually only the eigenvalue μ which is associated with the critical flutter condition will require this treatment. A careful examination of the computational steps leading to Eq. (11-98), compared with complete flutter analyses of the reference and n modified systems, indicates a considerable saving of labor, which should increase with an increasing number of design variables. The important facts are that the aerodynamic matrix need be constructed only once and that the complex eigenvalue solution (equivalent to a complete V-g determination) need be conducted only twice - for Eq. (11-94) and (11-95).

11.4 Concluding Discussion

The most obvious comment to be made about the subject of aeroelastic optimization, as comprehended in the foregoing sections, is that it is both presently incomplete and in a rapid state of evolution. It is clear that both continued research and practical applications will be necessary along the two major lines of development: discrete-element approximation of realistic light-weight structure; and the differential-equation idealization of simplified systems, which leads to a search of function space for optimal solutions by methods analogous to those used in modern control theory.

The latter approach is important as a general guide to the potentialities of this branch of optimization, as a reference source for checking more approximate results, and as a possible avenue to the proof of theorems illuminating certain questions that arise from the non-linear mathematics. Assuming that correct and meaningful problem statements can be achieved, one faces the single overriding difficulty of finding solutions by numerical integration of rather high-order differential-equation systems. Experience to date points to the transition-matrix scheme, together with the method of unit solutions for evaluating the required sensitivities, as most promising for this purpose. Very high-fidelity computer routines for matrix inversion are an essential adjunct. There are other alternatives for numerical solution discussed in books like Bryson and Ho [11.23], however; several among them, such as the method of backward sweeping, deserve further investigation.

In the area of discretization, the principal goal is to advance to very large numbers of degrees of freedom and design variables through the use of well-developed finite elements. There are many ways in which work done to date can be refined. Although it is not anticipated that extreme troubles will be encountered in the associated algebraic calculations, there remains a question whether better solutions far removed from the assumed initial design can be anticipated and/or realized. It is finally worth repeating that the imposition of aeroelastic inequality constraints should fit fairly routinely into several existing schemes for structural optimization under more conventional conditions on strength and stiffness.

One step in the direction of realism must be vigorously pursued in connection with both lines of development. This is the simultaneous imposition of multiple constraints of different types.

Although there are few general guideposts to determining when a problem is well-posed, it seems safe to conclude that any mathematically proper solution which also seems physically reasonable is an acceptable product of an optimal search. Two key and related questions stand unanswered, however, except in the most elementary cases. These concern when a result is unique and when it constitutes the absolute optimum, if such exists, among all possible extrema obtainable from a given algebraic or analytic statement. In the authors' view, these matters represent vital unfinished business for the applied mathematicians.

The benefits inherent in future exploration of this field are certainly not less than those foreseen from conventional minimum-weight structural design. One important basis for this opinion is the rather limited value of experience and intuition in the face of complicated configurations many of whose members must be sized by aeroelastic considerations. It has been said that modern structural optimization can produce results roughly comparable to the creations of an experienced designer confronted with the same requirements. There are few, if any, such designers in aeroelasticity.

With respect to what savings may be hoped for, it is first necessary to ask what one will choose as a figure of merit or reference solution. If the latter is (in some sense) a uniform structure with the same aeroelastic behavior, then fairly realistic - if simplified - examples have already been found where reductions in structurally-effective mass are possible in the order of 15-30%. (Those cases which save 70% to more than 90% are regarded either as suspect or physically unrealizable.) The small improvements so far achieved by finite-element methods are due to the near-optimality of the corresponding reference designs. There are many respects wherein the crude attempts of the past can be substantially improved, and this new tool is believed to hold substantial promise for the refinement of future aerospace structures.

Acknowledgements - This Chapter includes results obtained under a research program sponsored by the Air Force Office of Scientific Research and the National Aeronautics and Space Administration. Valuable assistance was provided by Messrs. M. J. Turner, Terrence A. Weisshaar, Jean-Louis Armand and William A. Vitte.

List of References

Ref.

11.1 Many Authors, *Manual of Aeroelasticity*, Vols.I through VI, issued over a period of years by NATO Advisory Group for Aeronautical Research and Development (revised versions of several volumes now in preparation under the editorship of R. Mazet)

11.2 R. L. Bisplinghoff and H. Ashley, *Principles of Aeroelasticity*, John Wiley and Sons, New York, 1962

11.3 W. M. Morrow II and L. A. Schmit, Jr., *Structural Synthesis of a Stiffened Cylinder*, NASA CR-1217, December 1968

11.4 L. A. Schmit and W. A. Thornton, *Synthesis of an Airfoil at Supersonic Mach Number*, NASA CR-144, January 1965

11.5 E. P. MacDonough, 'The Minimum Weight Design of Wings for Flutter Conditions,' *Journal of the Aeronautical Sciences*, Vol.20, No.8, August 1953, pp.573-574

11.6 A. L. Head, 'A Philosophy of Design for Flutter,' *Proceedings of AIAA National Specialists Meeting on Dynamics and Aeroelasticity*, Ft. Worth, Texas, November 1958, pp.59-65

11.7 H. Ashley and S. C. McIntosh, Jr., 'Application of Aeroelastic Constraints in Structural Optimization,' *Proceedings of the 12th International Congress of Applied Mechanics*, Springer, Berlin, 1969

11.8 W. H. Roberts, 'A Stiffness Criterion for Flutter and the Optimum Spanwise Skin Distribution for Aeroelastic Problems,' Rept. No.NA-54-619, 1954, North American Aviation, Inc., Los Angeles, California

11.9 R. E. Lunn et alii, 'A Method for Calculating the Optimum Bending Stiffness Required to Provide a Specified Lift-Curve Slope for a Flexible, Sweptback Wing,' Rept. No.NA-61-521, 1961, North American Aviation, Inc., Los Angeles, California

11.10 C. H. Hodson, 'Stiffness Requirements to Prevent Flutter of Moderate to High Aspect Ratio Surfaces,' Rept. No.NA-64-745, September 1964, North American Aviation, Inc., Los Angeles, California

11.11 C. H. Hodson, 'Estimation of Flutter Stiffness Requirements for Thin Low-Aspect-Ratio Wings,' Rept. No.NA-65-794, September 1965, North American Aviation, Inc., Los Angeles, California

11.12 M. J. Turner, 'Design of Minimum Mass Structures with Specified Natural Frequencies,' *AIAA Journal*, Vol.5, No.3, March 1967, pp.406-412

11.13 M. J. Turner, 'Optimization of Structures to Satisfy Flutter Requirements,' *Volume of Technical Papers on Structural Dynamics*, AIAA Structural Dynamics and Aeroelasticity Specialist Conference and ASME/AIAA 10th Structures, Structural Dynamics, and Materials Conference, American Institute of Aeronautics and Astronautics, New Orleans, 1969, pp.1-8

11.14 F. I. Niordson, 'On the Optimum Design of a Vibrating Beam,' *Quarterly of Applied Mathematics*, Vol.23, No.1, April 1965, pp.47-53

List of References (Contd.)

Ref.

11.15 J. E. Taylor, 'Minimum Mass Bar for Axial Vibration at Specified Natural Frequency,' *AIAA Journal*, Vol.5, No.10, October 1967, pp.1911-1913

11.16 W. Prager and J. E. Taylor, 'Problems of Optimal Structural Design,' *Journal of Applied Mechanics*, Vol.35, No.1, March 1968, pp.102-106

11.17 J. E. Taylor and E. F. Masur, 'A Global Condition for Optimal Structural Design,' Paper presented at 12th International Congress of Applied Mechanics, Stanford University, August 1968, (not yet published)

11.18 R. L. Halfman, *Dynamics*, Vol.II - *Systems, Variational Methods and Relativity*, Addison-Wesley, Reading, Mass., 1962, Chaps.10 and 11

11.19 R. H. Scanlan and R. Rosenbaum, *Aircraft Vibration and Flutter*, The Macmillan Company, New York, 1951

11.20 B. Smilg, 'The Instability of Pitching Oscillations of an Airfoil in Subsonic Incompressible Potential Flow,' *Journal of the Aeronautical Sciences*, Vol.16, No.11, November 1949, pp.691-696

11.21 J. E. Taylor, 'Optimum Design of a Vibrating Bar with Specified Minimum Cross Section,' *AIAA Journal*, Vol.6, No.7, July 1968, pp.1379-1381

11.22 S. C. McIntosh and F. E. Eastep, 'Design of Minimum-Mass Structures with Specified Stiffness Properties,' *AIAA Journal*, Vol.6, No.5, May 1968, pp.962-964

11.23 A. E. Bryson Jr. and Y.-C. Ho, *Applied Optimal Control*, Blaisdell, Waltham, Mass., 1969

11.24 S. C. McIntosh, Jr., T. A. Weisshaar, and H. Ashley, *Progress in Aeroelastic Optimization - Analytical Versus Numerical Approaches*, paper presented at AIAA Structural Dynamics and Aeroelasticity Specialist Conference, New Orleans, April 1969 (issued as SUDAAR No.383, Stanford University Department of Aeronautics and Astronautics)

11.25 R. L. Bisplinghoff, H. Ashley, and R. L. Halfman, *Aeroelasticity*, Addison-Wesley Publishing Company, Reading, Mass., 1955

11.26 F. Freudenstein and B. Roth, 'Numerical Solution of Systems of Nonlinear Equations,' *Journal of the Association for Computing Machinery*, Vol.10, 1963, pp.550-556

11.27 A. E. van de Vooren, *Theory and Practice of Flutter Calculations for Systems with Many Degrees of Freedom*, Doctoral Dissertation, Technical Institute of Delft, Holland, published in 1952 by Eduard Ijdo N. V., Leyden

11.28 M. J. Turner, *Proportioning Members of a Structure for Maximum Stiffness with Given Weight*, unnumbered report of Vought-Sikorsky Aircraft, January 1942 (adapted from work of S. J. Loring)

Appendix 11A - List of Principal Symbols

$a(x)$	function used for applying minimum-thickness constraint
A	cross-sectional area of bar
$[A]$	dimensionless aerodynamic matrix
$[B]$	$(= [K] - \omega^2[M])$ combined stiffness-inertia matrix
C	chord of wing
C_{L_α}	averaged or two-dimensional lift-curve slope of wing or airfoil
$[C]$	$(= -\omega^2 \rho C^4 L[A])$ abbreviated aerodynamic matrix
$e(\vec{q})$	strain energy per unit volume
E	Young's modulus; distance between wing aerodynamic center and elastic axis
f_i	function appearing in ith constraint differential equation
F	Euler-Lagrange function
g	structural damping parameter
$GJ(x)$	torsional rigidity of rod
H	Hamiltonian function
i	$(\equiv \sqrt{-1})$ the imaginary unit
I_o	mass moment of inertia per unit span (constant)
$I_\theta(x)$	optimum distribution of mass moment of inertia
ϑ	merit function (usually structural mass)
k	$(\equiv \omega C/2V)$ reduced frequency
K	a constant
$[K]$	matrix of stiffness elements
L	length of bar or beam; wingspan
m	dimensionless mass or thickness of structure (figure of merit)
m_j	added element of mass
$\bar{M}_X$	aerodynamic pitching moment about elastic axis (positive nose-up)
$[M]$	matrix of inertia elements
M	mass per unit length of bar
$\bar{M}_1$	tip mass attached to bar
n	order of vector of control or design variables
N	order of state vector of a system (total order of governing differential equations)
p,q,r,s,y	various auxiliary functions of x used in reducing a system to state-vector form
$\lfloor p \rfloor$	row matrix adjoint to $\{q\}$
$[P]$	square matrix of adjoint eigenvectors
q_i	discrete system coordinate or functional element of state vector
$\{q\}$	matrix of coordinates q_i
$[Q]$	matrix of q_i-eigenvectors
r	number of degrees of freedom in a discretized system
R	sensitivity of flutter speed to mass change
t	dimensionless thickness of structural material
t_o	specified minimum value of t
T	dimensional thickness
$[u]$	increment to mass-aerodynamic matrix
U	volume of elastic body
$U(X)$	displacement of points along a bar in extensional motion
$[U]$	mass/aerodynamic matrix in theory of [11.27]
V	airspeed or flight speed
V_D, V_F	critical values of V associated with divergence, flutter, respectively
w	$(\equiv Y/L)$ dimensionless bending deflection
x	dimensionless length coordinate in beam, bar, etc.

List of Principal Symbols (Contd.)

x_o	value of x at transition from varying thickness to minimum value t_o
X	dimensional length coordinate
y_i	element of state vector
Y(X)	bending deflection of beam or plate
α,β	quantities proportional to δ_1 and δ_2 [see Eq. (2-60)]
β	($\equiv \omega L \sqrt{\rho/E}$) dimensionless frequency parameter for vibrating bar
γ	used as a constant in Section 11.1; see also definition in (2-47)
δ	variation
δ_1,δ_2	fractions of mass, in reference uniform system, which are for primary structure and nonstructural, respectively
δ_1',δ_2',η	quantities related to δ_1 and δ_2 [see Eq. (2-37)]
$[\epsilon]$	increment to stiffness matrix
θ	elastic twist in rod or wing
λ	(with various subscripts) Lagrange multiplier or adjoint variable
μ	complex flutter eigenvalue
ρ	density of air or solid material
τ	time coordinate
ω	circular frequency of simple harmonic motion
ω_θ	frequency of torsional vibration or flutter
Ω	general dimensionless eigenvalue parameter

Subscripts, superscripts, etc.

$(\ )_o$	identifies reference system generally (uniform properties)
$(\ )^{(0)}$	superscript denoting reference system in theory of [11.13]
$(\ )^T$	transpose of matrix
$(\ )'$	derivative with respect to length coordinate
$(\ )', (\ )''$	real and imaginary parts of complex number
$(\ )^*$	complex conjugate
$\overline{(\ )}$	complex number or complex amplitude of simple harmonic quantity
$\vec{(\ )}$	vector

Chapter 12

OPTIMIZATION TECHNIQUES IN AIRCRAFT CONFIGURATION DESIGN

by

B. Silver and H. Ashley

12.1 Introduction

The present Chapter is a departure from the main theme of this book on structural optimization. The subject here is 'preliminary' or configuration design - and its optimization. Specifically, how is the 'best' combination of aircraft design parameters (such as wing loading, aspect ratio, tail area, etc.) to be selected in order to meet given system requirements? The study of configuration optimization is interactive and complementary to structural optimization; indeed, it is not difficult to foresee the day when both will coexist within a single computer program.

The two areas, structural optimization and configuration optimization, can both be classified under the heading of 'parameter optimization of non-linear systems', a rapidly growing domain of optimization theory. Thus the discussion of one area has implications for the other - and in fact, for many other optimization problems as well. In a broad sense, every engineering design problem is one of parameter optimization. At the risk of duplicating other material in this book, the present authors have sought to bring out the general nature of the design optimization problem, while still emphasizing the specific problem of aircraft configuration optimization. Further, they take the point of view that present-day methods of aircraft design optimization are a natural extension of past methods; that is, optimum-seeking computer programs have the same goal as the sliderule-wielding engineer of yore. Both want to find the 'best' airplane design. Both approaches rely on the designer's intuition for a first-guess, and both use iterative methods to improve this guess. The major advantage of optimization using the high-speed digital computer is that the space of design variables may be more exhaustively explored. The main disadvantage is that this space must be quantitatively defined - a process that de-emphasizes the roles of experience and intuition while inviting distortions and oversimplifications.

The iterative nature of the engineering design process is indicated in Fig.12.1. The 'search strategy' simply forms the feedback loop which attempts to improve the design. In the real-world ('reality') the feedback may come from operational experience with the actual airplane. Reality may be modeled, either with a physical model, such as a wind-tunnel model or a design mockup, or with a mathematical model ('quantitative abstraction') as shown. This abstraction requires inputs from the real world or from other models, and its results must be continuously compared with real-world results. This distinction between reality and abstraction should be kept in mind during computerized optimization for what is optimized is the model, not reality.

The 'model' involves the space of design variables which is searched. In Fig.12.1, the 'model' is comprised of (A') ('abstract domain of possible designs') and (B') (the value criterion: 'defined objective function and constraints'), which are analogues of (A) ('actual domain of possible designs') and (B) ('measure of actual value').

(A') contains the underlying physics of aircraft design, including aerodynamics, structures, propulsion, etc. (B') estimates the value of the aircraft design specified by (A'). Conceptually the model defined by (A') and (B') is the same one used in parametric analyses or tradeoff studies. The difference is that this model is driven in a sequential manner by (C'), the search strategy. Sequential search uses the results of previous iterations to select each new design, whereas a nonsequential search, such as a typical parametric analysis, iterates over a predetermined array of values.

12.1.1 A Comparison between 'Parametric Analysis' and Automated Search Methods

Parametric analysis methods are firmly rooted in the thinking of most aircraft designers. In this method a range of values of each of a number of parameters is analysed, the remaining parameters temporarily being held fixed. This is sometimes called a tradeoff study. The optimizer in this approach is the designer, whose judgment guides the selection of parameters. Of course the judgment of the designer must also remain active when he interprets the results of an automated optimization. Table 12.1 has been prepared to summarize the authors' views on the relative merits of these complementary approaches.

Selected references related to the use of parametric analysis in aircraft design are given in Reference Section 12A. One of the early efforts that brought together aerodynamics, structures, propulsion, performance and design into one aircraft synthesis program is SYNAC [12A.1] developed by General Dynamics Corporation. Having started as a parametric analysis program, SYNAC is moving toward an automated mode. Of course, once an integrated mathematical model such as this is developed, it is conceptually simple to 'drive' it with an automated optimizer. Every major aerospace company has a multiplicity of parametric computer programs. Both Hornburg of Douglas Aircraft [12F.1] and Hedrick of Grumman Aerospace [12F.2] believe that one of the major problems at present is obtaining a compatible integration of the various analysis and synthesis programs. A related problem is that the various computer programs within a company may not use a common data base. A reduction of the discipline-interface mismatches is one of the advantages of an integrated aerospace vehicle synthesis program.

Typical of present-day airplane parametric analysis programs is Boeing's 'Thumbprint' [12A.2]. The output of this program is plotted on transparent overlays to give the engineer a better view of the multi-dimensional space made up of the following: wing loading, thrust loading, gross weight, approach speed, maximum lift-to-drag ratio, takeoff noise level, initial altitude capability, takeoff field length and direct operating cost. This program is a useful tool for selecting proper tradeoffs in the preliminary sizing of a commercial transport.

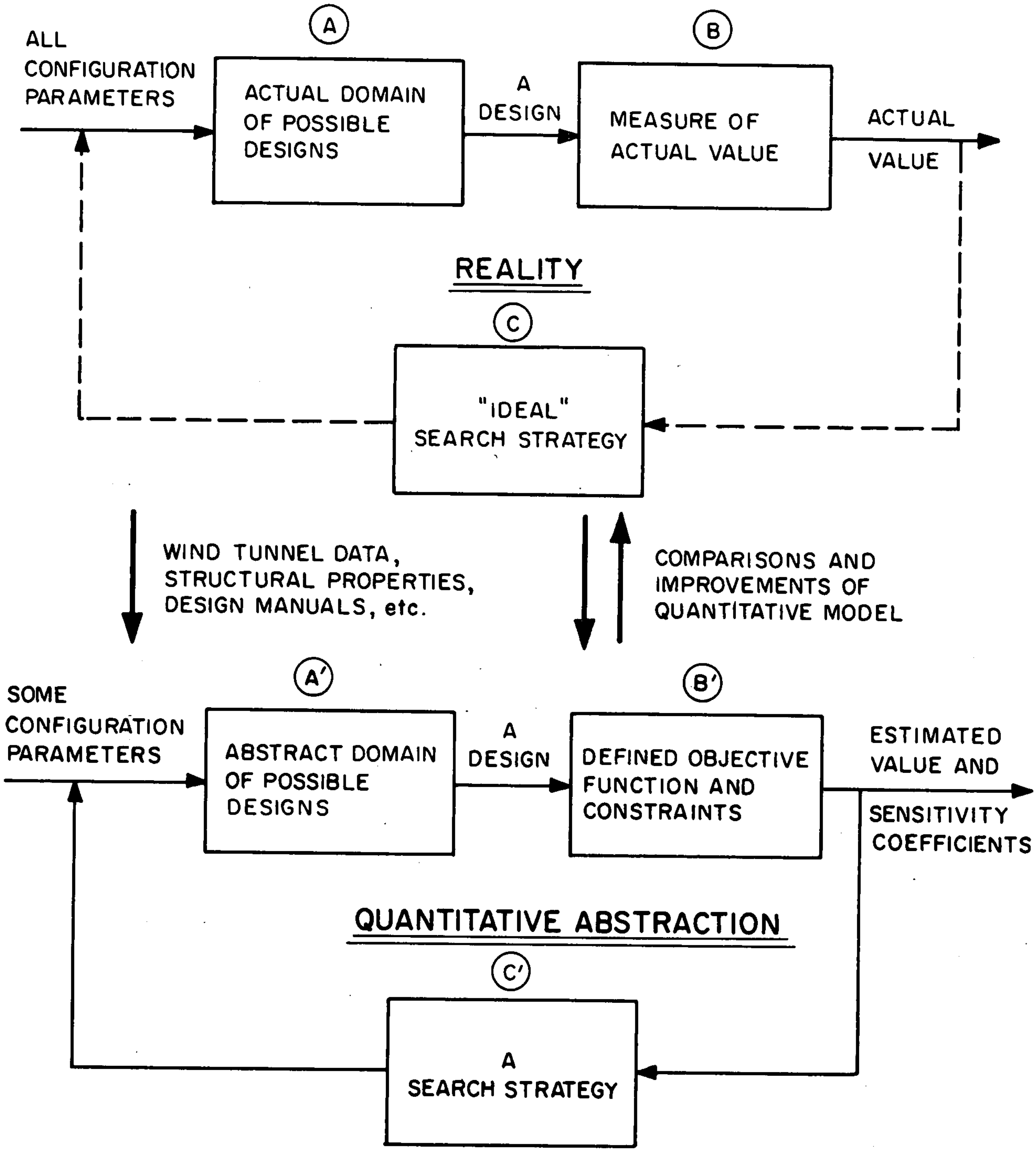


Fig.12.1 . The Aircraft Design Optimization Takes Place within a Quantitative Abstraction to Reality. It Contains Three Phases: (A′) the Quantitative Model (Underlying Physical Laws); (B′) a Definition of Value; and (C′) a Search Strategy. Each Phase must be Compared Against the Ideal of 'Reality'

Table 12.1

Comparison between Parametric Analysis and Automated Search Methods

	PARAMETRIC ANALYSIS	AUTOMATED SEARCH METHODS
ADVANTAGES:	1. Easy to program. 2. Traditional approach, matched to industry experience. 3. Maintains designer 'in the loop', thus exercising and strengthening his judgment. 4. Objective function does not need to be pre-specified. 5. Constraints do not have to be pre-specified. 6. Sensitivity to change in the parameters is apparent from the results ('maps' the space).	1. Potentially more efficient in locating optimum, particularly for higher-dimensional problems. 2. Potentially better convergence. 3. Applicable to a higher-dimensional space than man can manipulate. 4. Minimizes human bias. 5. Encourages more rigorous analysis. 6. Applicable to a wide class of problems, including those in which human intuition is not well developed (e.g. structural optimization with flutter constraints).
DISADVANTAGES:	1. Unwieldy and inefficient for higher-dimensional spaces. 2. Requires a good nominal or starting point. 3. Provides less certainty of obtaining optimum. 4. Must often be redone to extend the selected range of parameters. 5. Encourages the use of a restricted number of parameters.	1. Harder to program and debug. 2. 'Best' search method depends on the character of the problem. 3. Generally finds local rather than global optimum. 4. Can encounter convergence problems. 5. Search often tends to be driven outside the range of the mathematical model which is supported by physical data. 6. May not challenge and strengthen designer's intuition. 7. Sensitivity generally available only over linear range (does not 'map' the space).

The remainder of this Chapter will be devoted to methods of optimization that go beyond parametric analysis. Section 12.2 briefly discusses indirect methods of optimization and Section 12.3 describes some direct methods. The difference between these two approaches is that the indirect method solves an auxiliary problem $\left(\text{perhaps a generalized set of equations like } \frac{\partial y}{\partial X} = 0\right)$ while the direct method adopts a 'hill-climbing' strategy on the objective function $y(X)$ directly. Section 12.4 gives some operational results of direct search methods and Section 12.5 briefly describes the rapidly developing field of man-computer interactive design.

12.2 Indirect Methods of Optimization

Typical of indirect methods is the calculus of variations. The results of this approach can occasionally be elegant closed-form expressions, which represent solutions to a general class of problems. Unfortunately it has proved difficult to apply the calculus of variations to the optimization of systems as complex as an aircraft configuration. In fact, formal optimization of systems was not practical by any method until the advent of the modern computer and the powerful iterative methods typical of computer solutions.

Annotated references are included at the end of this Chapter. These references are divided into seven groups, two of which relate to indirect methods. Reference Section 12B lists a few recent books on the general subject of optimization (Ref. [12B.1] includes two introductory chapters on the use of variational calculus in optimization); Reference Section 12C deals specifically with indirect methods. Ref. [12C.1] is particularly concerned with extremal problems in the external shaping of aerodynamic surfaces, e.g. minimum-drag wings and bodies in supersonic and hypersonic flow. In [12C.2] the calculus of variations is used to obtain subsonic airfoil profiles of maximum section lift coefficient. Although drag is not constrained, the requirement of a fully attached boundary layer leads also to high predicted maximum lift-to-drag ratios (59 to 352, corresponding to a Reynolds number range of 10^6 to 10^7).

A recently developed indirect method is called 'geometric programming' and is associated with the names of Zener, Duffin and Peterson [12C.3]. Wilde and Beightler [12B.2], comment: "Geometric programming can now be used wherever a system is described by generalized polynomials. Potential applications abound because the technique is so new that only a few engineers have had time to put it to work. For problems with few degrees of difficulty, geometric programming promises to yield fast, accurate solutions to horribly non-linear problems. And when there are no degrees of difficulty at all, the method should produce rigorous 'rules of thumb' giving optimal component proportions that are completely independent of fluctuating prices and unit charges".

Apparently little has been done with geometric programming in aircraft optimization, although [12C.4] discusses its use in the design of V/STOL vehicles.

12.3 Direct Methods of Optimization

Methods of direct search are iterative methods which sequentially attempt to improve the defined objective function while satisfying given constraints. As shown in Fig.12.1, the search strategy (C') acts as a feedback loop driving the mathematical model made up of (A') and (B'). Although this Chapter is mainly concerned with (C'), it is worth noting that the real processes (A) and (B) are often more difficult to model than (C). Various researchers in the field of aircraft design (e.g. [12F.1] through [12F.6]) have singled out block (A') as presenting particular difficulties. Among the areas of difficulty mentioned were the following: inadequate theory for predicting aerodynamic forces on an arbitrary three-dimensional body; insufficiently accurate, simple methods of aircraft weight estimation; and inaccurate cost estimation techniques. Regarding the value criterion (B'), in general the aircraft configuration that is 'optimum' is very sensitive to the definition of this value criterion. Because of this behavior, the value criterion should continually be re-examined as the design evolves. This process forms an integral part of design optimization.

The value criterion in (B') is mathematically expressed as an 'objective function', which is to be optimized, as well as various constraints (inequality and/or equality) that are simultaneously to be satisfied. A conceptual difficulty with this formulation is that a single objective function must be defined. In the design of a commercial air transport, for example, this objective function might be direct operating cost (DOC). The designer might also wish to minimize takeoff distance (TOD). (Other parameters will be neglected for simplicity in this example.) A request to minimize both DOC and TOD independently would make no sense. How much TOD are you willing to give up for an improvement in DOC? At this point the designer has two alternatives: he can specify TOD as an inequality constraint, such as TOD $\leq$ 5000 ft, or he can multiply TOD by a selected constant (which implies the acceptable tradeoff between TOD and DOC) and add this quantity to DOC to form a new objective function. Either approach requires a constant to be selected beforehand. The implications of this selection should be tested by a sensitivity analysis. This refinement introduces an outside iteration loop to the process described by Fig.12.1.

Suppose that in the above example the designer chooses the constraint, TOD $\leq$ 5000 ft. After minimizing DOC while satisfying this constraint, a sensitivity analysis would determine the change in optimized DOC implied by a small change in the constraint. Suppose that a 1% decrease in TOD caused an increase in DOC of only 0.001%. In this case the sensitivity coefficient is $S_{DOC/TOD} = -0.001$. Such a small sensitivity would tempt the designer to select a shorter takeoff distance. On the other hand, if the sensitivity coefficient were -0.2 instead of -0.001, the designer might wish to relax the TOD constraint. Obviously the logical selection of the constraint depends on the associated sensitivity. In this case, the first author suggests that the constraint be placed on the sensitivity coefficient itself, rather than on the parameter. In the present example the designer might try replacing the constraint on TOD with the constraint, $S_{DOC/TOD} \geq -0.05$.

Many constraints can arise from an incompleteness of the model. If in the above example a higher level goal, say profitability to the manufacturer, were made the objective function, the constraint dealing with TOD might be eliminated. Of course, it would then be necessary greatly to expand the mathematical model to include a description of the selection criteria employed by airlines, whose representatives would consider takeoff distance. However, this process of extending the model to eliminate constraints can cause serious complications and increase the likelihood of error. The natural approach would be to start with simple models and proceed to greater complexity as required.

This discussion is intended as an introduction to the problem of setting up a model to be optimized using a direct search. Subsequent sub-sections will discuss in more detail the problem formulation, selected direct search methods and convergence criteria.

12.3.1 The Selection of Design Variables for Direct Methods

The efficiency of an optimization search and its chances of success depend strongly on the manner in which the problem is stated. This section will discuss some implications of direct search characteristics on the selection of the design variables. In every design problem there are alternate ways of describing the design. For example, any two of the following four wing variables imply the other two: span (b), mean chord ($\bar{c}$), area ($S = b\bar{c}$), and aspect ratio ($A = b/\bar{c}$). Any two, such as b and $\bar{c}$, could be selected as design variables. One would be insufficient, and three (or four) would be too many and would lead to what is known as 'ill-conditioning', i.e. non-independence of the variables.

The goal in selecting design variables is to reduce as much as possible the interaction between them. When a strong interaction exists there is a 'ridge' (or its inverse, a 'ravine') in the objective function. It has been found that most direct search methods encounter difficulty with ridges ([12B.1], p.268 and [12B.2], p.283). For all direct methods the easiest functions to optimize are those whose constant-value contours are circles. Of course this phenomenon rarely occurs in physical systems, but it is generally possible to make the contours more nearly circular (at least locally) by rescaling (and perhaps rotating) the variables. The designer should use his understanding of the problem in his selection and scaling of design variables.

For example, suppose the designer must choose two of the four wing variables mentioned above (b, $\bar{c}$, S and A). Many definitions of objective function would lead to a strong dependence on wing loading and thus on wing area, S. Since $S = b\bar{c}$, a strong interaction would exist between b and $\bar{c}$, and they would make a poor pair of design variables. In this case, contours of the objective function in the $b - \bar{c}$ plane would disclose a diagonal ridge. The selection of wing area and aspect ratio would be a better choice in this example.

Among the various ways in which the design variables can be defined, it usually proves most efficient to make as many as possible dimensionless. (In general, however, at least one design variable with dimensions will be required to specify the size of the aircraft.) The choice of dimensionless variables is particularly advantageous when they have direct physical utility, such as aspect ratio or tail volume coefficient [$\bar{V} = (S_t \ell_t)/(S\bar{c})$, where S_t = tail area and ℓ_t = tail length].

In aircraft optimization some design variables are more important than others. It is often worthwhile to order these variables according to their expected importance. Where computer time is a consideration, one then has the option of searching over just the most important variables. This is called a 'reduced-space search' and is useful either when the sophistication of a full-space search is not required or as a starting procedure for a full-space search.

The non-dimensionalization of the problem can be carried further, as discussed in the next section. One possibility is that all terms in the problem statement may be normalized, in order to decouple the search from the dimensions of the original problem. This accomplishes the rescaling suggested in this section.

12.3.2 Problem Statement and Constraint Formulation

The authors have attempted to limit the mathematical content of this Chapter, since its main purpose is to discuss concepts and report experience. However the normalized formulation presented here is considered to be of practical value in setting up a design problem for computer optimization. The reader uninterested in mathematical details may skip to Section 12.3.3.

The general problem of non-linear programming may be stated as follows: Find the vector α of n variables α_i, $i = 1,2,\ldots n$ (called the 'design variables' in this Chapter) which

(1) minimizes the scalar objective function $y(\alpha)$, subject to (12-1)

(2) design variable limits (α_{LO_i} and α_{HI_i} represent the low and high limits, respectively, and will be used later to normalize α_i),

$$\alpha_{LO_i} \leqslant \alpha_i \leqslant \alpha_{HI_i} \quad , \quad i = 1,2,\ldots n \; ; \tag{12-2}$$

(3) equality constraints,

$$e_j(\alpha) = b_j \quad , \quad j = 1,2,\ldots J \; ; \tag{12-3}$$

(4) and inequality constraints (sometimes called 'restraints'),

$$f_k(\alpha) \geqslant c_k \quad , \quad k = 1,2,\ldots K \; . \tag{12-4}$$

Here b_j and c_k are selected constants with dimensions appropriate to each constraint. These constants will be used to normalize the constraints, a technique which obviates the need for individual weighting constants in the development of the constraint penalty functions later in this section.

It is convenient to develop optimization search routines (block (C') in Fig.12.1) which do not have to consider dimensions and unequal orders of magnitude in the various design variables. For example, the large differences in magnitude between typical values of static longitudinal stability margin ($\sim$0.05), wing area ($\sim$1000 ft^2), and Reynolds number ($\sim 10^7$) would lead (if these parameters were used as design variables) to another form of ill-conditioning in most search methods. For use in the search subroutine it is cleaner to normalize the problem statement in the following manner:

(1) define normalized design variables*, X_i,

$$\left.\begin{aligned} X_i &\triangleq \frac{\alpha_i - \alpha_{LO_i}}{\alpha_{HI_i} - \alpha_{LO_i}} \; , \\ 0 &\leqslant X_i \leqslant 1 \; , \end{aligned}\right\} \; i = 1,2,\ldots n \tag{12-2'}$$

*The symbol $\triangleq$ means 'equal by definition'.

(2) define normalized equality constraints,

$$E_j \triangleq e_j/b_j = 1 \quad , \quad j = 1,2,\ldots J \quad ; \tag{12-3'}$$

(3) and define normalized inequality constraints,

$$F_k \triangleq f_k/c_k \geqslant 1 \quad , \quad k = 1,2,\ldots K \quad . \tag{12-4'}$$

Finally the objective function may be normalized within the program by the initial value of y at the input point (specified by α_{IN_i}, or equivalently, X_{IN_i}). Define the normalized objective function,

$$Y(X) \triangleq y(X)/y(X_{IN}) \quad . \tag{12-1'}$$

By choosing the normalizing constants in a physically meaningful way, the designer accomplishes the rescaling suggested in the previous section. For example, suppose that Reynolds number were a design variable whose range was limited by other considerations in the problem. The designer could then arbitrarily select a normalizing range ($\alpha_{HI} - \alpha_{LO}$) which included the actual range. That is, knowing that Reynolds number would never exceed 5×10^7, he could use this value for the upper limit and zero for the lower limit, thus rescaling this design variable to the range zero to one. Note that in the normalized formulation all variables are of order unity. Thus the designer has a good 'feel' for the percentage change in each variable. The normalization may expose similarities between various design optimization problems. These analogies might have the effect of improving the judgment of the design engineer.

12.3.2.1 Problem Statement Example

An interesting example of an aircraft whose design problem might have been simply stated for computer optimization is the Lockheed U-2 high-altitude research and reconnaissance aircraft. Although designed long before non-linear optimization techniques were well developed (the U-2 was designed in 1954, with first flight in 1955), the U-2 design problem might have been formalized as something like the following:

Maximize the service ceiling, h_s, or, equivalently, minimize $-h_s$, subject to constraints,

$$e_1 \triangleq \text{payload weight} = b_1$$

$$f_1 \triangleq \text{range} \geqslant c_1$$

$$f_2 \triangleq -\text{(static stability margin)} \geqslant c_2$$

$$f_3 \triangleq -C_{n\psi}\ \text{(directional stability)} \geqslant c_3 \quad .$$

The constraint constants might have been $b_1 = 2000$ lb, $c_1 = 4000$ miles, $c_2 = +0.05$ and $c_3 = +0.001$. The number of design variables is determined by the sophistication of one's mathematical model, but a few typical design variables are:

$\alpha_1 \triangleq$ wing area, S (ft^2)

$\alpha_2 \triangleq$ wing loading, W/S (lb/ft^2) (implies gross weight, W)

$\alpha_3 \triangleq$ wing aspect ratio

$\alpha_4 \triangleq$ wing chord taper ratio

$\alpha_5 \triangleq$ wing root thickness ratio, $(t/c)_{ROOT}$

$\alpha_6 \triangleq (t/c)_{TIP}/(t/c)_{ROOT}$

$\alpha_7 \triangleq$ wing twist ('washout')

$\alpha_8 \triangleq$ ratio of span to tail length

$\alpha_9 \triangleq$ ratio of stabilizer area to wing area

$\alpha_{10} \triangleq$ ratio of fin area to wing area

$\alpha_{11} \triangleq$ fuel weight fraction, W_{FUEL}/W .

Additional design variables which could be investigated include dihedral, incidence, sweep, design lift coefficient (root and tip) for the wing; aspect ratio, thickness ratio, taper ratio, sweep, control surface area for both stabilizer and fin; fineness ratio, cross-sectional area, camber for the fuselage; thrust vector angle; plus a number of other variables describing the propulsion unit, the inlets, the control system, etc. When these additional parameters are not design variables, their values would be assumed or determined by other analyses.

The appropriate design variable limits might be:

$$0 \leqslant \alpha_1 \leqslant 1000 \ (\mathrm{ft}^2)$$

$$0 \leqslant \alpha_2 \leqslant 50 \ (\mathrm{lb/ft}^2)$$

$$0 \leqslant \alpha_3 \leqslant 20$$

etc.

Note that most of the selected design variables are dimensionless. This approach improves the ease of scaling the aircraft, both within and without the computer program. For use within an optimization program, the entire problem statement could be normalized as in Section 12.3.2.

No actual solutions can yet be reported on this particular design.

12.3.2.2 Constraint Formulation

There are two basic constraint formulations. One changes the constrained problem into an unconstrained problem by adding a penalty function, based on constraint violations, to the original objective function. The other attempts to pick search directions that both satisfy the constraints and improve the objective function. The penalty function approach is discussed by Fox in Chapter 6. Examples of the second approach are described in Chapters 5 (sequence of linear programs) and 7 (feasible directions methods) by Pope and Kowalik respectively.

The normalized form developed in Section 12.3.2 is convenient for the following penalty function formulations:

'Interior' penalty function, P_I:

$$P_I \triangleq \sum_{k=1}^{K} \frac{1}{F_k - 1}, \qquad F_k \geqslant 1 \tag{12-5}$$

'Exterior' penalty function, P_E:

$$\left. \begin{aligned} P_E &\triangleq \sum_{j=1}^{J} (E_j - 1)^2 + \sum_{k=1}^{K} (F_k - 1)^2 \delta_k \\ \text{where} \quad \delta_k &= \begin{cases} 0, & F_k \geqslant 1 \\ 1, & F_k < 1 \end{cases} \end{aligned} \right\} \tag{12-6}$$

(E_j and F_k are defined in (12-3') and (12-4'), respectively).

A new objective function is formed by adding the weighted penalty function to the old objective function:

$$\text{'Interior' form: minimize } y_I = y + W_I P_I \tag{12-7a}$$

$$\text{'Exterior' form: minimize } y_E = y + W_E P_E \tag{12-7b}$$

The weighting factor, W_I or W_E, must be adjusted as the search progresses to insure that the constrained optimum is approached. Refer to Pierre [12B.1], p.338, and to Chapter 6 for discussions of weighting factor control. Either penalty formulation may be used, but the 'interior' form requires a feasible starting point and is difficult to use when equality constraints are present. Some engineers prefer the interior form, however, because solutions lie inside the constraint boundaries (hence the name, 'interior') and are thus conservative.

The penalty function warps the objective function and creates a two-sided 'ravine' for an equality constraint and a one-sided 'cliff' for an inequality constraint. These imposed non-linearities make the search more difficult. In many cases a constraint may be incorporated into the model to eliminate one of the design variables. In fact, the U-2 design example of the previous subsection was artificially constructed so that all of the constraints could be easily eliminated. The payload weight and range constraints could be used in an internal loop to scale the gross weight, thus eliminating α_2 and α_{11}, and the stability constraints could be used directly to size the tail areas, thus eliminating α_9 and α_{10}. This approach is preferred since it reduces both the number of constraints and design variables in the search driver (although they are still implicit in the mathematical model).

Finally it is noted that an inequality constraint is preferable to an equality constraint in the penalty formulation, because it reduces the likelihood of creating contradictory requirements (an equality constraint is always 'on') and because it forms only a one-sided warping of the space.

Most aircraft design constraints, such as take-off distance, rate of climb, stall speed, cruise speed, etc., can be more satisfactorily imposed as inequality constraints, anyway.

12.3.2.3 A Penalty Function for Integer Design Variables

Certain meaningful design variables can only take on integral values. Such variables include the number of engines, the crew size, and the number of windows, landing gears or wheels, hydraulic systems, etc. Many of these variables form an important part of a design analysis, and two techniques are suggested here for their introduction.

The first and simplest technique is to allow the integer variables to accept non-integer values during the first optimization run and then to fix their values at the nearest integer for a subsequent optimization run. The logic for this can be incorporated into the optimization program.

The second technique employs a penalty function which drives selected variables to integer values. As an example, the first author suggests a penalty (to be added to the objective function) of the form,

$$W_{INT} \sum_{i_{INT}} \sin^a \left(\pi \, \eta_{i_{INT}}\right)$$

where W_{INT} is a weighting factor, a is a selected constant, and $\eta_{i_{INT}}$ is the fractional part of the ith integer design variable, $X_{i_{INT}}$ (in Fortran IV, IBM 360, $\eta_{i_{INT}}$ = AMOD ($X_{i_{INT}}$, 1)).

As shown in Figs.12.2a and 12.2b, the exponent a can be used to smooth out the cusps at the integer values. The first order derivatives are continuous for $a > 1$ and the second order derivatives are continuous for $a > 2$, etc.

Because this penalty acts to prevent the natural migration of the selected design variables during an optimization, it is suggested that the weighting factor, W_{INT}, be kept at zero until 'rough convergence' is obtained ($X_{i_{INT}}$ are within ±0.5 of their estimated converged values), and then that W_{INT} be sequentially increased until all $X_{i_{INT}}$ fall within a given ε of being integers.

12.3.3 Summary of Selected Direct Search Methods

It is the intent of this section to discuss the interrelationship between aircraft design optimization problems and various direct search methods which may be applied to their solution. Previous sections have stressed the need to formulate the problem in a manner well suited for direct methods in general. This section briefly describes specific search methods and the types of problem for which each is suited. This description also lays the groundwork for Section 12.4, in which operational experience using these methods is reported.

The remainder of this Section attempts to develop the reader's intuitive understanding of direct methods; subsections 12.3.3.1 through 12.3.3.3 summarize specific methods.

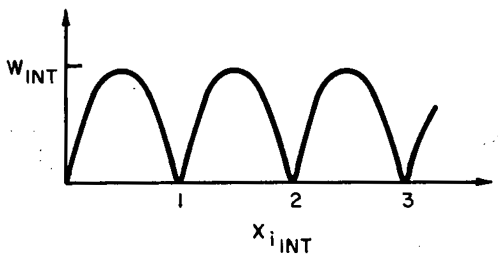
Most direct methods employ two distinct search strategies: one for direction selection and one for search along the direction. Direction selection strategies are discussed in subsections 12.3.3.1 and 12.3.3.2. A search along a direction is a one-dimensional search; techniques for this are listed in subsection 12.3.3.3. A semantic distinction is made in this Chapter between the one-dimensional iterations, called 'steps' and the overall movements made in each direction, called 'moves'. Fig.12.3a indicates this distinction. This figure shows the contours of the objective function in the plane of two design variables. For example, the objective function could be $(L/D)_{MAX}$ - that is, maximum lift to drag ratio - for a sailplane, and the two design variables could be normalized wing loading and normalized aspect ratio. This type of plot suggests the intuitive topographical images of peaks, ridges, passes (saddle points), ravines, etc. Fig.12.3a was drawn with two peaks. Direct methods generally stop at the first local optimum encountered, as shown (point 5).

Fig.12.3b indicates the dual-loop nature of direct methods. The inner loop is the one-dimensional search ('steps'), and the outer loop 'moves' to the best point thus found and then selects a new direction. The number of steps per move should be balanced against the requirements, both in time and resolution, of the direction strategy.

A numerical evaluation of the gradient at a point requires n perturbations, each of which is as time-consuming as one step since each requires one objective function evaluation. For aircraft design problems the number of function evaluations is generally a good measure of computer time expended. The time-consuming nature of numerical partials has encouraged the development of a number of search methods which do not require partials; these methods are discussed in subsection 12.3.3.1. When analytical partial derivatives are available - and often when they are not - methods employing the partials may be used; these are discussed in subsection 12.3.3.2.

12.3.3.1 Direct Search Methods without Derivatives

Mathematical models for aircraft configuration optimization often require many layers of computation, many times with internal loops. In addition, tabular input data are generally used. These factors make it difficult to obtain analytic partial derivatives. In this section, seven direct methods which do not require derivatives are briefly described. These short descriptions are intended for



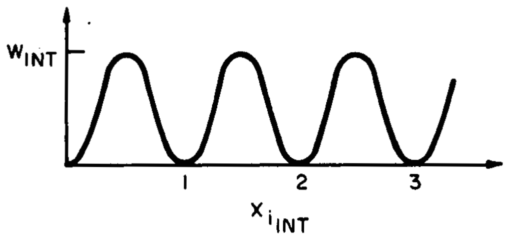


Fig.12.2a Penalty Function for Integer Programming (a = 1

Fig.12.2b Penalty Function for Integer Programming (a > 1

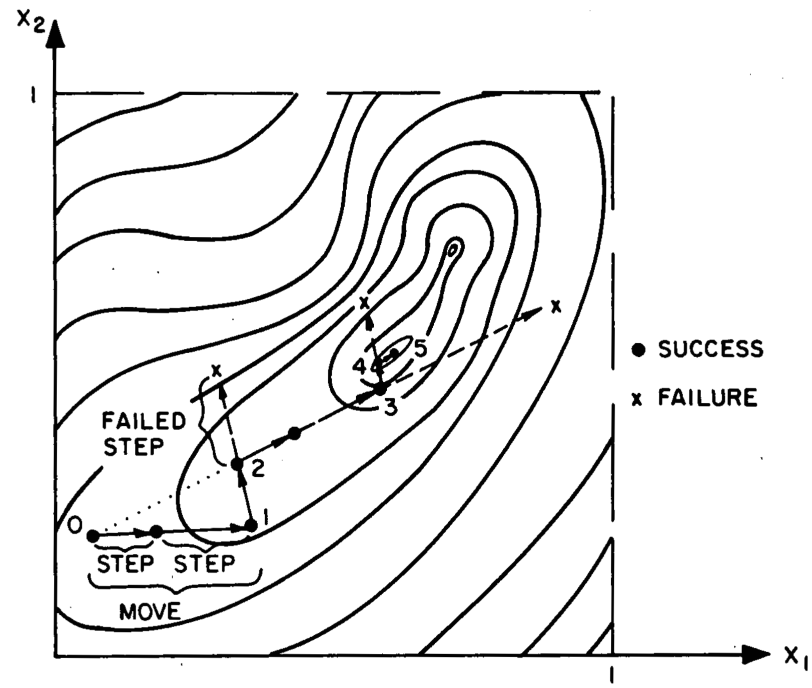


Fig.12.3a 'Steps' and 'Moves' in a Two-Dimensional Search

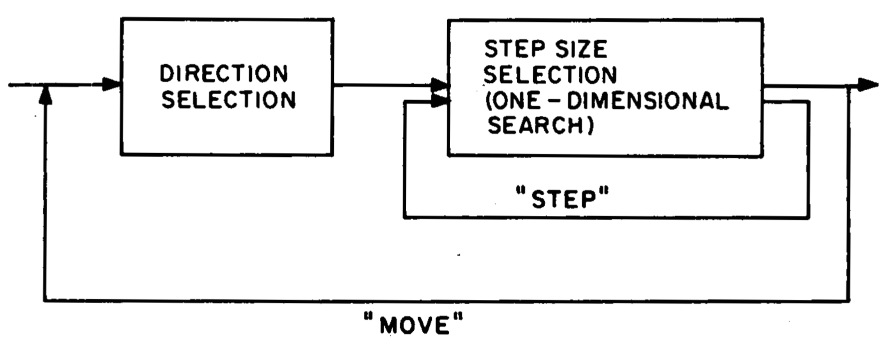


Fig.12.3b Conceptual Flow Diagram for the Two Search Phases: 'Steps' and 'Moves'

identification of the methods (often, different names are used for the same method) without going into the mathematical details. The interested reader is referred to the references cited.

Most of the direct methods discussed in this Chapter have been employed in a computer program called AESOP, 'Automated Engineering and Scientific Optimization Program'. This Boeing Company program is described in [12D.1], [12E.1], [12F.7] and [12F.8]. It is the most extensive parameter optimization driver known to the authors. It has nine search strategies, which may be used singly (except for 'Pattern') or in combination. Storage is provided for 100 design variables, and an unlimited number of constraints may be employed through a penalty function. AESOP has been applied to many aerospace problems including aircraft design, [12E.1], [12E.3], [12E.5], [12F.9]. The results of some of these studies will be discussed in Section 12.4.

The names and descriptions for methods 1 through 6 below are taken from AESOP, [12D.1], p.3:

(1) "SECTIONING - Succession of one-dimensional optimization calculations parallel to coordinate axes. Variables may be perturbed in random or natural order."

(2) "ADAPTIVE CREEPING - Search in small incremental steps parallel to the coordinate axes. Step-size adjusted automatically in the algorithm. Variables may be perturbed in random or natural order."

(3) "PATTERN - A Ray Search in the gross direction defined by a previous search or search combination." This is the acceleration move originally suggested by Forsythe and Motzkin in [12D.6]. Move 2-3 in Fig.12.3a is an acceleration move in the direction defined by the sum of the two previous moves.

(4) "MAGNIFICATION - Straightforward magnification or diminution about the origin."

(5) "RANDOM POINT - Function to be optimized is evaluated at a set of uniformly distributed random points in a specified region."

(6) "RANDOM RAY SEARCH - Function is optimized by search along a sequence of random rays having a uniformly distributed angular orientation in the multivariable parameter space."

(7) BEST-TRIAL SEARCH (Rastrigin, [12D.4]) - A cross between Random Ray and Statistical Gradient*, this method tests m random directions and then performs a one-dimensional search in the most promising direction.

12.3.3.2 Direct Search Methods with Derivatives

When analytic partial derivatives are available, the added computer time required for their calculation is generally small in comparison with the time required for the objective function evaluations alone (not to mention the time required for the n evaluations of numerical partials). The information contained in the partial derivatives can generally be used to improve the search procedure.

Again, the following short descriptions are intended to identify the methods only. The names and descriptions of methods 8 through 10 are taken from AESOP, [12D.1], p.3:

(8) "STEEPEST DESCENT - Search along the weighted gradient-direction. Several weighting options available."

(9) "DAVIDON'S METHOD - An attempt to achieve the advantages of second-order search from an ordered succession of first-order searches." Refer to Chapter 6 and to Davidon [12D.7], Fletcher and Powell [12D.8]; also [12B.1], p.320, and [12B.2], p.331.

(10) "QUADRATIC - Second-order multivariable curve fit to the function being optimized, followed by search in direction of second-order surface optimum."

(11) PARTAN (from "PARallel TANgents") - Introduced by Shah, et al. [12D.12]. This method selects directions which are parallel to tangent planes of previous moves; alternating moves are simple acceleration moves (Method 3). "Tangent plane" refers to the hyper-plane which is tangent to the objective function hyper-surface in n+1 space. This method is illustrated in Section 7-12 of [12B.2].

Some gradient methods, particularly Davidon and Partan, have proven quite powerful in certain applications, even when the partials were evaluated numerically.

12.3.3.3 One-Dimensional Search Methods

Most of the search methods discussed so far merely determine the direction to be searched without specifying how the one-dimensional search in this direction is to be conducted. In this subsection a few standard one-dimensional search methods will be described.

If a slice is taken of the objective function in a selected direction it might look like Fig.12.4. The starting point is the best point found on the previous move. The first step size may be determined by previously successful step sizes. Some strategy must be used to size subsequent steps. In Fig.12.4 the steps are simply doubled until an improvement is no longer obtained.

---

*The gradient is estimated as the scaled sum of m random perturbation vectors, where $m < n$.

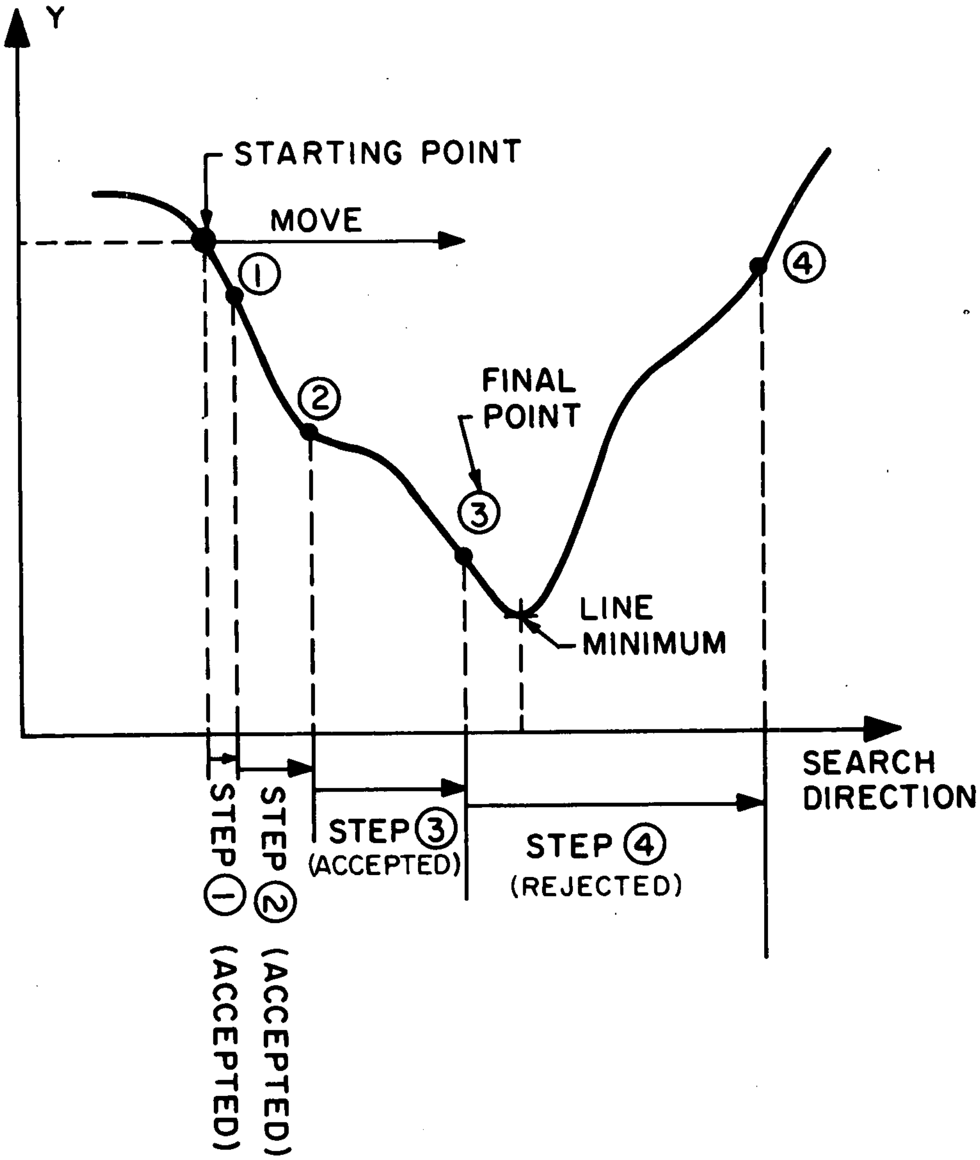


Fig.12.4 One-Dimensional Search. The Simple Search Strategy used here is to keep Doubling the Step Size until the Function no Longer Improves. The Move is made to the Last Accepted Point

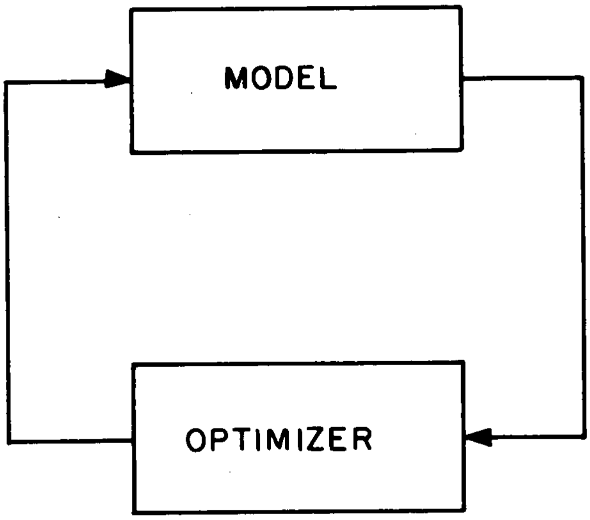


Fig.12.5 Separate Nature of Model and Optimizer

Since there may be many steps per move, the efficiency of the overall search is obviously dependent on the speed of the one-dimensional method employed.

Some one-dimensional strategies are:

(A) ONE-STEP - The simplest of all strategies; only one step is taken in each direction. The size of the step may be chosen on the basis of recent successes. For example, if the last move were a success, the present step could be double the last step; if a failure, one-half or one-fourth.

(B) ONE-STEP PLUS REVERSAL - A modification of the one-step, this strategy tries a step in the reverse direction if the first step is a failure. This method is used in AESOP's 'adaptive creeping'.

(C) STEP UNTIL FAILURE - This method continues stepping until the function stops improving, then moves to the last accepted point. Generally each step size is selected as some multiple of the previous step; that is, $S_j = KS_{j-1}$ where S_j is the jth step size on this move, and K is a selected constant. In Fig.12.4, $K = 2$. A rapidly increasing step size, such as $K = 10$, can be used to bound the line optimum, a necessary requisite for interval elimination methods.

(D) INTERVAL ELIMINATION - Fibonacci search and golden-section search both attempt to reduce the interval in which the line optimum lies to be small as possible, [12B.1], p.280; [12B.2], p.236 and [12B.3]. The difference between the methods is that the number of steps is pre-specified for the Fibonacci search and not for the golden-section search. The Fibonacci search reduces the interval by a factor of F_N. F_N, the Fibonacci number comes from the series (N = number of steps = 1,2,3,4...): 1,2,3,5,8,13,21,.... Note that $F_N = F_{N-1} + F_{N-2}$ for $N > 2$. The golden-section search is slightly less efficient and reduces the interval by $(0.618)^{N-1}$, the advantage being that the number of steps, N, is not pre-specified. For five steps the reduction ratio is 1/8 (0.125) for Fibonacci and 0.145 for golden section.

(E) CURVE FIT METHODS - A polynomial of degree N may be fitted to N points found in the one-dimensional search. The location of the optimum on this line can then be estimated by setting the first derivative equal to zero. If the function evaluated at this point matches the predicted value within some ε, the one-dimensional search is terminated; if not, this point is used to improve the polynomial, and a new optimum is estimated. A least-squares curve fit can be used where the number of points exceeds the degree of the polynomial. For further discussion see Chapter 6 (Fox), and also [12B.1], p.274.

12.3.4 Convergence Criteria for Direct Methods

As will be seen in Section 12.4 all direct methods do not converge on all problems. In fact, most commonly used stopping criteria do not guarantee convergence. It is often up to the engineer to determine how near his result is to the converged optimum. Strictly speaking a (local) optimum is assured only if necessary and sufficient conditions are fulfilled. The necessary condition requires the constrained partial derivatives of the objective function with respect to the design variables to be zero [12H.3], [12B.2]. The sufficient condition, which examines the matrix of second order partials of the constrained objective function for positive definiteness (for a minimum), can generally be forgone for aircraft configuration optimizations because it is obvious whether a maximum or a minimum is obtained (generally only one makes any sense). This leaves the requirement for only the first order partials, but even these are often not available. When they are available (subsection 12.3.3.2) the proper test is that all constrained partials be 'small', where 'small' is set by the engineer as a tradeoff between computer time and nearness of convergence.

When the test on partials can't be performed, some standard stopping criteria are:

(A) Function evaluations exceed a specified number.

(B) Sequential failures exceed a specified number.

(C) Step size drops below a given limit.

(D) Improvement in the objective function between iterations drops below a given level.

Two methods are available which often discriminate against a false optimum. One is to 'map' the region in the vicinity of the result; that can be done with a simple parametric analysis at the supposed optimum. Another method, which in the authors' experience is very effective, is to run the search again from a few different starting points (see for example, Table 12.2, Section 12.4.1).

Examination of the region of the supposed optimum is valuable not only for testing convergence but for refining the value criterion. For example, the design of a hypersonic cruise vehicle is discussed in Section 12.4.1. The objective function is number of passengers. It appears from Table 12.2 that the optimum is rather flat with respect to wing loading. Other considerations, such as approach speed, might cause one to select a lower wing loading without much degradation of the objective function, number of passengers.

12.4 Operational Experience with Direct Methods

The number of parameter optimization studies is growing rapidly. Applications in the field of aircraft design, however, have lagged those in certain other fields. Stepniewski and Kalmbach [12F.9] comment: "... there is much less optimization activity in the domain of aeronautics than in astronautics, or even in chemical processes".

It may be expected that this situation will change as more aeronautical engineers become aware of the power of optimization methods. One realization that will aid in this process is the fact that existing mathematical models presently used in parametric analyses can be rather simply modified to be driven by separate optimization packages. In Fig.12.5, the 'Optimizer' completes a feedback loop on the mathematical 'Model' which previously may have been used open-loop for a parametric study. General Dynamic's SYNAC, a large parametric aircraft design program, is moving toward automated search capabilities.

Boeing, recognizing the separate nature of the Model and the Optimizer, has developed the aforementioned large, general purpose parameter optimization package called AESOP [12D.1], [12E.1], which has been applied to a variety of optimization problems. This program and its results will be discussed in more detail in Section 12.4.1.

Some foresight in the structuring of the mathematical model, as suggested in Section 12.3, can ease the transition to the optimization mode. In particular, it is required that all design variables be available for external manipulation.

Before comparing results of various direct methods it is important to note that more criteria than speed should be applied in comparing methods. Speed of convergence is certainly important but other considerations are the following:

(1) Degree of convergence.

(2) Robustness. Convergence is reliably obtained for a variety of initial conditions, constraints and noise in the data.

(3) Computer memory requirements.

(4) Ease of programming and debugging.

(5) Output capabilities. Does the method calculate sensitivity coefficients, partial derivatives, etc.?

12.4.1 Operational Experience with AESOP

AESOP was developed by the Boeing Company starting in 1965 under the direction of D. S. Hague [12D.1]. This general-purpose optimizer can drive up to 100 non-linear parameters, using requested combinations of nine search methods (described in Section 12.3.3 as Methods 1 through 6 and 9 through 11). AESOP has been applied to several aeronautical problems including [12D.1]:

(1) Two-dimensional minimum drag supersonic airfoil shaping.

(2) Minimum drag supersonic bodies of revolution.

(3) Minimum drag hypersonic bodies of revolution.

(4) Three-dimensional supersonic airfoil shaping.

(5) STOL preliminary design [12F.9].

(6) Hypersonic cruise vehicle preliminary design [12F.7].

The hypersonic cruise vehicle optimization problem is reported in detail in [12F.7]. A nominal vehicle was designed by Ames Research Center (NASA) using conventional preliminary design techniques. This vehicle, shown in Fig.12.6, had the following specifications: 500 000 lb gross weight, 5500 nm range and a speed of Mach 6. The nominal payload turned out to be 220.3 passengers*. Five design variables ('parameters') had nominal values as listed in Table 12.2. This table also shows an off-nominal starting point (to check sensitivity of the optimum to initial conditions) and the final design points, obtained from each of these starting points by using an 'adaptive creeping' search method (Method 2 of subsection 12.3.3.1). The results indicate an improvement in the objective function, number of passengers, of 33 or 15%. The fact that the design variables do not converge to the same values for the two different starting points indicates either that true convergence has not been obtained or that the optimum is relatively flat in some direction in the design space. Since a number of other optimization methods resulted in almost the same maximum number of passengers, it might appear that the optimum is indeed rather flat. Note that the final wing loading varies 6% between the two cases.

The hypersonic cruise vehicle optimization problem is probably representative of aircraft configuration designs in regard to the relative success of various optimization methods. In this study four methods were used to optimize the hypersonic vehicle. Two simple univariate (only one design variable is changed at a time) methods, 'sectioning' and 'adaptive creeping' were found to be superior to two more complicated methods, steepest descent and 'quadratic'. The results of these four methods are shown in Table 12.3.

Although the adaptive creeping technique worked best on this example, it should be pointed out that five of AESOP's nine search options were not tried, and one of them might have proved better. The univariate methods which worked so well in this example cannot be expected to show similar success against a coupled surface, i.e. a ridge. When a univariate method encounters a ridge it starts to

*Number of passengers is treated as a continuous function. In the final design this number would be rounded off to an integer.

'zig-zag' along the ridge and is very slow to reach convergence. In fact it may stop quite far from the optimum. An example of this phenomenon for the sectioning method is shown in Fig.12.7. The fact that the hypersonic cruise vehicle response surface is relatively uncoupled is pointed out in [12F.7] and [12F.9].

Table 12.3 shows that the steepest descent method was much slower than either of the univariate methods. The actual performance was, in a sense, worse than indicated, because it was necessary to develop a complete empirical weighting matrix even to obtain these results. Unweighted steepest descent and steepest descent with a weighting matrix based on first derivatives both failed to converge. In view of the fact that steepest descent has been one of the most popular methods in parameter optimization, these results are a warning. (The accelerated steepest descent, that is, Methods 8 and 3 of Section 12.3.3, was not tried and might be expected to provide better results.)

The quadratic method, which requires second-order numerical perturbation about a point to fit a quadratic surface, gave the worst performance. A clue as to why the two sophisticated methods failed to perform well on this problem is provided in [12F.7], p.42: "... at this scale the response surface is quite irregular". (Numerical experiments had indicated that the mathematical model gave noisy results at the fine scale.) "... numerical derivative calculations in the steepest descent, Davidon, or quadratic searches could also be in serious error if the control parameter perturbations used in their calculation is too small."

Additional parameters and constraints were introduced into the hypersonic cruise vehicle optimization. The successive numbers of design variables used were 6, 11, 17 and 28. Actually, in the two latter cases, certain 'design variables' were climb-trajectory parameters. This selection indicates that a continuous function, such as a trajectory, can be optimized using discrete elements. One of the interesting results of the higher-dimension study was that the number of evaluations required for the 11 variable optimization was only slightly higher than that for the 6 variable optimization.

Table 12.2

Hypersonic Cruise Vehicle Optimization
Using Adaptive Creeping Search [12F.7]

Parameter	Ames nominal		Off-nominal	
	Start	Finish	Start	Finish
Wing loading (lb/ft^2)	80	108.5	120	115.2
Aspect ratio	1.455	1.499	2	1.563
Fuselage fineness	14	15.8	20	15.46
Engine parameter	4	3.30	5	3.36
Pressure limit	200	150.0	200	150.4
Number of passengers	220.3	253.3	192.8	253.4

Table 12.3

Hypersonic Cruise Vehicle Optimization
Using Four Search Methods [12F.7]

Method	Number of passengers	Function Evaluations*
Sectioning	253.1	70
Adaptive creeping	253.3	52
Steepest descent	252.0	150
Quadratic	253.3	220

*Measure of total computer time.

An interesting conclusion made by Hague and Glatt [12F.7], which supports their multi-method approach,is "... the more 'sophisticated' searches typified by the steepest descent ... the second order searches, converge less rapidly and reliably than the straightforward creeping search wherever comparisons are made. This behavior is in contradiction to several of the numerical experiments performed in [12D.1]. This emphasizes a point long known to practicing optimization specialists, that no single universal search technique is best suited to solution of all conceivable optimization problems. Conversely, given a particular search algorithm, one can almost always define a surface on which the particular search will appear superior to other searches".

Other applications of AESOP are reported by Stepniewski and Kalmbach of Boeing's Vertol Division [12F.9]. One optimization problem investigated was the maximization of the efficiency of a hovering rotor, the design variables being parameters describing the twist and chordlength distributions along the blade span. This problem was set up to provide a man-computer interface (with an IBM 2250 graphic display scope), in order that an engineer could monitor and control the optimization. The results indicated a 46% improvement in the 'static figure of merit' of the rotor in approximately 15 minutes elapsed time (computer time was less).

Other applications reported in [12F.9] are: prop/rotor design for tilt-wing and tilt-rotor aircraft; helciopter rotor design; and the design of an STOL transport.

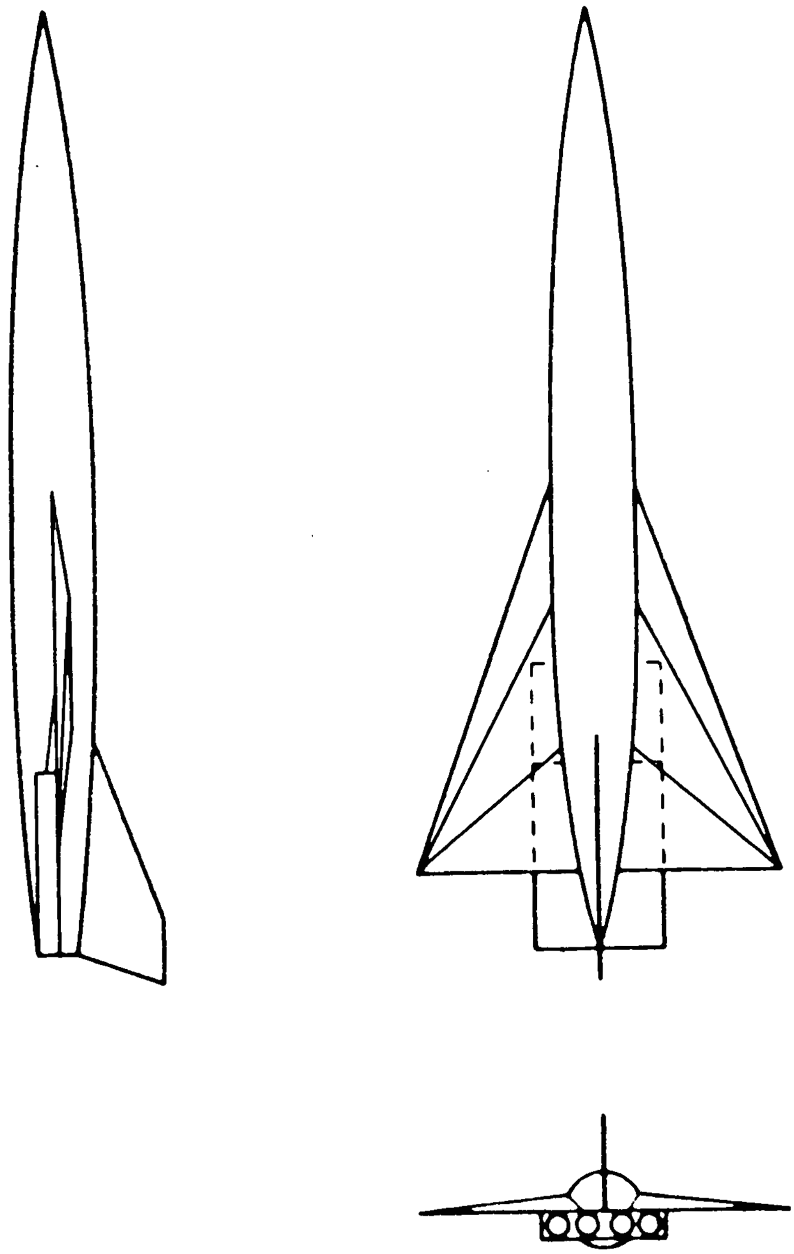


Fig.12.6 NASA Hypersonic Cruise Vehicle (Ref. [12F.7])

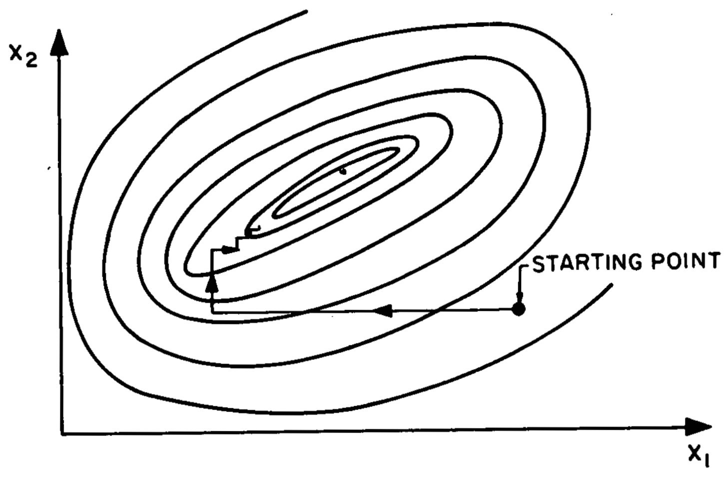


Fig.12.7 Search by "Sectioning" Note Premature Stopping due to Presence of a Ridge

Some results of using various search methods are presented for the STOL transport problem, which may be stated as follows: maximize the payload weight for a fixed takeoff gross weight (100 000 lb), fixed fuselage and empennage, fixed takeoff distance (1000 ft over a 50 ft obstacle), and fixed cruise speed (400 kt at 30000 ft). The design variables are

α_1 wing loading (lb/ft^2)

α_2 aspect ratio (upper limit of 12 from aeroelastic considerations)

α_3 wing thickness ratio

α_4 lift engine thrust/gross weight

α_5 lift engine thrust angle

α_6 added lift coefficient ΔC_L due to high-lift devices.

Another study added 30 flight path angles (at 1000 ft altitude increments) as 'design variables' in order also to optimize the climb.

It is obvious from the formulation above that wing design is one of the most critical factors. Wing weight was estimated by an empirical formula based on the four wing parameters; for example, wing weight was predicted to increase 22% for each unit increment in ΔC_L.

Equivalent to maximizing the payload is minimizing the sum of the wing weight, engine weights and mission fuel weight; this sum is defined as y. The results of the various search techniques are reproduced from [12F.9] in Figs.12.8, 12.9 and 12.10. For this particular study single search methods, such as steepest descent, quadratic and sectioning, did not perform well. Combinations of search methods, particularly those employing random ray, showed much better results. The fastest combination tried is random point + random ray + quadratic + pattern. The philosophy behind the combination strategy is based on an observation of Wilde [12B.2], who compared optimization search with the three phases of chess: opening game, middle game and end game. Each requires a different strategy.

The results of the [12F.9] study lead Stepniewski and Kalmbach to state that "... it appears that gradient procedures like steepest descent, quadratic and Davidon are of limited help with engineering problems of the class being investigated here". This experience correlates with the conclusions of Ref. [12F.7].

Other investigators have reported satisfactory results using a combination of just random ray and pattern from AESOP [12F.10], [12F.11]. In fact they indicate they have standardized on this combination for the present. It should be added that AESOP is quite a large computer program, almost filling core memory (IBM 7094) in some applications [12D.1], [12F.10].

The somewhat surprising success evidenced by strategies employing the random ray search, one of the most unsophisticated of methods, may arise from its very simplicity. It has no innate bias and is not confined to predetermined directions, nor can it be fooled by inaccurate gradient calculations or noisy data. The method is also of interest for controlling noisy dynamic systems, Rastrigin [12D.4].

12.4.2 Other Operational Experience

The acceleration move (called 'pattern' in AESOP) has proven useful ([12B.1], p.309 and [12B.2], p.305) in overcoming the 'zig-zagging' tendency of certain methods, such as steepest descent, particularly on a ridge. In addition to its ability to overcome a ridge, it should be noted that the acceleration move does not require evaluation of partial derivatives.

Other accelerated searches are the pattern search of Hooke and Jeeves, [12D.9] and [12B.2] p.307, which is not the same as AESOP's pattern search, and also Rosenbrock's 'method of rotating coordinates', [12B.2], p.312. Rosenbrock's method has shown good results in overcoming curved ridges. W. B. Herbst of McDonnell-Douglas Aircraft Corporation reports success with this method when applied to the FX tactical fighter development and also to the follow-on F-15 design [12F.5]. For the latter application a specific program called CASE ('computerized systems engineering') was written.

The method of Davidon [12D.7], as extended by Fletcher and Powell [12D.8], has proven to be very powerful in solving problems with either analytic partials or smooth data from which numerical partials may be estimated [12B.1], p.320 or [12B.2], p.331. Jameson of Grumman Aerospace Corporation [12F.13] has described effective applications of this method. That Davidon failed to provide good convergence in the AESOP results reported in the previous section is probably due to noise in the mathematical model. Since Davidon obtains an estimate of the second-order derivatives from the changes in the first-order derivatives, it is obvious that any noise in the data would create spurious results. Kelley and Myers [12D.10] have discovered that this method is also sensitive to roundoff errors in the computer. They suggest either the use of double-precision arithmetic or a modified technique in which the procedure is restarted every n moves. Davidon has demonstrated success on a helical ridge and on functions with up to 100 variables, [12B.1], p.349. Moreover, Fletcher and Powell [12D.8] have demonstrated that the number of moves increases only linearly with the number of variables when this method is employed on a quadratic function.

Experience with various one-dimensional search methods has not been fully reported. The Fibonacci search has been called 'best' at various times, and some have taken this as being literally true [12F.9]. However, the only claim made for this method is that it is the minimax interval elimination strategy, that is, it guarantees the best reduction for the worst possible outcomes. Moreover, it assumes very little about the objective function except that it is unimodal, i.e. has one optimum in the interval.

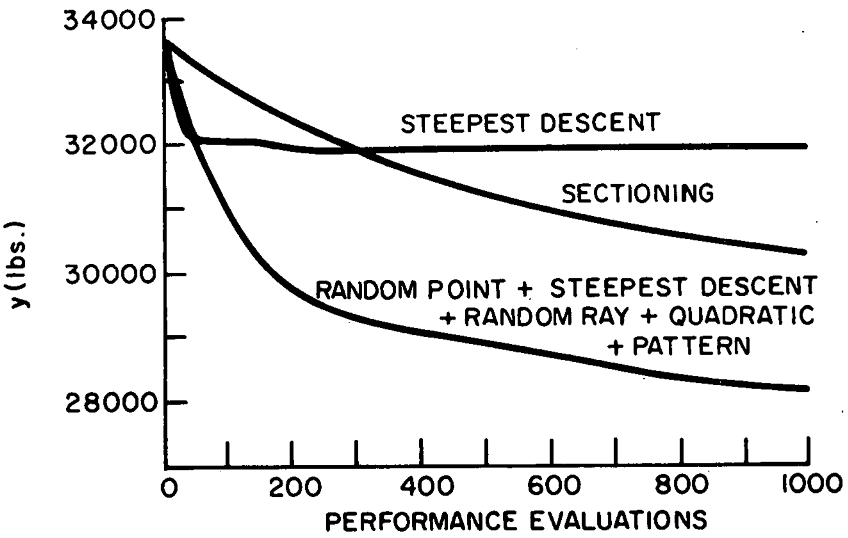


Fig. 12.8 STOL Optimization Results

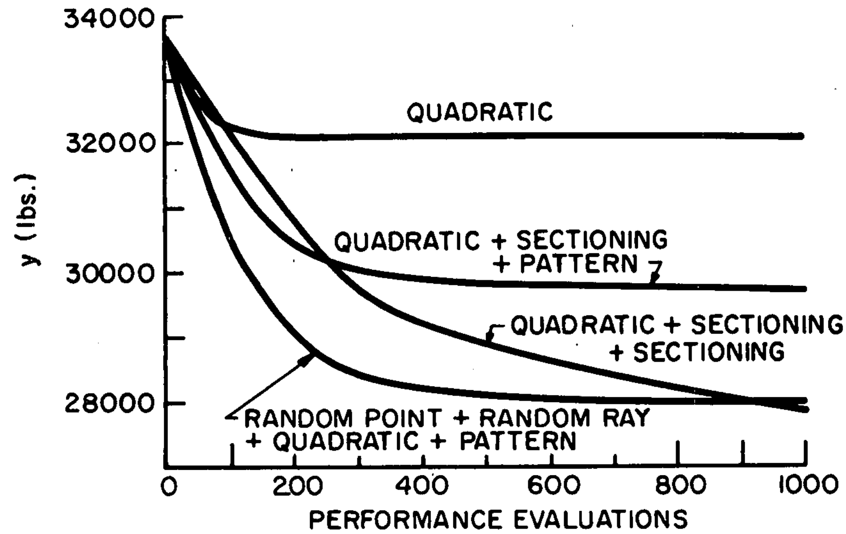


Fig. 12.9 STOL Optimization Results

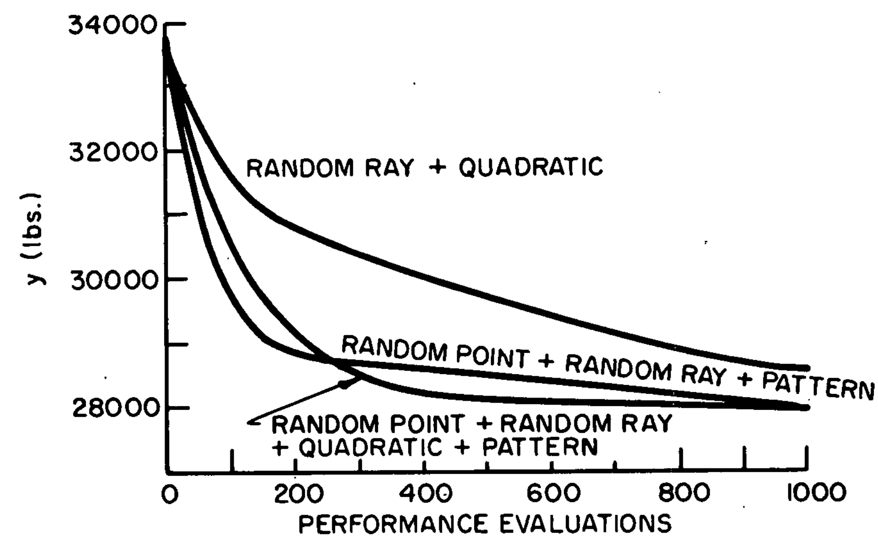


Fig. 12.10 STOL Optimization Results

For fairly well-behaved functions it is obviously possible to improve over this method; in fact, the quadratic curve fit finds the optimum in three steps for a quadratic function. This difference is not trivial, since many search methods spend the bulk of their time in the one-dimensional mode.

When using the curve-fit method for a one-dimensional search, it is generally desirable to select a low-degree polynomial, such as a quadratic or a cubic. This method works best when the interval of points (to which the curve is fitted) contains the optimum. To this end, the 'step until failure' method with $K = 10$ can be used to find the proper scale before the curve-fit method is applied.

12.5 Man-Computer Interactive Design

There is a semantic problem in titling this section. The area discussed here is commonly called 'computer aided design' (CAD) or 'computer graphics'. These terms do not, in the authors' opinion, sufficiently delineate the concept which is the interaction or conversation between man and computer allowed by new techniques of time-sharing. The first-generation of digital computers were so slow and difficult to program that a more-or-less continuous man-computer interaction was required. The second-generation computers emphasized the 'batch-process'. The disadvantage of this approach has been made frustratingly clear to most engineers and programmers: control of the program is lost until the run is returned. The time that this takes is called the 'turn-around time', and it may be of the order of a day. However efficient this appears to be from the point of view of computer operations, it is often very wasteful of the engineer's time. The interface between the man and the computer shows up the basic differences in their capabilities: man is much slower and more prone to mistakes than is the computer - but man does have more adaptability and judgment.

Third-generation computers have the capability of 'time-sharing', which allows many users to be simultaneously serviced by a single data-processing center. This feature greatly improves the man-computer interface; each operates at his (its) best speed. It also enables the engineer to retain control of the program. These advantages do not come without added cost: quoting Narahara [12G.1], "If engineers were as rigorous as is often claimed, batch processing would be good enough, and the large investment in interactive systems wouldn't be necessary or even worthwhile. But batch processing really requires the engineer to know what he wants (beforehand).... The program must be thought out to the point that the problem-solving method can be specified in intricate detail." This is particularly significant for aircraft design, which often includes considerations that are difficult to state analytically.

A terminal for interactive design may be simply a typewriter- or teletype-like terminal which allows the engineer to converse with the computer. A more versatile and commonly used device, however, involves a graphic display, such as an IBM 2250 [12G.2], which can display graphs, three-view drawings, numerical results, etc. This device also has a light-pen and a keyboard, which provide in combination for the engineer to control the operation of his program. By contrast with batch processing, this arrangement gives instant turn-around time. Thus more efficient use is made of the engineer's time.

Since time is generally critical in aircraft design development, interactive programming can help compress the design cycle - if the programming has been prepared beforehand. Chasen of Lockheed-Georgia [12G.3] suggests that the design process consists of an interactive sequence of events and decisions involving many specialists, say, individuals A, B and C. A design sequence might be ADBACADCDCBD..., where D is the computer. It is obvious that this sequence would take many days in a batch-processing usage of the computer. If these three specialists can be brought together in a conversation with the computer, however, the design cycle might be compressed greatly.

Rapid growth is predicted in interactive design activities by experts in this field, [12F.2], [12F.4], [12F.6], [12F.14] and [12F.15]. Among existing applications of interactive design are:

(1) Aircraft preliminary design, especially of large cargo types, Lockheed-Georgia, [12G.6].

(2) Wing/body aerodynamic design, Lockheed-Georgia, [12G.7].

(3) Aircraft sizing, Douglas Aircraft Company, [12G.8].

(4) Helicopter vibration analysis, Boeing Vertol Division, [12G.9].

(5) Integrated wing design (ORACLE), The Boeing Company, [12G.10].

(6) Aircraft and missile preliminary design, The Boeing Company, [12G.10].

Many aerospace companies have developed capabilities for drafting, parts-design, and fabrication by numerically-controlled tools using computer graphics [12G.3]. A number of other applications in many fields have been found for computer graphics, including integrated circuit design, automobile design, animated dynamics, curve fitting, etc. [12G.3]. A recent instance of aeronautical interest involves aircraft flight-path optimization [12G.11].

List of References

12A Parametric Analysis and Aircraft Design

Ref.

12A.1 Following three references refer to General Dynamics' computer program SYNAC, (SYNthesis of AirCraft). This is a comprehensive preliminary aircraft design program in the parametric analysis vein.

12A.1a V. A. Lee and H. G. Ball, "Parametric Aircraft Synthesis and Performance Analysis", AIAA Paper 66-795, October 1966

12A.1b V. A. Lee, et al., "Computerized Aircraft Synthesis", *J. Aircraft,* Vol.4, No.5, September-October 1967

12A.1c H. R. Anderson and J. D. McLeod, "Computerized Aircraft Design Studies", Symposium on Aircraft Systems Design Synthesis, California Institute of Technology, December 1968. (Reports that SYNAC, in a parametric analysis mode, ran over 1000 designs in 16 hours of IBM 360/65 computer time. Results showed a 'considerable' improvement over original configurations proposed by preliminary designers.)

12A.2 R. A. Davis, "Computer Application to the Airplane Design Selection Process", (Thumbprint), Boeing Company Document D6-24222TN, September 1969

12A.3 F. D. Orazio, "From Technology to Systems in Military Aircraft", *Astronautics and Aeronautics,* July 1969, p.48. (This and the following article give the flavor of present aircraft design. Of course, a large number of other articles of this type could be given.)

12A.4 E. R. Schuberth and L. Celniker, "Synthesizing Aircraft Design", *Space/Aeronautics,* April 1969, p.60

12B General Books on Optimization

12B.1 D. A. Pierre, *Optimization Theory with Applications,* John Wiley and Sons, New York, 1969. (Covers calculus of variations, linear programming, parameter optimization, dynamic programming and the maximum principle; with particular applications in the field of electrical engineering.)

12B.2 D. J. Wilde and C. S. Beightler, *Foundations of Optimization,* Prentice-Hall, Englewood Cliffs, New Jersey, 1967. (Presents a unified view of various optimization approaches; content similar to previous reference with examples typically in chemical engineering.)

12B.3 D. J. Wilde, *Optimal Seeking Methods,* Prentice-Hall, Englewood Cliffs, New Jersey, 1964. (An earlier book by Wilde, of narrower scope than Ref. [12B.2].)

12B.4 A. Lavi and T. P. Vogl, eds., *Recent Advances in Optimization Techniques,* John Wiley and Sons, New York, 1966. (A collection of articles on various aspects of optimization, many of which deal with design optimization of optical and electrical systems.)

12B.5 A. E. Bryson, Jr. and Yu-Chi Ho, *Applied Optimal Control,* Blaisdell Publishing Company, Waltham, Massachusetts, 1969. ("Emphasis is on determining the best way to control complicated dynamic systems." An introductory chapter is on parameter optimization.)

12B.6 G. Hadley, *Non-linear and Dynamic Programming,* Addison-Wesley, Reading, Massachusetts, 1964. ("Presents a rather detailed development of the theory and computational techniques.... Topics include: approximation methods, stochastic programming, integer programming and gradient methods.")

12C Indirect Methods of Optimization

12C.1 A. Miele, ed., *Theory of Optimum Aerodynamic Shapes,* Academic Press, New York, 1965. (Subtitled "Extremal Problems in the Aerodynamics of Supersonic, Hypersonic, and Free-Molecular Flows.")

12C.2 R. H. Liebeck and A. I. Ormsbee, "Optimization of Airfoils for Maximum Lift", AIAA Paper 69-739, July 1969. (Derives subsonic airfoil profiles using calculus of variations.)

12C.3 R. J. Duffin, E. L. Petersen and C. Zener, Geometric Programming, John Wiley and Sons, New York, 1967

12C.4 Z. M. v. Krzywoblocki and W. Z. Stepniewski, "Application of Optimization Techniques to the Design and Operation of V/STOL Aircraft", Annals of the New York Academy of Sciences, New York, Vol.154, Art.2, p.982, November 1968

12D Direct Methods of Optimization

12D.1 D. S. Hague and C. R. Glatt, "An Introduction to Multivariable Search Techniques for Parameter Optimization (and Program AESOP)", NASA CR-73200, April 1968. (A description of the direct search methods used in Boeing's optimization program AESOP.)

12D.2 Ref.[12B.1], Chapter Six. (Entitled "Search Techniques and Non-linear Programming", this Chapter includes a rather extensive discussion of methods of direct search.

List of References (Contd.)

Ref.

12D.3 Ref. [12B.2], Chapters Six and Seven. (Entitled "Direct Elimination", i.e. one-dimensional interval elimination methods, and "Direct Climbing", these Chapters cover methods of direct search.)

12D.4 L. A. Rastrigin, "Random Search in Optimization Problems for Multiparameter Systems", Clearinghouse AD 669 542, Springfield, Virginia. Translated from the Russian. (Originally published by Isdatel'stvo 'Zinatne', Riga, 1965.) (This English translation may be ordered for $3 from the Clearinghouse. 252 pages.)

12D.5 S. M. Movshovich, "Random Search and the Gradient Method in Optimization Problems", pp.60-72 of Technical Cybernetics, No.6, 1966, USSR. (Available in English translation from the Clearinghouse, TT: 67-30538, $3.)

12D.6 G. E. Forsythe and T. S. Motzkin, "Acceleration of the Optimum Gradient Method - Preliminary Report (Abstract)", Bulletin of the American Mathematical Society, p.304, July 1951

12D.7 W. C. Davidon, "Variable Metric Method for Minimization", AEC R and D Rep. Anl-5990, December 1959

12D.8 R. Fletcher and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization", *Computer Journal*, Vol.6, No.2, pp.163-168, 1963. (This work further developed and interpreted Davidon's Method.)

12D.9 R. Hooke and T. A. Jeeves, "Direct Search Solution of Numerical and Statistical Problems", *Journal of the Association for Computing Machinery*, Vol.8, pp.212-229, April 1961. (Introduces an accelerated ridge-following method called Pattern Search; not the same as the pattern method used in AESOP.)

12D.10 H. J. Kelly and G. E. Meyers, "Conjugate Direction Methods for Parameter Optimization", 18th International Astronautical Federation, Belgrade, Yugoslavia, September 1967

12D.11 H. J. Kelly, et al., "An Accelerated Gradient Method for Parameter Optimization with Nonlinear Constraints", American Astronautical Society Preprint 66-118, July 1966

12D.12 B. V. Shah, et al., "Some Algorithms for Minimizing a Function of Several Variables", *Journal of the S.I.A.M.*, Vol.12, No.1, pp.74-92, March 1964

12E Some Automated Aircraft Design Programs

12E.1 D. S. Hague and C. R. Glatt, "A Guide to the Automated Engineering and Scientific Optimization Program", NASA CR-73201, June 1968. (Boeing's AESOP is a search driver which has been applied to a number of aircraft design problems, including a hypersonic cruise vehicle, Ref. [12F.7].)

12E.2 J. Czinczenheim and M. Pottier, "Integrated Airplane Design Optimization", Breguet Aviation (France), (Undated but about 1967). (Develops a steepest descent formulation in which constraints are adjoined to the objective function with Lagrange multipliers. When applied to a 275 passenger airbus, this method gave a 9% improvement over the traditional hand method. This application took three minutes per iteration and about one hour for convergence on an IBM 7094 II. Contains a good bibliography.)

12E.3 R. J. White, "A Digital Program Useful for Airplane Integration and Design (AID)", Boeing Company Document D6-23592, January 1969. (Coupled with AESOP, this forms a preliminary design optimization program.)

12E.4 W. B. Herbst, McDonnell Douglas Corporation, St. Louis, Missouri, October 17, 1969, private communication. (Describes a computer program called CASE which uses Rosenbrock's Method of direct search to optimize the design "during the definition phase of the F-15 Tactical Fighter development".)

12E.5 H. M. Drake, Ames Research Center (NASA), Moffett Field, California, October 17, 1969, private communication. (Describes two aircraft synthesis programs which may be coupled with AESOP. Each program contains aerodynamics, propulsion, performance and weights sections.)

12E.6 W. Z. Stepniewski and C. F. Kalmbach, Jr., "Multivariable Search and its Application to Aircraft Design Optimization", Boeing Vertol Division, September 1969. (Describes the application of AESOP to V/STOL design problems.)

12F Operational Experience in Aircraft Optimization

12F.1 R. C. Hornburg, Douglas Aircraft Company, Long Beach, California, November 3, 1969, private communication

12F.2 I. G. Hedrick, Grumman Aerospace Corporation, Bethpage, New York, October 28, 1969, private communication

12F.3 V. A. Lee, General Dynamics, Fort Worth, Texas October 16, 1969, private communication

List of References (Contd.)

Ref.

12F.4 J. A. Thelander, Douglas Aircraft Company, Long Beach, California, November 3, 1969, private communication

12F.5 W. B. Herbst, McDonnell Douglas Corporation, St. Louis, Missouri, October 17, 1969, private communication

12F.6 H. M. Drake, Ames Research Center (NASA), Moffett Field, California, October 17, 1969, private communication

12F.7 D. S. Hague and C. R. Glatt, "Application of Multivariable Search Techniques to the Optimal Design of a Hypersonic Cruise Vehicle", NASA CR-73202, April 1968. (Applies Boeing's AESOP to the design of a hypersonic cruise vehicle. Finds a 15% improvement over NASA supplied nominal obtained by conventional hand methods.)

12F.8 D. S. Hague, "Application of Multivariable Search Techniques to the Shaping of Minimum Total Heat Reentry Bodies at Hypersonic Velocity", NASA CR-73203, April 1968

12F.9 W. Z. Stepniewski and C. F. Kalmbach, Jr., "Multivariable Search and its Application to Aircraft Design Optimization", Boeing Vertol Division, September 1969

12F.10 R. H. Petersen, Ames Research Center (NASA), Moffett Field, California, private communication

12F.11 R. J. White, Boeing Company, Seattle, Washington, January 14, 1970, private communication

12F.12 D. J. Wilde, Stanford University, Stanford, California, private communication

12F.13 A. Jameson, Grumman Aerospace Corporation, Bethpage, New York, October 29, 1969 private communication

12F.14 R. Q. Boyles, Lockheed-Georgia Company, Marietta, Georgia, private communication

12F.15 F. D. Orazio, Sr., Wright-Patterson Air Force Base, Ohio, December 16, 1969, private communication

12G Man-Computer Interactive Design

12G.1 R. M. Narahara, "Computer-Aided Design", *Space/Aeronautics*, December 1969

12G.2 Many authors, "Interactive Graphics in Data Processing", IBM Systems Journal, Vol.7, Nos.3 and 4, 1968

12G.3 S. H. Chasen, "The Role of Man-Computer Graphics in the Design Process", AIAA Professional Study Series Volume, 1969

12G.4 S. H. Chasen and B. Herzog, "Applied Computer Aided Design and Interactive Graphics", AIAA Professional Study Series Volume, 1969

12G.5 B. Herzog, "Lectures in Computer Aided Design", AIAA Professional Study Series Volume, 1969

12G.6 R. Q. Boyles, "Aircraft Design Augmented by a Man-Computer Graphic System", *Journal of Aircraft*, Vol.5, No.5, September-October 1968

12G.7 J. A. Bennett, W. A. Stevens and R. C. Davis, "A Computer-Aided Wing/Body Aerodynamic Design Concept for Subsonic Vehicles of the 1970-1980 Period", AIAA Paper No.69-1130, October 1969

12G.8 G. D. Buell, Jr., "Aircraft Sizing Using Computer Graphics", Douglas Aircraft Group IRAD T.R. DAC 67140, July 1968

12G.9 J. J. Sciarra, "Vibration Analysis in 3D with Computer Graphics", *Sound and Vibration*, Vol.4, No.1, January 1970

12G.10 Many authors, "Computer Aided Design Workshop", Boeing Company, Attachment to M-7130-025, September 1969

12G.11 J. A. Thelander, "Variation Analysis - Applications and Solution Techniques Related to Aircraft Optimization Problems", AIAA Paper No.67-557, August 1967

12H Other References

12H.1 "An Indexed Bibliography of Optimization Literature Related to Engineering Design", Vol.3, Appendix 1 of "Advanced Decoy Technology Program, Final Report (U)", Avco Missiles, Space and Electronics Group, Wilmington, Massachusetts, February 1968, (636 entries).

12H.2 A. Leon, "A Classified Bibliography on Optimization", *Recent Advances in Optimization Techniques*, pp.599-649 of Ref. [12B.4]. (377 entries)

12H.3 G. V. Reklaitis and D. J. Wilde, "Necessary Conditions for a Local Optimum without Prior Constraint Qualification", to be published in the Journal Of Optimization Theory

APPENDIX A

SELECTIVE BIBLIOGRAPHY

Appendix A

Selective Bibliography

(Each Section is listed in reverse chronological order.)

Engineering Applications

Books

1 Fox, R. L., *An Introduction to Optimization Methods for Engineers*, to be published by Addison-Wesley, Reading, Massachusetts, 1970

2 Cohn, M. Z., ed., *An Introduction to Structural Optimization*, University of Waterloo, Waterloo, Canada, 1969

3 Au, T. and Stelson, T. E., *Introduction to Systems Engineering-Deterministic Models*, 1st ed., Addison-Wesley, Reading, Massachusetts, 1969

4 Cox, H. L., *The Design of Structures of Least Weight*, Pergamon, Oxford, 1965

5 Gerard, G., *Minimum Weight Analysis of Compressive Structures*, 1st ed., New York University Press, New York, 1956

6 Shanley, F. R., *Weight-Strength Analysis of Aircraft Structures*, 1st ed., McGraw-Hill, New York, 1952

Reviews

1 Barnett, R. L., "Survey of Optimum Structural Design", *Experimental Mechanics*, Vol.6, No.12, December 1966, pp.19A-26A

2 Gerard, G., "Optimum Structural Design Concepts for Aerospace Vehicles: Bibliography and Assessment", USAF, AFFDL TR-66-188, December 1966

3 Kowalik, J., "Non-linear Programming Procedures and Design Optimization", Mathematics and Computing Machinery Series NR 13, 1966, Acta Polytechnica Scandinavica, Trondheim, Norway

4 Gerard, G., "Optimum Structural Design Concepts for Aerospace Vehicles: Bibliography and Assessment", USAF, AFFDL TR-65-9, June 1965

5 Wasiutynski, Z. and Brandt, A., "The Present State of Knowledge in the Field of Optimum Design of Structures", *Applied Mechanics Review*, Vol.16, No.5, May 1963, pp.341-350

6 Micks, W. R., "Bibliography of Literature on Optimum Design of Structures and Related Topics", RM2304, December 15, 1958, The Rand Corporation, Santa Monica, California

7 Hemp, W. S., "Theory of Structural Design", Report No.115, August 1958, The College of Aeronautics, Cranfield, England

Papers

1 Felton, L. P. and Hofmeister, L. D., "Synthesis of Waffle Plates with Multiple Rib Sizes", *AIAA Journal*, Vol.7, No.12, December 1969, pp.2193-2199

2 Shinozuka, M. and Yang, J. N., "Optimum Structural Design Based on Reliability and Proof-Load Test", *Annals of Assurance Sciences, Proceedings of Reliability and Maintainability Conference*, Vol.8, July 1969, pp.375-391

3 Zarghamee, M. S., "Minimum Weight Design of Enclosed Antennas", *Journal of the Structural Division, ASCE*, Vol.95, No.ST6, June 1969, pp.1139-1152

4 McIntosh, S. C., Weisshaar, T. A. and Ashley, H., "Progress in Aeroelastic Optimization - Analytical vs. Numerical Approaches", AIAA Structural Dynamics and Aeroelasticity Specialist Conference, New Orleans, La., April 16-17, 1969

5 Rubin, C. P., "Dynamics Optimization of Complex Structures", AIAA Structural Dynamics and Aeroelasticity Specialist Conference, New Orleans, La., April 16-17, 1969, pp.9-14

6 Dayaratnam, P. and Patnaik, S., "Feasibility of Full Stress Design", *AIAA Journal*, Vol.7, No.4, April 1969, pp.773-774

7 Romstad, K. M. and Wang, C. K., "Optimum Design of Framed Structures", *Journal of the Structural Division, ASCE*, Vol.94, No.ST12, December 1968, pp.2817-2845

8 Switzky, H., "Designing for Minimum Flexibility or Weight", *Journal of Spacecraft and Rockets*, Vol.5, No.12, December 1968, pp.1473-1476

9 Fox, R. L. and Kapoor, M. P., "A Minimization Method for the Solution of the Eigenproblem Arising in Structural Dynamics", Proc. of the Second Conference on Matrix Methods in Structural Mechanics, WPAFB, Ohio, October 1968, AFFDL-TR-68-150, December 1969, pp.271-306

Appendix A (Contd.)

10 Marcal, P. V. and Gellatly, R. A., "Application of the Created Response Surface Technique to Structural Optimization", Second Conference on Matrix Methods in Structural Mechanics, WPAFB, Ohio, October 1968, AFFDL-TR-68-150, December 1969, pp.83-110

11 Fox, R. L. and Stanton, E., "Developments in Structural Analysis by Direct Energy Minimization", *AIAA Journal*, Vol.6, No.6, June 1968, pp.1036-1042

12 McIntosh, S. C. and Eastep, F. E., "Design of Minimum Mass Structures with Specified Stiffness Properties", *AIAA Journal*, Vol.6, No.5, May 1968, pp.962-964

13 Toakley, A. R., "Optimum Design Using Available Section", *Journal of the Structural Division, ASCE*, Vol.94, No.ST5, May 1968, pp.1219-1241

14 Luik, R. and Melosh, R. J., "An Allocation Procedure for Structural Design", Preprint No.68-329, AIAA/ASME 9th Structures, Structural Dynamics and Materials Conference, Palm Springs, Calif., April 1968

15 Pope, G. G., "The Design of Optimum Structures of Specified Basic Configuration", *International Journal of Mechan. Sci.*, Vol.10, No.4, April 1968, pp.251-263

16 Prager, W., and Shield, R. T., "Optimal Design of Multipurpose Structures", *International Journal of Solids and Structures*, Vol.4, No.4, April 1968, pp.469-475

17 Zarghamee, M. S., "Optimum Frequency of Structures", *AIAA Journal*, Vol.6, No.4, April 1968, pp.749-750

18 Prager, W. and Taylor, J. E., "Problems in Optimal Structural Design", *Journal of Applied Mechanics*, Vol.35, No.1, March 1968, pp.102-106

19 Kicher, T. P., "Structural Synthesis of Integrally Stiffened Cylinders", *Journal of Spacecraft and Rockets*, Vol.5, No.1, January 1968, pp.62-67

20 Moe, J. and Lund, S., "Cost and Weight Minimization of Structures with Special Emphasis on Longitudinal Strength Members of Tankers", Transactions of the Royal Institution of Naval Architects, Vol.110, No.1, January 1968, pp.43-70

21 Turner, M. J., "Design of Minimum Mass Structures with Specified Natural Frequencies", *AIAA Journal*, Vol.5, No.3, March 1967, pp.406-412

22 Goble, G. G. and DeSantis, P. V., "Optimum Design of Mixed Steel Composite Girders", *Journal of the Structural Division, ASCE*, Vol.92, No.ST6, December 1966, pp.25-43

23 Kicher, T. P., "Optimum Design - Minimum Weight Versus Fully Stress", *Journal of the Structural Division, ASCE*, Vol.92, No.ST6, December 1966, pp.265-279

24 Young, J. W., Jr. and Christiansen, H. N., "Synthesis of a Space Truss Based on Dynamic Criteria", *Journal of the Structural Division, ASCE*, Vol.92, No.ST6, December 1966, pp.425-442

25 Ghista, D. N., "Fully-Stress Design for Alternative Loads", *Journal of the Structural Division, ASCE*, Vol.92, No.ST5, October 1966, pp.237-260

26 Ghista, D. N., "Optimum Frameworks Under Single Load System", *Journal of the Structural Division, ASCE*, Vol.92, No.ST5, October 1966, pp.261-286

27 Gellatly, R. A. and Gallagher, R. H., "A Procedure for Automated Minimum Weight Structural Design, Part I - Theoretical Basis, Part II - Applications", *Aeronautical Quarterly*, Vol.17, No.3, August 1966, pp.216-230 and No.4, November 1966, pp.332-342

28 Fox, R. L. and Schmit, L. A., "Advances in the Integrated Approach to Structural Synthesis", *Journal of Spacecraft and Rockets*, Vol.3, No.6, June 1966, pp.858-866

29 Razani, R. and Goble, G. G., "Optimum Design of Constant Depth Girders", *Journal of the Structural Division, ASCE*, Vol.92, No.ST2, April 1966, pp.253-281

30 Porter Goff, R. F. D., "Decision Theory and the Shape of Structures", *Journal of the Royal Aeronautical Society*, Vol.70, No.663, March 1966, pp.448-452

31 Razani, Reza, "The Behavior of Fully-Stressed Design of Structures and its Relationship to Minimum Weight Design", *AIAA Journal*, Vol.3, No.12, December 1965, pp.2262-2268

32 Schmit, L. A. and Fox, R. L., "An Integrated Approach to Structural Synthesis and Analysis", *AIAA Journal*, Vol.3, No.6, June 1965, pp.1104-1112

33 Schmit, L. A., "Comment on Completely Automatic Weight-Minimization Method for High-Speed Digital Computers", *Journal of Aircraft*, Vol.1, No.6, November/December 1964, pp.375-377

34 Best, G. C., "Completely Automatic Weight-Minimization Method for High-Speed Digital Computers", *Journal of Aircraft*, Vol.1 No.3, May/June 1964, pp.129-133

Appendix A (Contd.)

35 Switsky, H., "Minimum Weight Design with Structural Reliability", *AIAA Fifth Annual Structures and Materials Conference,* April 1964, pp.316-322

36 Dorn, W. S., Gomory, R. E., and Greenberg, H. J., "Automatic Design of Optimal Structures", *Journal de Mechanique,* Vol.3, No.1, March 1964, pp.25-52

37 Best, G. C., "A Method of Structural Weight Minimization Suitable for High Speed Digital Computers", *AIAA Journal,* Vol.1, No.2, February 1963, pp.478-479

38 Kalaba, R., "Design of Minimal-Weight Structures for Given Reliability and Cost", *Journal of the Aerospace Sciences,* Vol.29, No.3, March 1962, pp.355-356

39 Hilton, H. H. and Feigen, M., "Minimum Weight Analysis Based on Structural Reliability", *Journal of the Aerospace Sciences,* Vol.27, No.9, September 1960, pp.641-652

40 Schmidt, L. C., "Fully-Stressed Design of Elastic Redundant Trusses under Alternative Load Systems", *Australian Journal of Applied Science,* Vol.9, No.4, December 1958, pp.337-348

41 Heyman, J. and Prager, W., "Automatic Minimum Weight Design of Steel Frames", *Journal of the Franklin Institute,* Vol.266, No.5, November 1958, pp.339-364

42 Livesley, R. K., "The Automatic Design of Structural Frames", *Quarterly Journal of Mechanics and Applied Mathematics,* Vol.9, Pt.3, September 1956, pp.257-278

43 Sved, G., "The Minimum Weight of Certain Redundant Structures", *Australian Journal of Applied Science,* Vol.5, No.1, March 1954, pp.1-9

44 Foulkes, J. D., "Minimum Weight Design and the Theory of Plastic Collapse", *Quarterly of Applied Mathematics,* Vol.10, No.4, January 1953, pp.347-358

45 Heyman, J., "Plastic Design of Beams and Frames for Minimum Material Consumption", *Quarterly of Applied Mathematics,* Vol.8, No.4, January 1951, pp.373-381

Reports

1 Venkayya, V. B., Knot, N. S. and Reddy, V. S., "Energy Distribution in an Optimum Structural Design", USAF, AFFDL-TR-68-156, March 1969

2 Kapoor, M. P., "Automated Optimum Design of Structures Under Dynamic Response Restrictions", Thesis for Degree of Doctor of Philosophy, Thesis Advisor, R. L. Fox, Case Western Reserve University, January 1969

3 Morrow, II, W. M., and Schmit, L. A., "Structural Synthesis of a Stiffened Cylinder", NASA CR-1217, December 1968

4 Thornton, W. A. and Schmit, L. A., "The Structural Synthesis of an Ablating Thermostructural Panel, NASA CR-1215, December 1968

5 Moses, F. and Stevenson, J. D., "Reliability based Structural Design", SMSMD Report No.16, January 1968, Case Western Reserve University, Cleveland, Ohio

6 Tocher, J. L. and Karnes, R. N., "Automatic Design of Optimum Hole Reinforcement", No.D6-23359, May 21, 1968, The Boeing Company, Commercial Airplane Division, Renton, Washington

7 Melosh, R. J. and Luik, R., "Approximate Multiple Configuration Analysis and Allocation for Least Weight Structural Design", USAF, AFFDL-TR-67-29, April 1967

8 Moses, F., "Some Notes and Ideas on Mathematical Programming Methods for Structural Optimization", Meddelelse SKB II/M8, January 1967, Norges Tekniske Høgskole, Trondheim, Norway

9 Toakley, A. R., "The Optimum Elastic-Plastic Design of Rigid Jointed Sway Frames", Fourth Report, Study of Analytical and Design Procedures for Elastic and Elastic-Plastic Structures, 1967, Dept. of Civil Engineering, University of Manchester, England

10 Gellatly, R. A., "Development of Procedures for Large Scale Automated Minimum Weight Structural Design", USAF, AFFDL-TR-66-180, December 1966

11 Ghista, D. N., "Structural Optimization with Probability of Failure Constraints", NASA TN D-3777, December 1966

12 Kavlie, D., Kowalik, J. and Moe, J., "Structural Optimization by Means of Non-linear Programming", Meddelelse SKB II/M4, 1966, Norges Tekniske Høgskole, Trondheim, Norway

13 Toakley, A. R., "Studies in Minimum Weight Rigid Plastic Design with Particular Reference to Discrete Sections", Second Report, Study of Analytical and Design Procedures for Elastic and Elastic-Plastic Structures, 1966, Dept. of Civil Engineering, University of Manchester, England

14 Brown, D. M. and Ang, A. H. S., "A Non-linear Programming Approach to the Minimum Weight Elastic Design of Steel Structures", Structural Research Series No.298, October 1965, Civil Engineering Studies, University of Illinois, Urbana, Illinois

Appendix A (Contd.)

15 Cornell, C. A., Reinschmidt, K. F. and Brotchie, J. F., "Structural Optimization", Research Report R65-26, Part 2, September 1965, Dept. of Civil Engineering, Mass. Inst. of Tech., Cambridge, Mass.

16 Schmit, L. A. and Thornton, W. A., "Synthesis of an Airfoil at Supersonic Mach Number", NASA CR 144, January 1965

17 Gellatly, R. A., Gallagher, R. H. and Luberacki, W. A., "Development of a Procedure for Automated Synthesis of Minimum Weight Structures", USAF, FDL-TDR-64-141, October 1964

18 Schmit, L. A. and Kicher, T. P., "Structural Synthesis of Symmetric Waffle Plate", NASA TN D-1691, December 1962

Mathematical Methods

Books

1 Kowalik, J. and Osborne, M. R., *Methods for Unconstrained Optimization Problems,* 1st ed., American Elsevier, New York, 1968

2 Fiacco, A. and McCormick, G. P., *Non-linear Programming; Sequential Unconstrained Minimization Techniques,* 1st ed., Wiley, New York, 1968

3 Wilde, D. J. and Beightler, C. S., *Foundations of Optimization,* 1st ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1967

4 Lavi, A. and Vogl, T. P., eds., *Recent Advances in Optimization Techniques,* 1st ed., Wiley, New York, 1966

5 Hadley, G., *Non-linear and Dynamic Programming,* 1st ed., Addison-Wesley, Reading, Massachusetts, 1964

6 Dantzig, G. B., *Linear Programming and Extensions,* 1st ed., Princeton University Press, Princeton, New Jersey, 1963

7 Hadley, G., *Linear Programming,* 1st ed., Addison-Wesley, Reading, Massachusetts, 1962

8 Zoutendijk, G., *Methods of Feasible Directions,* 1st ed., Elsevier, Amsterdam, 1960

9 Bellman, R., *Dynamic Programming,* 1st ed., Princeton University Press, Princeton, New Jersey, 1957

Reviews

1 Zoutendijk, G., "Non-linear Programming: A Numerical Survey", *SIAM Journal on Control,* Vol.4, No.1, February 1966, pp.194-210

2 Fletcher, R., "Function Minimization without Evaluating Derivatives - A Review", *The Computer Journal,* Vol.8, No.1, April 1965, pp.33-41

3 Spang, H. A., "A Review of Minimization Techniques for Non-linear Functions", *SIAM Review,* Vol.4, No.4, October 1962, pp.343-365

4 Brooks, S. H., "A Comparison of Maximum Seeking Methods", *Operations Research,* Vol.7, No.4, July-August 1959, pp.430-457

Papers

1 Bard, Y., "On A Numerical Instability of Davidon-Like Methods", *Mathematics of Computation,* Vol.22, No.103, July 1968, pp.665-666

2 Zangwill, W. I., "Minimizing a Function without Calculating Derivatives", *The Computer Journal,* Vol.10, No.3, November 1967, pp.293-296

3 Broyden, C. G., "Quasi-Newton Methods and their Application to Function Minimization", *Mathematics of Computation,* Vol.21, No.99, July 1967, pp.368-381

4 Daniel, J. W., "Convergence of the Conjugate Gradient Method with Computationally Convenient Modifications", *Numerische Mathematik,* Vol.10, No.2, July 1967, pp.125-131

5 Daniel, J. W., "The Conjugate Gradient Method for Linear and Non-linear Operator Equations", *SIAM Journal on Numerical Analysis,* Vol.4, No.1, March 1967, pp.10-26

6 Stewart, III, G. W., "A Modification of Davidon's Minimization Method to Accept Difference Approximations of Derivatives", *Journal ACM,* Vol.14, No.1, January 1967, pp.72-83

7 Zangwill, W. I., "Non-linear Programming via Penalty Functions", *Management Science, Series A,* Vol.13, No.5, January 1967, pp.344-358

8 Bradbury, W. W. and Fletcher, R., "New Iterative Methods for Solution of the Eigenproblem", *Numerische Mathematik,* Vol.9, No.3, December 1966, pp.259-267

Appendix A (Contd.)

9 Wilde, D. J., "Objective Function Indistinguishability in Unimodal Optimization", *Recent Advances in Optimization Techniques,* Lavi, A. and Vogl, T., eds., Wiley, New York, 1966, pp.341-349

10 Broyden, C. G., "A Class of Methods for Solving non-linear Simultaneous Equations", *Mathematics of Computation,* Vol.19, No.92, October 1965, pp.577-593

11 Bauer, F. L., "Optimally Scaled Matrices", *Numerische Mathematik,* Vol.5, No.1, March 1963 pp.73-87

12 Mugele, R. A., "A Non-linear Digital Optimizing Program for Process Control Systems", *Proceedings of Spring Joint Computer Conference,* National Press, Palo Alto, Calif., Vol.21, 1962, pp.15-31

13 Powell, M. J. D., "An Iterative Method for Finding Stationary Values of a Function of Several Variables", *The Computer Journal,* Vol.5, No.2, July 1962, pp.147-151

14 Hooke, R. and Jeeves, T. A., "Direct Search Solution of Numerical and Statistical Problems", *Journal Assoc. Comp. Mach.,* Vol.8, 1961, pp.212-229

15 Rosen, J. B., "The Gradient Projection Method for Non-linear Programming, Part II, Non-linear Constraints", *Journal SIAM,* Vol.9, No.4, December 1961, pp.514-532

16 Rosenbrock, H. H., "An Automatic Method for Finding the Greatest or Least Value of a Function", *The Computer Journal,* Vol.3, No.3, October 1960, pp.175-184

17 Rosen, J. B., "The Gradient Projection Method for Non-linear Programming, Part I, Linear Constraints", *Journal SIAM,* Vol.8, No.1, March 1960, pp.181-217

18 Brooks, S. H., "A Discussion of Random Methods for Seeking Maxima", *Operations Research,* Vol.6, No.2, April 1958, pp.244-251

19 Hestenes, M. R., "The Conjugate-Gradient Method for Solving Linear Systems", *Proceedings of Symposia in Applied Mathematics,* McGraw-Hill, New York, Vol. VI, 1956, pp.83-102

20 Crocket, J. B. and Chernoff, Herman, "Gradient Methods of Maximization", *Pacific Journal of Mathematics,* Vol.5, 1955, pp.33-50

21 Hestenes, M. R. and Stiefel, E., "Methods of Conjugate Gradients for Solving Linear Systems", *Journal Res. Natl. Bureau Standards,* Vol.49, No.6, December 1952, pp.409-436

22 Curry, H. B., "The Method of Steepest Descent for Non-linear Minimization Problems", *Quarterly of Applied Mathematics,* Vol.2, No.3, October 1944, pp.258-261

23 Levenberg, K., "A Method for the Solution of Certain Non-linear Problems in Least Squares", *Quarterly of Applied Mathematics,* Vol.2, No.2, July 1944, pp.164-168

Reports

1 Pearson, J. D., "On Variable Metric Methods of Minimization", RAC-TP-302, February 1968, Research Analysis Corporation, McLean, Virginia

2 Fiacco, A. V. and McCormick, G. P., "Programming under Non-linear Constraints by Unconstrained Minimization: A Primal-Dual Method", RAC-TP-96, September 1963, Research Analysis Corporation, Bethesda, Maryland

3 Gomory, R. E., "Large and Nonconvex Problems in Linear Programming", RC-765, 1962, IBM Research Report, Yorktown Heights, New York

4 Davidon, W. C., "Variable Metric Method for Minimization", ANL-5990 Rev., November 1959, Argonne National Laboratory, University of Chicago, Lemont, Illinois

AGARDograph No.149
North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development
STRUCTURAL DESIGN APPLICATIONS OF MATHEMATICAL PROGRAMMING TECHNIQUES
G.G.Pope and L.A.Schmit
Published February 1971
208 pages, incl. figs

The application of mathematical programming techniques in the optimum design of aerospace and similar structures is described starting from basic concepts and proceeding in a logical manner to derivations of the most powerful techniques currently available, and to descriptions of typical recent applications.

P.T.O.

624.07:681.3.06

AGARDograph No.149
North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development
STRUCTURAL DESIGN APPLICATIONS OF MATHEMATICAL PROGRAMMING TECHNIQUES
G.G.Pope and L.A.Schmit
Published February 1971
208 pages, incl. figs

The application of mathematical programming techniques in the optimum design of aerospace and similar structures is described starting from basic concepts and proceeding in a logical manner to derivations of the most powerful techniques currently available, and to descriptions of typical recent applications.

P.T.O.

624.07:681.3.06

AGARDograph No.149
North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development
STRUCTURAL DESIGN APPLICATIONS OF MATHEMATICAL PROGRAMMING TECHNIQUES
G.G.Pope and L.A.Schmit
Published February 1971
208 pages, incl. figs

The application of mathematical programming techniques in the optimum design of aerospace and similar structures is described starting from basic concepts and proceeding in a logical manner to derivations of the most powerful techniques currently available, and to descriptions of typical recent applications.

P.T.O.

624.07:681.3.06

AGARDograph No.149
North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development
STRUCTURAL DESIGN APPLICATIONS OF MATHEMATICAL PROGRAMMING TECHNIQUES
G.G.Pope and L.A.Schmit
Published February 1971
208 pages, incl. figs

The application of mathematical programming techniques in the optimum design of aerospace and similar structures is described starting from basic concepts and proceeding in a logical manner to derivations of the most powerful techniques currently available, and to descriptions of typical recent applications.

P.T.O.

624.07:681.3.06

Emphasis is placed on design for minimum weight and the relationship between classical and mathematical programming approaches is demonstrated in that context.

The text is divided into four main sections entitled: Fundamental Concepts and Literature Review, Algorithmic Tools, Simple applications, and Future trends and research needs. Aeroelastic constraints and reliability constraints are considered in the final section which concludes with a chapter on aircraft configuration design.

This AGARDograph was sponsored by the Structures and Material Panel of AGARD-NATO.

Emphasis is placed on design for minimum weight and the relationship between classical and mathematical programming approaches is demonstrated in that context.

The text is divided into four main sections entitled: Fundamental Concepts and Literature Review, Algorithmic Tools, Simple applications, and Future trends and research needs. Aeroelastic constraints and reliability constraints are considered in the final section which concludes with a chapter on aircraft configuration design.

This AGARDograph was sponsored by the Structures and Material Panel of AGARD-NATO.

Emphasis is placed on design for minimum weight and the relationship between classical and mathematical programming approaches is demonstrated in that context.

The text is divided into four main sections entitled: Fundamental Concepts and Literature Review, Algorithmic Tools, Simple applications, and Future trends and research needs. Aeroelastic constraints and reliability constraints are considered in the final section which concludes with a chapter on aircraft configuration design.

This AGARDograph was sponsored by the Structures and Material Panel of AGARD-NATO.

Emphasis is placed on design for minimum weight and the relationship between classical and mathematical programming approaches is demonstrated in that context.

The text is divided into four main sections entitled: Fundamental Concepts and Literature Review, Algorithmic Tools, Simple applications, and Future trends and research needs. Aeroelastic constraints and reliability constraints are considered in the final section which concludes with a chapter on aircraft configuration design.

This AGARDograph was sponsored by the Structures and Material Panel of AGARD-NATO.

AGARDograph No.149
North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development
STRUCTURAL DESIGN APPLICATIONS OF MATHEMATICAL PROGRAMMING TECHNIQUES
G.G.Pope and L.A.Schmit
Published February 1971
208 pages, incl. figs

The application of mathematical programming techniques in the optimum design of aerospace and similar structures is described starting from basic concepts and proceeding in a logical manner to derivations of the most powerful techniques currently available, and to descriptions of typical recent applications.

P.T.O.

624.07:681.3.06

AGARDograph No.149
North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development
STRUCTURAL DESIGN APPLICATIONS OF MATHEMATICAL PROGRAMMING TECHNIQUES
G.G.Pope and L.A.Schmit
Published February 1971
208 pages, incl. figs

The application of mathematical programming techniques in the optimum design of aerospace and similar structures is described starting from basic concepts and proceeding in a logical manner to derivations of the most powerful techniques currently available, and to descriptions of typical recent applications.

P.T.O.

624.07:681.3.06

AGARDograph No.149
North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development
STRUCTURAL DESIGN APPLICATIONS OF MATHEMATICAL PROGRAMMING TECHNIQUES
G.G.Pope and L.A.Schmit
Published February 1971
208 pages, incl. figs

The application of mathematical programming techniques in the optimum design of aerospace and similar structures is described starting from basic concepts and proceeding in a logical manner to derivations of the most powerful techniques currently available, and to descriptions of typical recent applications.

P.T.O.

624.07:681.3.06

AGARDograph No.149
North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development
STRUCTURAL DESIGN APPLICATIONS OF MATHEMATICAL PROGRAMMING TECHNIQUES
G.G.Pope and L.A.Schmit
Published February 1971
208 pages, incl. figs

The application of mathematical programming techniques in the optimum design of aerospace and similar structures is described starting from basic concepts and proceeding in a logical manner to derivations of the most powerful techniques currently available, and to descriptions of typical recent applications.

P.T.O.

624.07:681.3.06

Emphasis is placed on design for minimum weight and the relationship between classical and mathematical programming approaches is demonstrated in that context.

The text is divided into four main sections entitled: Fundamental Concepts and Literature Review, Algorithmic Tools, Simple applications, and Future trends and research needs. Aeroelastic constraints and reliability constraints are considered in the final section which concludes with a chapter on aircraft configuration design.

This AGARDograph was sponsored by the Structures and Material Panel of AGARD-NATO.

Emphasis is placed on design for minimum weight and the relationship between classical and mathematical programming approaches is demonstrated in that context.

The text is divided into four main sections entitled: Fundamental Concepts and Literature Review, Algorithmic Tools, Simple applications, and Future trends and research needs. Aeroelastic constraints and reliability constraints are considered in the final section which concludes with a chapter on aircraft configuration design.

This AGARDograph was sponsored by the Structures and Material Panel of AGARD-NATO.

Emphasis is placed on design for minimum weight and the relationship between classical and mathematical programming approaches is demonstrated in that context.

The text is divided into four main sections entitled: Fundamental Concepts and Literature Review, Algorithmic Tools, Simple applications, and Future trends and research needs. Aeroelastic constraints and reliability constraints are considered in the final section which concludes with a chapter on aircraft configuration design.

This AGARDograph was sponsored by the Structures and Material Panel of AGARD-NATO.

Emphasis is placed on design for minimum weight and the relationship between classical and mathematical programming approaches is demonstrated in that context.

The text is divided into four main sections entitled: Fundamental Concepts and Literature Review, Algorithmic Tools, Simple applications, and Future trends and research needs. Aeroelastic constraints and reliability constraints are considered in the final section which concludes with a chapter on aircraft configuration design.

This AGARDograph was sponsored by the Structures and Material Panel of AGARD-NATO.

NATIONAL DISTRIBUTION CENTRES FOR UNCLASSIFIED AGARD PUBLICATIONS

Unclassified AGARD publications are distributed to NATO Member Nations through the unclassified National Distribution Centres listed below

BELGIUM
General J.DELHAYE
Coordinateur AGARD – V.S.L.
Etat Major Forces Aériennes
Caserne Prince Baudouin
Place Dailly, Bruxelles 3

CANADA
Director of Scientific Information Services
Defence Research Board
Department of National Defence – 'A' Building
Ottawa, Ontario

DENMARK
Danish Defence Research Board
Østerbrogades Kaserne
Copenhagen Ø

FRANCE
O.N.E.R.A. *(Direction)*
29, Avenue de la Division Leclerc
92, Châtillon-sous-Bagneaux

GERMANY
Zentralstelle für Lüftfahrtdokumentation
und Information
Maria-Theresia Str. 21
8 München 27
Attn: Dr Ing. H.J.RAUTENBERG

GREECE
Hellenic Armed Forces Command
D Branch, Athens

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG
Obtainable through BELGIUM

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
Attn: Mr A.H.GEUDEKER
P.O. Box 126
Delft

NORWAY
Norwegian Defense Research Establishment
Main Library, c/o Mr P.L.EKERN
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Direccao do Servico de Material
da Forca Aerea
Rua de Escola Politecnica 42
Lisboa
Attn: Brig. General Jose de Sousa OLIVEIRA

TURKEY
Turkish General Staff (ARGE)
Ankara

UNITED KINGDOM
Ministry of Technology Reports Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

* * *

If copies of the original publication are not available at these centres, the following may be purchased from:

Microfiche or Photocopy
National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22151, USA

Microfiche
ESRO/ELDO Space
Documentation Service
European Space
Research Organization
114, Avenue de Neuilly
92, Neuilly-sur-Seine, France

Microfiche
Ministry of Technology
Reports Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE
England

The request for microfiche or photocopy of an AGARD document should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number.

Full bibliographical references and abstracts of the newly issued AGARD publications are given in the following bi-monthly abstract journals with indexes:

Scientific and Technical Aerospace Reports (STAR)
published by NASA,
Scientific and Technical Information Facility,
P.O. Box 33, College Park,
Maryland 20740, USA

United States Government Research and Development
Report Index (USGDRI), published by the
Clearinghouse for Federal Scientific and Technical
Information, Springfield, Virginia 22151, USA

Printed by Technical Editing and Reproduction Ltd
Harford House, 7-9 Charlotte St, London. W1P 1HD