CHAPTER

ONE

BASIC CONCEPTS

1-1 INTRODUCTION

The concept of optimization is basic to much of what we do in our daily lives. The desire to run a faster race, win a debate, or increase corporate profit implies a desire to do or be the best in some sense. In engineering, we wish to produce the "best quality of life possible with the resources available." Thus in "designing" new products, we must use design tools which provide the desired results in a timely and economical fashion. Numerical optimization is one of the tools at our disposal.

In studying design optimization, it is important to distinguish between analysis and design. Analysis is the process of determining the response of a specified system to its environment. For example, the calculation of stresses in a structure that result from applied loads is referred to here as analysis. Design, on the other hand, is used to mean the actual process of defining the system. For example, structural design entails defining the sizes and locations of members necessary to support a prescribed set of loads. Clearly, analysis is a sub-problem in the design process because this is how we evaluate the adequacy of the design.

Much of the design task in engineering is quantifiable, and so we are able to use the computer to analyze alternative designs rapidly. The purpose of numerical optimization is to aid us in rationally searching for the best design to meet our needs.

While the emphasis here is on design, it should be noted that these methods can often be used for analysis as well. Nonlinear structural analysis is an example where optimization can be used to solve a nonlinear energy minimization problem.

Although we may not always think of it this way, design can be defined as the process of finding the minimum or maximum of some parameter which may be called the objective function. For the design to be acceptable, it must also satisfy a certain set of specified requirements called constraints. That is, we wish to find the constrained minimum or maximum of the objective function. For example, assume we wish to design an internal-combustion engine. The design objective could be to maximize combustion efficiency. The engine may be required to provide a specified power output with an upper limit on the amount of harmful pollutants which can be emitted into the atmosphere. The power requirements and pollution restrictions are therefore constraints on the design.

Various methods can be used to achieve the design goal. One approach might be through experimentation where many engines are built and tested. The engine providing maximum economy while satisfying the constraints on the design would then be chosen for production. Clearly this is a very expensive approach with little assurance of obtaining a true optimum design. A second approach might be to define the design process analytically and then to obtain the solution using differential calculus or the calculus of variations. While this is certainly an attractive procedure, it is seldom possible in practical applications to obtain a direct analytical solution because of the complexities of the design and analysis problem.

Most design organizations now have computer codes capable of analyzing a design which the engineer considers reasonable. For example, the engineer may have a computer code which, given the compression ratio, air-fuel mixture ratio, bore and stroke, and other basic design parameters, can analyze the internal-combustion engine to predict its efficiency, power output, and pollution emissions. The engineer could then change these design variables and rerun the program until an acceptable design is obtained. In other words, the physical experimentation approach where engines are built and tested is replaced by numerical experimentation, recognizing that the final step will still be the construction of one or more prototypes to verify our numerical results.

With the availability of computer codes to analyze the proposed design, the next logical step is to automate the design process. In its most basic form, design automation may consist of a series of loops in the computer code which cycle through many combinations of design variables. The combination which provides the best design satisfying the constraints is then termed optimum. This approach has been used with some success and may be quite adequate if the analysis program uses a small amount of computer time. However, the cost of this technique increases dramatically as the number of design variables to be changed increases and as the computer time for a single analysis increases.

Consider, for example, a design problem described by three variables. Assume we wish to investigate the designs for 10 values of each variable.

Assume also that any proposed design can be analyzed in one-tenth of a central processing unit (CPU) second on a digital computer. There are then 10 combinations of design variables to be investigated, each requiring one-tenth second for a total of 100 CPU seconds to obtain the desired optimum design. This would probably be considered an economical solution in most design situations. However, now consider a more realistic design problem where 10 variables describe the design. Again, we wish to investigate 10 values of each variable. Also now assume that the analysis of a proposed design requires 10 CPU seconds on the computer. The total CPU time now required to obtain the optimum design is 10^{11} seconds, or roughly 3200 years of computer time! Clearly, for most practical design problems, a more rational approach to design automation is needed.

Numerical optimization techniques offer a logical approach to design automation, and many algorithms have been proposed in recent years. Some of these techniques, such as linear, quadratic, dynamic, and geometric programming algorithms, have been developed to deal with specific classes of optimization problems. A more general category of algorithms referred to as nonlinear programming has evolved for the solution of general optimization problems. Methods for numerical optimization are referred to collectively as mathematical programming techniques.

Though the history of mathematical programming is relatively short, roughly 45 years, there has been an almost bewildering number of algorithms published for the solution of numerical optimization problems. The author of each algorithm usually has numerical examples which demonstrate the efficiency and accuracy of the method, and the unsuspecting practitioner will often invest a great deal of time and effort in programming an algorithm, only to find that it will not in fact solve the particular optimization problem being attempted. This often leads to disenchantment with these techniques which can be avoided if the user is knowledgeable in the basic concepts of numerical optimization. There is an obvious need, therefore, for a unified, non-theoretical presentation of optimization concepts.

The purpose here is to attempt to bridge the gap between optimization theory and its practical applications. The remainder of this chapter will be devoted to a discussion of the basic concepts of numerical optimization. We will consider the general statement of the nonlinear constrained optimization problem and some (slightly) theoretical aspects regarding the existence and uniqueness of the solution to the optimization problem. Finally, we will consider some practical advantages and limitations to the use of these methods.

Numerical optimization has traditionally been developed in the operations research community. The use of these techniques in engineering design was popularized in 1960 when Schmit [1] applied nonlinear optimization techniques to structural design and coined the phrase "structural synthesis." While the work of Ref. 1 was restricted to structural optimization, the concepts presented there offered a fundamentally new approach to engineering design which is applicable to a wide spectrum of design problems. The basic concept is that the purpose of design is the allocation of scarce resources [2]. The purpose of numerical optimization is to provide a computer tool to aid the designer in this task.

1-2 OPTIMIZATION CONCEPTS

Here we will briefly describe the basic concepts of optimization by means of two examples.

Example 1-1 Unconstrained function minimization
Assume we wish to find the minimum value of the following simple algebraic function.

$$F(\mathbf{X}) = 10X_1^4 - 20X_1^2X_2 + 10X_2^2 + X_1^2 - 2X_1 + 5 \quad (1\text{-}1)$$

$F(\mathbf{X})$ is referred to as the objective function which is to be minimized, and we wish to determine the combination of the variables X_1 and X_2 which will achieve this goal. The vector $\mathbf{X}$ contains X_1 and X_2 and we call them the design, or decision, variables. No limits are imposed on the values of X_1 and X_2 and no additional conditions must be met for the "design" to be acceptable. Therefore, $F(\mathbf{X})$ is said to be unconstrained. Figure 1-1 is a graphical representation of the function, where lines of constant value of $F(\mathbf{X})$ are drawn. This function is often referred to as the *banana function* because of its distinctive geometry. Figure 1-1 is referred to as a two-variable design space, where the design variables X_1 and X_2 correspond to the coordinate axes. In general, a design space will be n dimensional, where n is the number of design variables of which the objective is a function. The two-variable design space will be used throughout our discussion of optimization techniques to help visualize the various concepts.

From Figure 1-1 we can estimate that the minimum value of $F(\mathbf{X})$ will occur at $X_1^* = 1$ and $X_2^* = 1$. We know also from basic calculus that at the optimum, or minimum, of F($\mathbf{X}$), the partial derivatives with respect to X_1 and X_2 must vanish. That is

$$\frac{\partial}{\partial X_1}F(\mathbf{X}) = 40X_1^3 - 40X_1X_2 + 2X_1 - 2 = 0 \quad (1\text{-}2)$$

$$\frac{\partial}{\partial X_2} F(\mathbf{X}) = -20X_1^2 + 20X_2 = 0 \tag{1-3}$$

Solving for X_1 and X_2, we find that indeed $X_1^* = 1$ and $X_2^* = 1$. We will see later that the vanishing gradient is a necessary but not sufficient condition for finding the minimum.

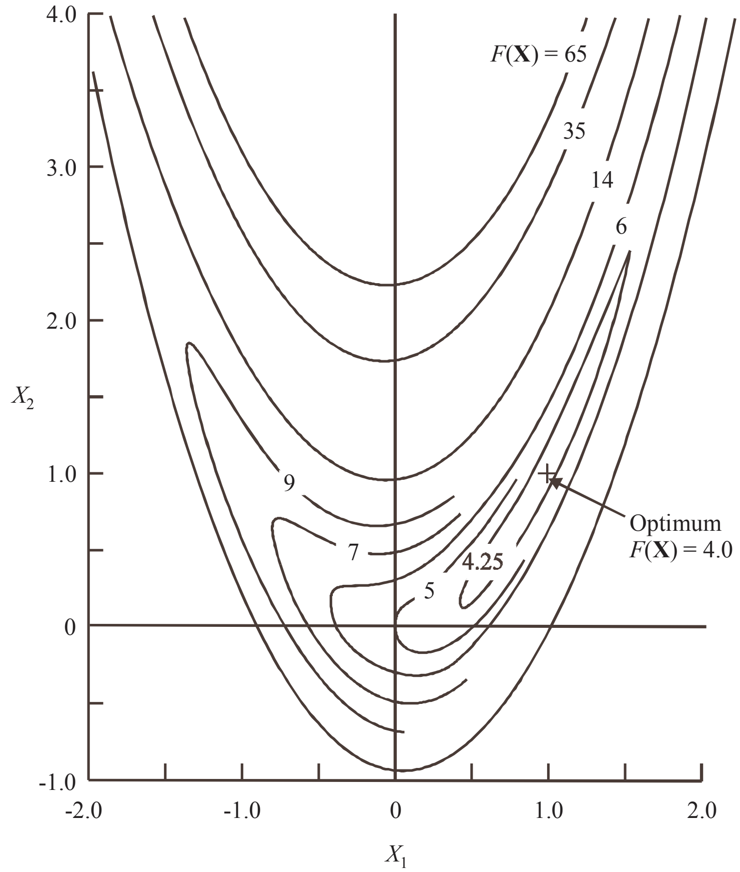

Figure 1-1 Two-variable function space.

In this example, we were able to obtain the optimum both graphically and analytically. However, this example is of little engineering value, except for demonstration purposes. In most practical engineering problems the minimum of a function cannot be determined analytically. The problem is further complicated if the decision variables are restricted to values within a specified range or if other conditions are imposed in the minimization problem. Therefore, numerical techniques are usually resorted to. We will now consider a simple design example where conditions (constraints) are imposed on the optimization problem.

Example 1-2 Constrained function minimization

Figure 1-2a depicts a tubular column of height h which is required to support a concentrated load P as shown. We wish to find the mean diameter D and the wall thickness t to minimize the weight of the column. The column weight is given by

$$W = \rho A h = \rho \pi D t h \tag{1-4}$$

where A is the cross-sectional area and ρ is the material's unit weight.

We will consider the axial load only, and for simplicity will ignore any eccentricity, lateral loads, or column imperfections. The stress in the column is given by

$$\sigma = \frac{P}{A} = \frac{P}{\pi D t} \tag{1-5}$$

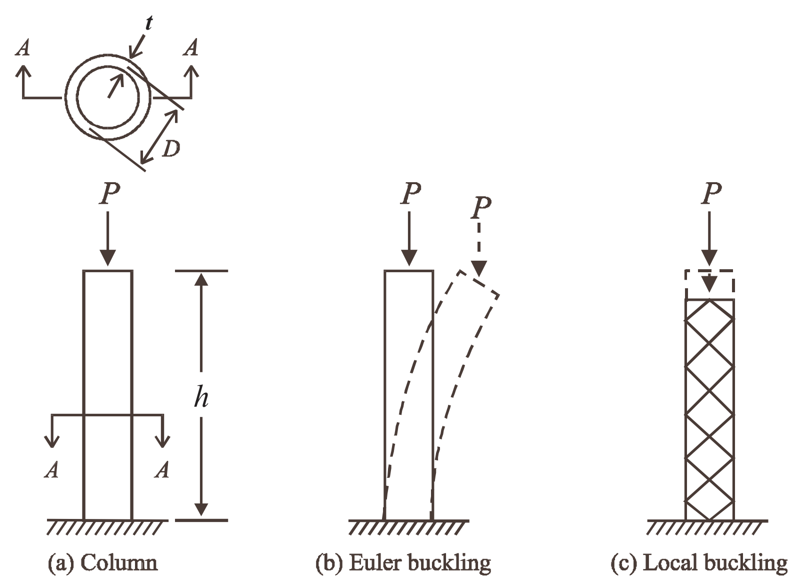

Figure 1-2 Column design for least weight.

where stress is taken as positive in compression. In order to prevent material failure, this stress must not exceed the allowable stress $\bar{\sigma}$. In addition to preventing material failure, the stress must not exceed that at which Euler buckling or local shell buckling will occur, as shown in Figs. 1-2b and c. The stress at which Euler buckling occurs is given by

$$\sigma_b = \frac{\pi^2 EI}{4Ah^2} = \frac{\pi^2 E(D^2 + t^2)}{32h^2} \tag{1-6}$$

where E = Young's modulus
I = moment of inertia
The stress at which shell buckling occurs is given by

$$\sigma_s = \frac{2Et}{D\sqrt{3(1-\nu^2)}} \tag{1-7}$$

where ν = Poisson's ratio
The column must now be designed so that the magnitude of the stress is less than the minimum of $\bar{\sigma}$, σ_b, and σ_s. These requirements can be written algebraically as

$$\sigma \le \bar{\sigma} \tag{1-8}$$

$$\sigma \le \sigma_b \tag{1-9}$$

$$\sigma \le \sigma_s \tag{1-10}$$

In addition to the stress limitations, the design must satisfy the geometric conditions that the mean diameter be greater than the wall thickness and that both the diameter and thickness be positive

$$D \ge t \tag{1-11}$$

$$D \ge 10^{-6} \tag{1-12}$$

$$t \ge 10^{-6} \tag{1-13}$$

Bounds of 10^{-6} are imposed on D and t to ensure that σ in Eq. (1-5) and σ_s in Eq. (1-7) will be finite.

The design problem can now be stated compactly as

Minimize: $$W(D,t) = \rho\pi Dth \tag{1-14}$$

Subject to:

$$g_1(D, t) = \frac{\sigma}{\bar{\sigma}} - 1 \le 0 \tag{1-15a}$$

$$g_2(D, t) = \frac{\sigma}{\sigma_b} - 1 \le 0 \tag{1-15b}$$

$$g_3(D, t) = \frac{\sigma}{\sigma_s} - 1 \le 0 \tag{1-15c}$$

$$g_4(D, t) = t - D \le 0 \tag{1-15d}$$

$$D \ge 10^{-6} \tag{1-16a}$$

$$t \ge 10^{-6} \tag{1-16b}$$

where $\bar{\sigma}$, σ_b, and σ_s are given by Eqs. (1-5), (1-6), and (1-7), respectively. To summarize, Eq. (1-14) defines the objective function and Eqs. (1-15a) - (1-15d) and (1-16a, 1-16b) define the constraints on the design problem. Note that Eq. (1-15a to c) is just a normalized form of Eqs. (1-8) to (1-10). The constraints given by Eq. (1-16) are often referred to as side constraints because they directly impose bounds on the value of the design variables. Figure 1-3 is the design space associated with the column design problem. In addition to contours of constant objective, the constraint boundaries $[g_j(\mathbf{X}) = 0]$ are also drawn in the design space. That portion of the design space inside the constraint boundaries defined by the hatched lines is referred to as the feasible design space, and all designs in this region are acceptable. Any design which violates these constraint boundaries is unacceptable and is referred to as infeasible. This figure represents a simple example of the general nonlinear constrained optimization problem.

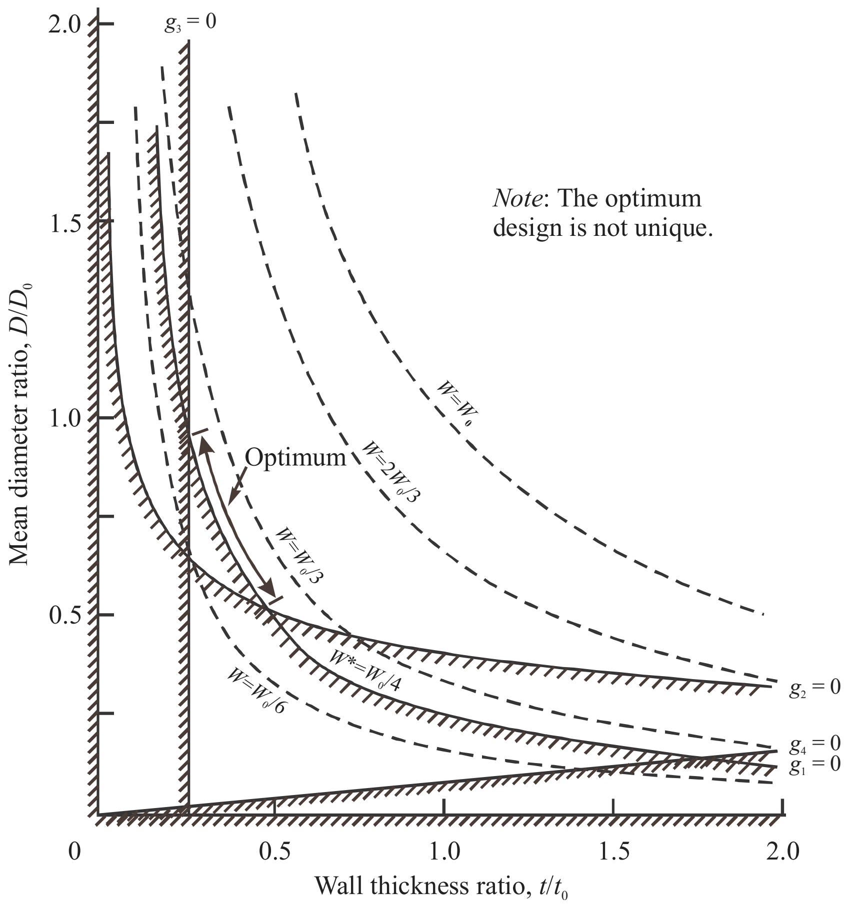

Figure 1-3 Two-variable function space for column.

1-3 GENERAL PROBLEM STATEMENT

We can now write the nonlinear constrained optimization problem mathematically as follows:

$$\text{Minimize: } F(\mathbf{X}) \qquad \text{objective function} \tag{1-17}$$

Subject to:

$$g_j(\mathbf{X}) \le 0 \qquad j=1,m \qquad \text{inequality constraints} \tag{1-18}$$

$$h_k(\mathbf{X}) = 0 \qquad k=1,l \qquad \text{equality constraints} \tag{1-19}$$

$$X_i^l \le X_i \le X_i^u \qquad i=1,n \qquad \text{side constraints} \tag{1-20}$$

$$\text{where } \mathbf{X} = \begin{Bmatrix} X_1 \\ X_2 \\ X_3 \\ . \\ . \\ . \\ X_n \end{Bmatrix} \qquad \text{design variables}$$

The vector **X** is referred to as the vector of design variables. In the column design given above, this vector would contain the two variables D and t. The objective function F(**X**) given by Eq. (1-17), as well as the constraint functions defined by Eqs. (1-18) and (1-19) may be linear or nonlinear functions of the design variables **X.** These functions may be explicit or implicit in **X** and may be evaluated by any analytical or numerical techniques we have at our disposal. However, except for special classes of optimization problems, it is important that these functions be continuous and have continuous first derivatives in **X.**

In the column design example, we considered only inequality constraints of the form given by Eq. (1-18). Additionally, we now include the set of equality constraints $h_k(\mathbf{X})$ as defined by Eq. (1-19). If equality constraints are explicit in **X,** they can often be used to reduce the number of design variables considered. For example, in the column design problem, we may wish to require the thickness be one-tenth the value of the diameter, that is, $t = 0.1D$. This information could be substituted directly into the problem statement to reduce the design problem to one in diameter D only.

In general, $h(\mathbf{X})$ may be either a very complicated explicit function of the design variables **X** or may be implicit in **X.**

Equation (1-20) defines bounds on the design variables **X** and so is referred to as a side constraint. Although side constraints could be included in the inequality constraint set given by Eq. (1-18), it is usually convenient to treat them separately because they define the region of search for the optimum.

The above form of stating the optimization problem is not unique, and various other statements equivalent to this are presented in the literature. For example, we may wish to state the problem as a maximization problem where we desire to maximize $F(\mathbf{X})$. Similarly, the inequality sign in Eq. (1-18) can be reversed so that $g(\mathbf{X})$ must be greater than or equal to zero. Using our notation, if a particular optimization problem requires maximization, we simply minimize $-F(\mathbf{X})$. The choice of the non-positive inequality sign on the constraints has the geometric significance that, at the optimum, the gradients of the objective and all critical constraints point away from the optimum design.

1-4 THE ITERATIVE OPTIMIZATION PROCEDURE

Most optimization algorithms require that an initial set of design variables, $\mathbf{X}^0$**,** be specified. Beginning from this starting point, the design is updated iteratively. Probably the most common form of this iterative procedure is given by

$$\mathbf{X}^q = \mathbf{X}^{q-1} + \alpha^* \mathbf{S}^q \tag{1-21}$$

where q is the iteration number and **S** is a vector search direction in the design space. The scalar quantity α^* defines the distance that we wish to move in direction **S**.

To see how the iterative relationship given by Eq. (1-21) is applied to the optimization process, consider the two-variable problem shown in Figure 1-4.

Assume we begin at point $\mathbf{X}^0$ and we wish to reduce the objective function. We will begin by searching in the direction $\mathbf{S}^1$ given by

$$\mathbf{S}^1 = \begin{Bmatrix} -1.0 \\ -0.5 \end{Bmatrix} \tag{1-22}$$

The choice of **S** is somewhat arbitrary as long as a small move in this direction will reduce the objective function without violating any constraints. In this case, the $\mathbf{S}^1$ vector is approximately the negative of the gradient of the objective function, that is, the direction of steepest descent. It is now necessary to find the scalar α^* in Eq. (1-21) so that the objective is minimized in this direction without violating any constraints.

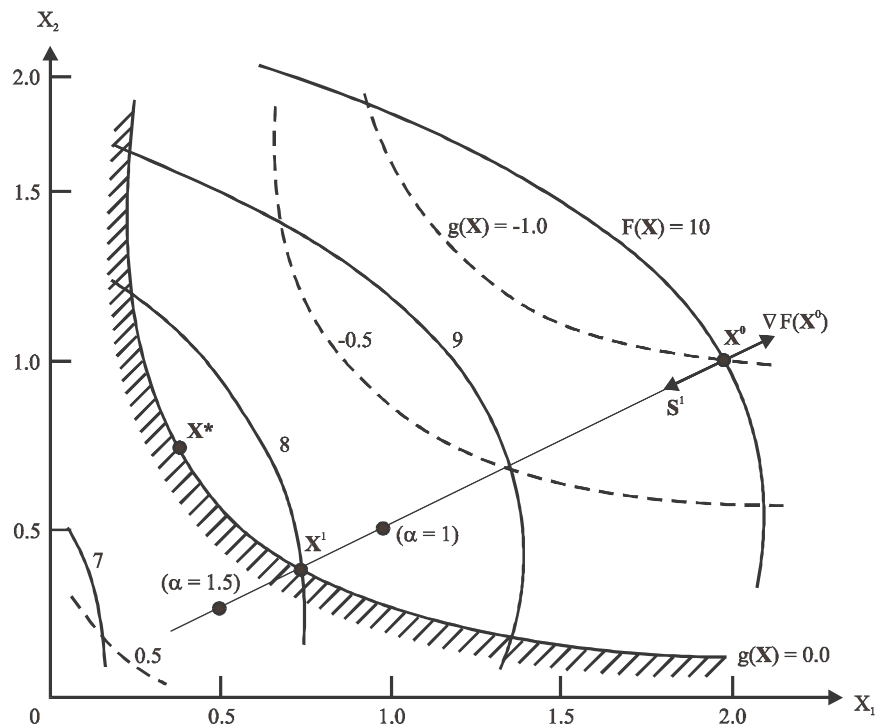

Figure 1-4 Search in direction S.

We now evaluate **X** and the corresponding objective and constraint functions for several values of α to give

$$\alpha = 0 \qquad \mathbf{X} = \begin{Bmatrix} 2.0 \\ 1.0 \end{Bmatrix}$$

$$F(\alpha) = 10.0 \qquad g(\alpha) = -1.0 \tag{1-23a}$$

$$\alpha = 1.0 \qquad \mathbf{X} = \begin{Bmatrix} 2.0 \\ 1.0 \end{Bmatrix} + 1.0 \begin{Bmatrix} -1.0 \\ -0.5 \end{Bmatrix} = \begin{Bmatrix} 1.0 \\ 0.5 \end{Bmatrix}$$

$$F(\alpha) = 8.4 \qquad g(\alpha) = -0.2 \tag{1-23b}$$

$$\alpha = 1.5 \qquad \mathbf{X} = \begin{Bmatrix} 2.0 \\ 1.0 \end{Bmatrix} + 1.5 \begin{Bmatrix} -1.0 \\ -0.5 \end{Bmatrix} = \begin{Bmatrix} 0.50 \\ 0.25 \end{Bmatrix}$$

$$F(\alpha) = 7.6 \qquad g(\alpha) = 0.2 \tag{1-23c}$$

$$\alpha^* = 1.25 \qquad \mathbf{X}^* = \begin{Bmatrix} 2.0 \\ 1.0 \end{Bmatrix} + 1.25 \begin{Bmatrix} -1.0 \\ -0.5 \end{Bmatrix} = \begin{Bmatrix} 0.750 \\ 0.375 \end{Bmatrix}$$

$$F(\alpha^*) = 8.0 \qquad g(\alpha^*) = 0.0 \tag{1-23d}$$

where the objective and constraint values are estimated using Figure 1-4. In practice, we would evaluate these functions on the computer, and, using several proposed values of α, we would apply a numerical interpolation scheme to estimate α^*. This would provide the minimum $F(\mathbf{X})$ in this search direction which does not violate any constraints. Note that by searching in a specified direction, we have actually converted the problem from one in n variables $\mathbf{X}$ to one variable α. Thus, we refer to this as a one-dimensional search. At point $\mathbf{X}^1$, we must find a new search direction such that we can continue to reduce the objective without violating constraints. In this way, Eq. (1-21) is used repetitively until no further design improvement can be made.

From this simple example, it is seen that nonlinear optimization algorithms based on Eq. (1-21) can be separated into two basic parts. The first is determination of a direction of search $\mathbf{S}$, which will improve the objective function subject to constraints. The second is determination of the scalar parameter α^* defining the distance of travel in direction $\mathbf{S}$. Each of these components plays a fundamental role in the efficiency and reliability of a given optimization algorithm, and each will be discussed in detail in later chapters.

1-5 EXISTENCE AND UNIQUENESS OF AN OPTIMUM SOLUTION

In the application of optimization techniques to design problems of practical interest, it is seldom possible to ensure that the absolute optimum design will be found subject to the constraints. This may be because multiple solutions to the optimization problem exist or simply because numerical ill-conditioning in setting up the problem results in extremely slow convergence of the optimization algorithm. From a practical standpoint, the best approach is usually to start the optimization process from several different initial vectors, and if the optimization results in essentially the same final design, we can be reasonably assured that this is the true optimum. It is, however, possible to check mathematically to determine if we at least have a relative minimum. In other words, we can define necessary conditions for an optimum, and we can show that under certain circumstances these necessary conditions are also sufficient to ensure that the solution is the global optimum.

1-5.1 Unconstrained Problems

First consider the unconstrained minimization problem where we only wish to minimize $F(\mathbf{X})$ with no constraints imposed. We know that for $F(\mathbf{X})$ to be minimum, the gradient of $F(\mathbf{X})$ must vanish

$$\nabla \mathbf{F}(\mathbf{X}) = \mathbf{0} \tag{1-24a}$$

where

$$\nabla \mathbf{F}(\mathbf{X}) = \begin{Bmatrix} \frac{\partial}{\partial X_1} F(\mathbf{X}) \\ \frac{\partial}{\partial X_2} F(\mathbf{X}) \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial}{\partial X_n} F(\mathbf{X}) \end{Bmatrix} \tag{1-24b}$$

However, this is only a necessary condition and is not sufficient to ensure optimality. This is easily seen by referring to Figure 1-5, which is a function of one variable. Here the gradient of $F(\mathbf{X})$ is simply the first derivative of $F(\mathbf{X})$ with respect to the single variable X. Clearly the gradient of $F(X)$ vanishes at three points in the figure, A, B, and C. Point A defines the minimum and point C defines the maximum. Point B is neither the minimum nor the maximum.

We also know from calculus that in order for a function of one variable to be a minimum, its second derivative with respect to the independent variable must be positive, and this is certainly true at point A in Figure 1-5. In the general n dimensional case this translates into the requirement that the Hessian matrix be positive definite, where the Hessian matrix is the matrix of second partial derivatives of the objective with respect to the design variables

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 F(\mathbf{X})}{\partial X_1^2} & \dfrac{\partial^2 F(\mathbf{X})}{\partial X_1 \partial X_2} & \cdots & \dfrac{\partial^2 F(\mathbf{X})}{\partial X_1 \partial X_n} \\ \dfrac{\partial^2 F(\mathbf{X})}{\partial X_2 \partial X_1} & \dfrac{\partial^2 F(\mathbf{X})}{\partial X_2^2} & \cdots & \dfrac{\partial^2 F(\mathbf{X})}{\partial X_2 \partial X_n} \\ \cdots & \cdots & \ddots & \cdots \\ \dfrac{\partial^2 F(\mathbf{X})}{\partial X_n \partial X_1} & \dfrac{\partial^2 F(\mathbf{X})}{\partial X_n \partial X_2} & \cdots & \dfrac{\partial^2 F(\mathbf{X})}{\partial X_n^2} \end{bmatrix} \tag{1-25}$$

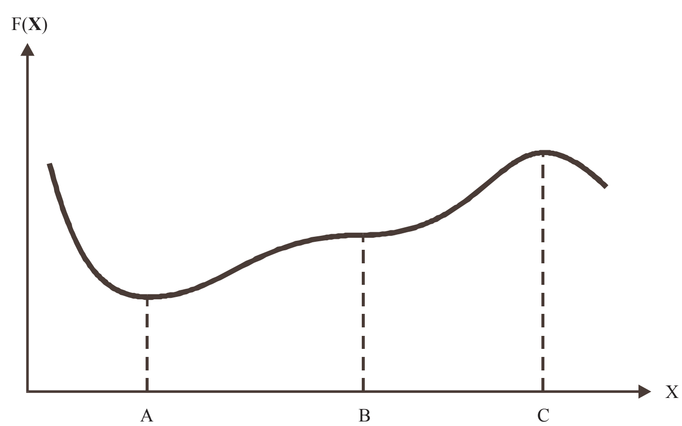

Figure 1-5 Relative optima of an unconstrained function.

Positive definiteness means that this matrix has all positive eigenvalues. If the gradient is zero and the Hessian matrix is positive definite for a given **X**, this insures that the design is at least a relative minimum, but again it does not insure that the design is a global minimum. The design is only guaranteed to be a global minimum if the Hessian matrix is positive definite for all possible values of the design variables **X.** This can seldom be demonstrated in practical design applications. We must usually be satisfied with starting the design from various initial points to see if we can obtain a consistent optimum and therefore have reasonable assurance that this is the true minimum of the function. However, an understanding of the requirements for a unique optimal solution is important to provide insight into the optimization process. Also, these concepts provide the basis for the development of many of the more powerful algorithms which we will be discussing in later chapters.

1-5.2 Constrained Problems

Now, consider the constrained minimization problem and assume that, at the optimum, at least one constraint on the design is active. It is no longer necessary that the gradient of the objective vanish at the optimum. Referring to Figure 1-3, this is obvious. At the optimum, the objective function, being the weight of the tubular column, could be reduced by either reducing the diameter or the wall thickness so that the components of the gradient of the objective function are clearly not zero at this point. However, any reduction in the dimensions of the column in order to reduce the weight would result in constraint violations.

We can, at least intuitively, define the necessary conditions for a constrained optimum by referring to Figure 1-6. Assume a design is specified at point *A*. In order to improve on this design, it will be necessary to determine a direction vector **S** which will reduce the objective function and yet not violate the active constraint. We define any direction which will reduce the objective as a usable direction. Clearly the line tangent to a line of constant value of the objective function at this point will bound all possible usable directions. Any direction on the side of this line (hyperplane) which will reduce the objective function is defined as a usable direction, and this portion of the design space is referred to as the usable sector. Note that if we take the scalar product of any direction vector **S** which we choose in the usable sector, with the gradient of the objective function, the result will be negative to zero (that is, $\boldsymbol{S}^T \nabla \boldsymbol{F}(\boldsymbol{X}) \leq 0$). In other words, the angle between these two vectors must be between 90° and 270°. This is because the scalar product is the product of the magnitudes of the vectors and the cosine of the angle between them. The cosine is only negative for angles between 90° and

270^{o}. Similarly, a hyperplane tangent to the constraint surface at point A will bound the feasible sector, where a direction vector **S** is defined as feasible if a small move in this direction will not violate the constraint. In this case, the scalar product of **S** with the gradient of the constraint is negative or zero ($\boldsymbol{S}^T \nabla \boldsymbol{g}(\boldsymbol{X}) \leq 0$), so the angle between these two vectors must also be between 90^{o} and 270^{o}. In order for a direction vector **S** to yield an improved design without violating the constraint, this direction must be both usable and feasible, and any direction in the usable feasible sector of the design space satisfies this criterion. It is noteworthy here that if the direction vector is nearly tangent to the hyperplane bounding the feasible sector, a small move in this direction will result in a constraint violation but will reduce the objective function quite rapidly. On the other hand, if the direction vector is chosen nearly tangent to the line of constant objective, we can move some distance in this direction without violating the constraint, but the objective function will not decrease rapidly and, if the objective is nonlinear, may in fact begin to increase. At point A in Figure 1-6 a direction does exist which reduces the objective function without violating the constraint for a finite move.

In the general optimization problem there may be more than one active constraint at a given time in the design process, where a constraint is considered active if its value is within a small tolerance of zero. If this is the case, a direction vector must be found which is feasible with respect to all active constraints. The requirement that a move direction be both usable and feasible is stated mathematically as

Usable direction:

$$\nabla \mathbf{F}(\mathbf{X})^T \mathbf{S} \leq 0 \tag{1-26}$$

Feasible direction:

$$\nabla \mathbf{g_j}(\mathbf{X})^T \mathbf{S} \leq 0 \qquad \text{for all } j \text{ for which } g_j(\mathbf{X}) = 0 \tag{1-27}$$

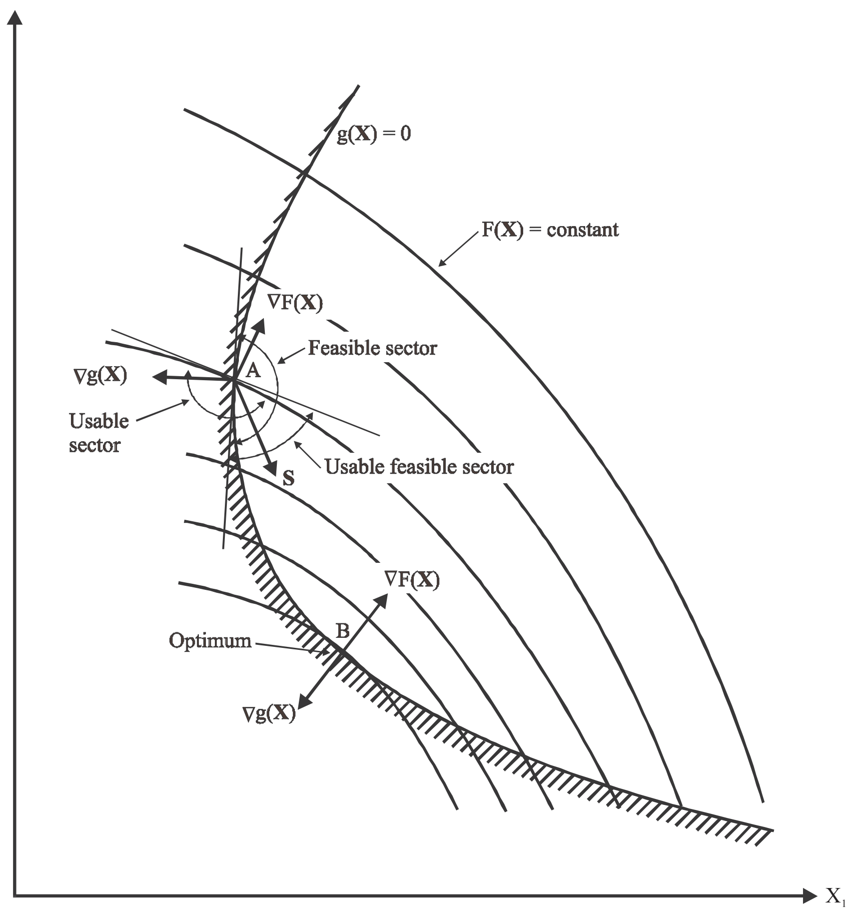

Figure 1-6 Usable-feasible search direction.

Now consider point ***B*** in Figure 1-6, which is the optimum design for this example. Here the gradient of the objective and the gradient of the constraint point in exactly the opposite directions. Therefore, the only possible vector **S** which will satisfy the requirements of usability and feasibility will be precisely tangent both to the constraint boundary and to a line of constant objective function and will form an angle of 90° with both gradients. This condition can be stated mathematically as

$$\nabla \mathbf{F}(\mathbf{X}) + \sum_{j=1}^{m} \lambda_j \nabla \boldsymbol{g_j}(\mathbf{X}) + \sum_{k=1}^{l} \lambda_{m+k} \nabla \boldsymbol{h_k}(\mathbf{X}) = \mathbf{0} \tag{1-28}$$

$$\lambda_j \geq 0 \tag{1-29}$$

$$\lambda_{m+k} \text{ unrestricted in sign} \tag{1-30}$$

Note that if there are no constraints on the design problem, Eq. (1-28) reduces to the requirement that the gradient of the objective function must vanish, as was already discussed in the case of unconstrained minimization.

While Eq. (1-28) defines a necessary condition for a design to be a constrained optimum, it is by no means sufficient, as can be seen from Figure 1-7. Points ***A***, ***B***, and ***C*** in the figure all satisfy the requirements of Eq. (1-28) and yet only point ***C*** is the true global optimum.

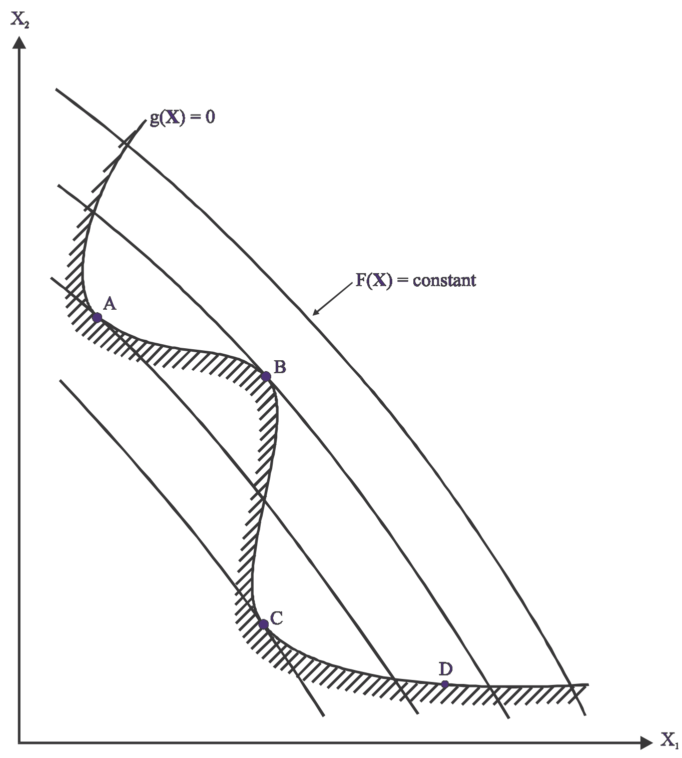

Figure 1-7 Relative optima of a constrained function.

1-5.3 The Kuhn-Tucker Conditions

Equation (1-28) is actually the third of a set of necessary conditions for constrained optimality. These are referred to as the Kuhn-Tucker necessary conditions.

The Kuhn-Tucker conditions define a stationary point of the Lagrangian

$$\mathbf{L}(\mathbf{X}, \lambda) = \mathbf{F}(\mathbf{X}) + \sum_{j=1}^{m} \lambda_j \boldsymbol{g_j}(\mathbf{X}) + \sum_{k=1}^{l} \lambda_{m+k} \boldsymbol{h_k}(\mathbf{X}) \tag{1-31}$$

All three conditions are listed here for reference and state simply that if the vector **X*** defines the optimum design, the following conditions must be satisfied:

$$1.\ \mathbf{X}^* \text{ is feasible} \tag{1-32}$$

$$2.\ \lambda_j g_j(\mathbf{X}^*) = 0 \qquad j = 1, m \qquad \lambda_j \geq 0 \tag{1-33}$$

$$3.\ \nabla\mathbf{F}(\mathbf{X}^*) + \sum_{j=1}^{m} \lambda_j \nabla \boldsymbol{g_j}(\mathbf{X}^*) + \sum_{k=1}^{l} \lambda_{m+k} \nabla \boldsymbol{h_k}(\mathbf{X}^*) = \mathbf{0} \tag{1-34}$$

$$\lambda_j \geq 0 \tag{1-35}$$

$$\lambda_{m+k} \text{ unrestricted in sign} \tag{1-36}$$

Equation (1-32) is a statement of the obvious requirement that the optimum design must satisfy all constraints. Equation (1-33) imposes the requirement that if the constraint $g_j(\mathbf{X})$ is not precisely satisfied [that is, $g_j(\mathbf{X}) < 0$] then the corresponding Lagrange multiplier must be zero. Equations (1-34) to (1-36) are the same as Eqs. (1-28) to (1-30).

The geometric significance of the Kuhn-Tucker conditions can be understood by referring to Figure 1-8, which shows a two-variable minimization problem with three inequality constraints. At the optimum, constraint $g_3(\mathbf{X}^*)$ is not critical and so, from Eq. (1-33), $\lambda_3 = 0$. Equation (1-34) requires that, if we multiply the gradient of each critical constraint [$g_1(\mathbf{X}^*)$ and $g_2(\mathbf{X}^*)$] by its corresponding Lagrange multiplier, the vector sum of the result must equal the negative of the gradient of the objective function. Thus we see from Figure 1-8 that

$$\nabla \mathbf{F}(\mathbf{X}^*) + \lambda_1 \nabla \boldsymbol{g}_1(\mathbf{X}^*) + \lambda_2 \nabla \boldsymbol{g}_2(\mathbf{X}^*) = \mathbf{0} \tag{1-37a}$$

$$\lambda_1 \geq 0 \qquad \lambda_2 \geq 0 \tag{1-37b}$$

Since $g_1(\mathbf{X}^*) = 0$ and $g_2(\mathbf{X}^*) = 0$, the second Kuhn-Tucker condition is satisfied identically with respect to these constraints. Also, we see from Figure 1-8 that **X*** is feasible and so each of the Kuhn-Tucker necessary conditions is satisfied.

In the problem of Figure 1-8, the Lagrange multipliers are uniquely determined from the gradients of the objective and the active constraints. However, we can easily imagine situations where this is not so. For example, assume that we have defined another constraint, $g_4(\mathbf{X})$, which happens to be identical to $g_1(\mathbf{X})$ or perhaps a constant times $g_1(\mathbf{X})$. The constraint boundaries $g_1(\mathbf{X}) = 0$ and $g_4(\mathbf{X}) = 0$ would now be the same and the Lagrange multipliers λ_1 and λ_4 can have any combination of values which satisfy the vector addition shown in the figure. Thus, we can say that one of the constraints is redundant. As another example, we can consider a constraint $g_5(\mathbf{X})$ which is independent of the other constraints, but at the optimum in Figure 1-8, constraints $g_1(\mathbf{X})$, $g_2(\mathbf{X})$ and $g_5(\mathbf{X})$ are all critical. Now, we may pick many combinations of λ_1, λ_2, and λ_5 which will satisfy the Kuhn-Tucker conditions so that, while all constraints are independent, the Lagrange multipliers are not unique. These special cases do not detract from the usefulness of the Kuhn-Tucker conditions in optimization theory. It is only necessary that we account for these possibilities when using algorithms that require calculation of the Lagrange multipliers.

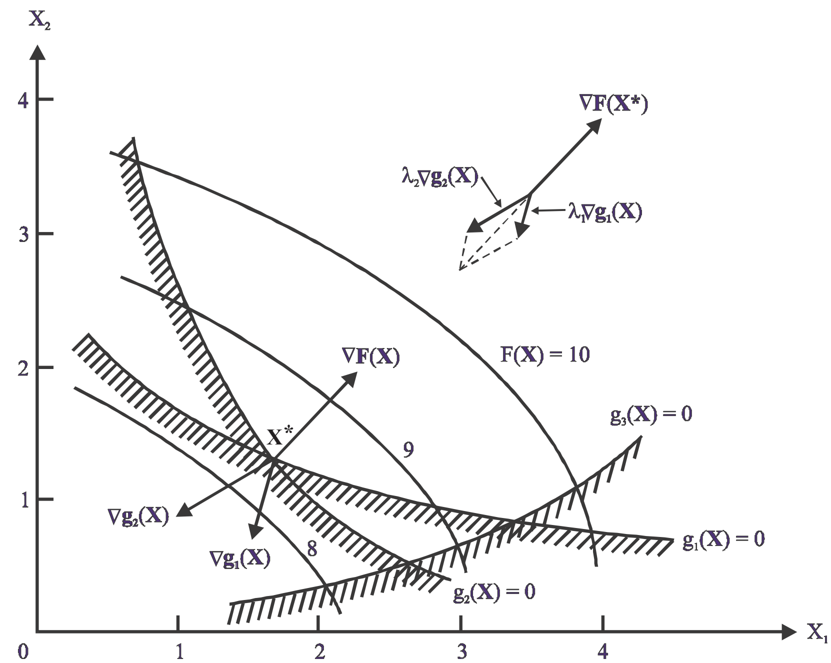

Figure 1-8 Geometric interpretation of the Kuhn-Tucker conditions.

The question now arises, when are the Kuhn-Tucker conditions both necessary and sufficient to define a global optimum? We can intuitively define these requirements by referring to Figs. 1-6 and 1-7. Assume we draw a straight line between points ***A*** and ***B*** in Figure 1-6. Any design which lies on this line will be a feasible design. On the other hand, in Figure 1-7 if we draw a straight line between points ***A*** and **C** or points ***A*** and ***D***, at least some portion of this line will be outside the feasible region. If we pick a line connecting any two points in the feasible region and all points on the line lie within the feasible region, as in Figure 1-6, we say that the constraint surfaces are convex.

A similar argument holds when we consider the objective function. If the objective function satisfies the requirements that it can only have one global optimum as discussed previously in the case of unconstrained minimization, then the objective is said to be convex, that is, the Hessian matrix of the objective function is positive definite for all possible designs. If the objective function and all constraint surfaces are convex, then the design space is said to be convex and the necessary Kuhn-Tucker conditions are also sufficient to guarantee that if we obtain an optimum it is the global optimum. This definition of sufficient conditions for a global optimum is actually more restrictive than is theoretically required but is adequate to provide a basic understanding of the necessary and sufficient conditions under which a design is the true global optimum. A detailed discussion of the Kuhn-Tucker conditions as well as a concise definition of the sufficiency requirements can be found in Ref. 3.

Just as in the case of unconstrained minimization, it is seldom possible in practical applications to know whether the sufficiency conditions are met. However, most optimization problems can be stated in many different forms, and an understanding of the desirable characteristics of the design space is often useful in helping us to cast the design problem in a form conducive to solution using numerical optimization.

Calculating the Lagrange Multipliers

Now consider how we might calculate the values of the Lagrange Multipliers at the optimum. First, we know that if a constraint value is non-zero (within a small tolerance), then from Eq. (1-33), the corresponding Lagrange multiplier is equal to zero. For our purposes here, both inequality and equality constraints are treated the same, so we can treat them all together. It is only important to remember that the equality constraints will always be active the optimum and that they can have positive or negative Lagrange Multipliers. Also, assuming all constraints are independent, the number of active constraints will be less than or equal to the number of design variables. Thus, Eq. (1-34) often has fewer unknown parameters, λ_j than equations.

Because precise satisfaction of the Kuhn-Tucker conditions may not be reached, we can rewrite Eq. (1-34) as

$$\nabla \mathbf{F}(\mathbf{X}^*) + \sum_{j=1}^{m} \lambda_j \nabla \boldsymbol{g_j}(\mathbf{X}^*) = \mathbf{R} \tag{1-38}$$

where the equality constraints are omitted for brevity and $\mathbf{R}$ is the vector of residuals.

Now, because we want the residuals as small as possible (if all components of $\mathbf{R} = 0$, the Kuhn-Tucker conditions are satisfied precisely), we can minimize the square of the magnitude of $\mathbf{R}$.
Let

$$\mathbf{B} = \nabla \mathbf{F}(\mathbf{X}^*) \tag{1-39a}$$

and

$$\mathbf{A} = \left[\nabla \boldsymbol{g}_1(\mathbf{X}^*)\ \nabla \boldsymbol{g}_2(\mathbf{X}^*)\ \ldots\ \nabla \boldsymbol{g}_M(\mathbf{X}^*)\right] \tag{1-39b}$$

where M is the set of active constraints.

Substituting Eqs. (1-39a) and (1-39b) into Eq. (1-38),

$$\mathbf{B} + \mathbf{A}\lambda = \mathbf{R} \tag{1-40}$$

Now, because we want the residuals as small as possible (if all components of $\mathbf{R} = 0$, the Kuhn-Tucker conditions are satisfied precisely), we can minimize the square of the magnitude of $\mathbf{R}$.

$$\text{Minimize} \quad \mathbf{R}^T\mathbf{R} = \mathbf{B}^T\mathbf{B} + 2\lambda^T\mathbf{A}^T\mathbf{B} + \lambda^T\mathbf{A}^T\mathbf{A}\lambda \tag{1-41}$$

Differentiating Eq. (1-41) with respect to λ and setting the result to zero gives

$$2\mathbf{A}^T\mathbf{B} + \mathbf{A}^T\mathbf{A}\lambda = 0 \tag{1-42}$$

from which

$$\lambda = -2[\mathbf{A}^T\mathbf{A}]^{-1}\mathbf{A}^T\mathbf{B} \tag{1-43}$$

Now if all components of λ corresponding to inequality constraints are non-negative, we have an acceptable estimate of the Lagrange multipliers. Also, we can substitute Eq. (1-43) into Eq. (1-40) to estimate how precisely the Kuhn-Tucker conditions are met. If all components of the residual vector, $\mathbf{R}$, are very near zero, we know that we have reached at least a relative minimum.

Sensitivity of the Optimum to Changes in Constraint Limits

The Lagrange multipliers have particular significance in estimating how sensitive the optimum design is to the active constraints. It can be shown that the derivative of the optimum objective with respect to a constraint is just the value of the Lagrange multiplier of that constraint, so

$$\frac{\partial}{\partial g_j(\mathbf{X}^*)} F(\mathbf{X}^*) = \lambda_j \tag{1-44}$$

or, in more useful form;

$$F(\mathbf{X}^*, \delta g_j) = \mathrm{F}(\mathbf{X}^*) + \lambda_j \delta g_j \tag{1-45}$$

If we wish to change the limits on a set of constraints, J, Eq. (1-45) is simply expanded as

$$F(\mathbf{X}^*, \delta g_j) = \mathrm{F}(\mathbf{X}^*) + \sum_{j \varepsilon J} \lambda_j \delta g_j \tag{1-46}$$

Remember that Eq. (1-45) is the sensitivity with respect to g_j. In practice, we may want to know the sensitivity with respect to bounds on the response.

Assume we have normalized an upper bound constraint

$$g_j = \frac{R - R_u}{|R_u|} \tag{1-47}$$

where R is the response and R_u is the upper bound.

$$F(\mathbf{X}^*, \delta R_u) = \mathrm{F}(\mathbf{X}^*) - \lambda_j \frac{\delta R_u}{|R_u|} \tag{1-48}$$

Similarly, for lower bound constraints

$$F(\mathbf{X}^*, \delta R_l) = \mathrm{F}(\mathbf{X}^*) + \lambda_j \frac{\delta R_l}{|R_l|} \tag{1-49}$$

Therefore, the Lagrange multipliers tell us the sensitivity of the optimum with respect to a relative change in the constraint bounds, while the Lagrange multipliers divided by the scaling factor (usually the matnitude of the bound) give us the sensitivity to an absolute change in the bounds.

Example 1-3 Sensitivity of the Optimum

Consider the constrained minimization of a simple quadratic function with a single linear constraint.

Minimize $$F(\mathbf{X}) = X_1^2 + X_2^2 \tag{1-50}$$

Subject to; $$g(\mathbf{X}) = \frac{2 - (X_1 + X_2)}{2} \le 0 \tag{1-51}$$

At the optimum;

$$\mathbf{X}^* = \begin{Bmatrix} 1.0 \\ 1.0 \end{Bmatrix} \qquad F(\mathbf{X}^*) = 2.0 \qquad g(\mathbf{X}^*) = 0.0 \tag{1-52a}$$

and

$$\nabla F(\mathbf{X}^*) = \begin{Bmatrix} 2.0 \\ 2.0 \end{Bmatrix} \qquad \nabla g(\mathbf{X}^*) = \begin{Bmatrix} -0.5 \\ -0.5 \end{Bmatrix} \qquad \lambda = 4.0 \tag{1-52b}$$

Now, assume we wish to change the lower bound on g from 2.0 to 2.1. From Eq. (1-49) we get

$$F(\mathbf{X}^*, \delta R_l) = 2.0 + 4.0\left\{\frac{0.1}{2.0}\right\} = 2.20 \tag{1-53}$$

The true optimum for this case is

$$\mathbf{X}^* = \begin{Bmatrix} 1.05 \\ 1.05 \end{Bmatrix} \qquad F(\mathbf{X}^*, \delta R_l) = 2.205 \tag{1-54}$$

1-6 CONCLUDING REMARKS

In assessing the value of optimization techniques to engineering design, it is worthwhile to review briefly the traditional design approach. The design is often carried out through the use of charts and graphs which have been developed over many years of experience. These methods are usually an efficient means of obtaining a reasonable solution to traditional design problems. However, as the design task becomes more complex, we rely more heavily on the computer for analysis. If we assume that we have a computer code capable of analyzing our proposed design, the output from this program will provide a quantitative indication of the acceptability and optimality of the design. We may change one or more design variables and rerun the computer program to see if any design improvement can be obtained. We then take the results of many computer runs and plot the objective and constraint values versus the various design parameters. From these plots we can interpolate or extrapolate to what we believe to be the optimum design. This is essentially the approach that was used to obtain the optimum constrained minimum of the tubular column shown in Figure 1-3, and this is certainly an efficient and viable approach when the design is a function of only a few variables. However, if the design exceeds three variables, the true optimum may be extremely difficult to obtain graphically. Then, assuming the computer code exists for the analysis of the proposed design, automation of the design process becomes an attractive alternative. Mathematical programming simply provides a logical framework for carrying out this automated design process. Some advantages and limitations to the use of numerical optimization techniques are listed here.

1-6.1 Advantages of Numerical Optimization

- A major advantage is the reduction in design time – this is especially true when the same computer program can be applied to many design projects.
- Optimization provides a systematized logical design procedure.
- We can deal with a wide variety of design variables and constraints which are difficult to visualize using graphical or tabular methods.
- Optimization virtually always yields some design improvement.
- It is not biased by intuition or experience in engineering. Therefore, the possibility of obtaining improved, nontraditional designs is enhanced.
- Optimization requires a minimal amount of human-machine interaction.

1-6.2 Limitations of Numerical Optimization

- Computational time increases as the number of design variables increases. If one wishes to consider all possible design variables, the cost of automated design is often prohibitive. Also, as the number of design variables increases, these methods tend to become numerically ill-conditioned.
- Optimization techniques have no stored experience or intuition on which to draw. They are limited to the range of applicability of the analysis program.
- If the analysis program is not theoretically precise, the results of optimization may be misleading, and therefore the results should always be checked very carefully. Optimization will invariably take advantage of analysis errors in order to provide mathematical design improvements.

- Most optimization algorithms have difficulty in dealing with discontinuous functions. Also, highly nonlinear problems may converge slowly or not at all. This requires that we be particularly careful in formulating the automated design problem.
- It can seldom be guaranteed that the optimization algorithm will obtain the global optimum design. Therefore, it may be desirable to restart the optimization process from several different points to provide reasonable assurance of obtaining the global optimum.
- Because many analysis programs were not written with automated design in mind, adaptation of these programs to an optimization code may require significant reprogramming of the analysis routines.

1-6.3 Summary

Optimization techniques, if used effectively, can greatly reduce engineering design time and yield improved, efficient, and economical designs. However, it is important to understand the limitations of optimization techniques and use these methods as only one of many tools at our disposal.

Finally, it is important to recognize that, using numerical optimization techniques, the precise, absolute best design will seldom if ever be achieved. Expectations of achieving the absolute "best" design will invariably lead to "maximum" disappointment. We may better appreciate these techniques by replacing the word "optimization" with "design improvement," and recognize that a convenient method of improving designs is an extremely valuable tool.

REFERENCES

1. Schmit, L. A.: Structural Design by Systematic Synthesis, *Proceedings, 2nd Conference on Electronic Computation, ASCE,* New York, pp. 105 – 122, 1960.
2. Schmit, L. A.: Structural Synthesis – Its Genesis and Development, *AIAA J.,* vol. 10, no. 10, pp. 1249 – 1263, October 1981.
3. Zangwill, W. I.: "Nonlinear Programming: A Unified Approach," Prentice-Hall, Englewood Cliffs, N.J., 1969.

PROBLEMS

1-1 Consider the 500-N weight hanging by a cable, as shown in Figure 1-9. A horizontal force, $F = 100$ N, is applied to the weight. Under this force, the weight moves from its original position at A to a new equilibrium position at B. Ignore the cable weight. The equilibrium position is the one at which the total potential energy PE is a minimum, where PE = $WY - FX$.

a. Write an expression for PE in terms of the horizontal displacement X alone.

b. Write an expression for PE in terms of the angle θ alone.

c. Plot a graph of PE versus θ between $\theta = 0°$ and $\theta = 45°$.

d. Find the angle corresponding to the minimum value of PE both graphically and analytically. Prove that this is a minimum.

e. Using statics, verify that the θ at which PE is minimum is indeed the equilibrium position.

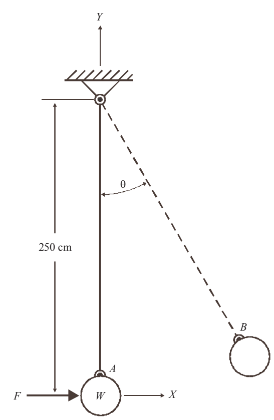

Figure 1-9

1-2 Given the unconstrained function

$$F = X_1 + \frac{1}{X_1} + X_2 + \frac{1}{X_2}$$

a. Calculate the gradient vector and the Hessian matrix.

b. At what combinations of X_1 and X_2 is the gradient equal to zero?

c. For each point identified in part *b,* is the function a minimum, a maximum, or neither?

1-3 Given the unconstrained function,

$$F = X_1^2 + \frac{1}{X_1} + X_2 + \frac{1}{X_2}$$

a. At $X_1 = 2$ and $X_2 = 2$, calculate the gradient of F.

b. At $X_1 = 2$ and $X_2 = 2$, calculate the direction of steepest descent.

c. Using the direction of steepest descent calculated in part *b,* update the design by the standard formula

$$\boldsymbol{X}^1 = \boldsymbol{X}^0 + \alpha \boldsymbol{S}^1$$

Evaluate X_1, X_2 and F for $\alpha = 0, 0.2, 0.5$, and 1.0 and plot the curve of F versus α.

d. Write the equation for F in terms of α alone. Discuss the character of this function.

e. From part *d,* calculate $dF/d\alpha$ at $\alpha = 0$.

f. Calculate the scalar product $\nabla \mathbf{F} \bullet \mathbf{S}$ using the results of parts *a* and *b* and compare this with the result of part *e.*

1-4 Consider the constrained minimization problem:

Minimize: $F = (X_1 - 1)^2 + (X_2 - 1)^2$

Subject to:

$$X_1 + X_2 \le 0.5$$

$$X_1 \ge 0.0$$

a. Sketch the two-variable function space showing contours of $F = 0$, 1, and 4 as well as the constraint boundaries.

b. Identify the unconstrained minimum of F on the figure.

c. Identify the constrained minimum on the figure.

d. At the constrained minimum, what are the Lagrange multipliers?

1-5 Given the ellipse $(X/2)^2 + Y^2 = 4$, it is desired to find the rectangle of greatest area which will fit inside the ellipse.

a. State this mathematically as a constrained minimization problem. That is, set up the problem for solution using numerical optimization.

b. Analytically determine the optimum dimensions of the rectangle and its corresponding area.

c. Draw the ellipse and the rectangle on the same figure.

1-6 Given the following optimization problem;

Minimize: $F = X_1 + X_2$

Subject to:

$$g_1 = 2 - X_1^2 - X_2 \le 0$$
$$g_2 = 4 - X_1 - 3X_2 \le 0$$
$$g_3 = -30 + X_1 + X_2^4 \le 0$$

The current design is $X_1 = 1$, $X_2 = 1$.

a. Does this design satisfy the Kuhn-Tucker necessary conditions for a constrained optimum? Explain.

b. What are the values of the Lagrange multipliers at this design point?

1-7 Given the following optimization problem:

Minimize: $F = 10 + X_1 + X_2$

Subject to:

$$g_1 = 5 - X_1 - 2X_2 \le 0$$
$$g_2 = \frac{1}{X_1} + \frac{1}{X_2} - 2 \le 0$$

a. Plot the two-variable function space showing contours of $F = 10$, 12, and 14 and the constraint boundaries $g_1 = 0$ and $g_2 = 0$.

b. Identify the feasible region.

c. Identify the optimum.