Norm Optimization Method in Structural Design

Moshe B. Fuchs*
Technion, Israel Institute of Technology, Haifa, Israel
and
Maurice A. Brull†
Tel-Aviv University, Ramat-Aviv, Israel

A general method for the minimization of a class of nondifferentiable merit functions is presented. The merit functions are defined as the maximum absolute value of the components of a vector of functions. These merit functions have gradient discontinuities in the design space and cannot be minimized by efficient algorithms of mathematical programming. The technique consists of sequential minimizations of an appropriate family of substitute merit functions, namely, the *p*th order norm of the vector. The efficiency of the technique is illustrated by the design of continuous beams for optimum geometry and is shown to give good results. It is further indicated that the method could be applied to general nonlinear inequality constrained mathematical programming problems and a few encouraging numerical examples are presented.

Nomenclature

a_i = cross section of truss member i
a_{pj} = attenuation factor
A = vector of components a_i
f_j = nonlinear function of X
F = vector of components f_j
g_j = constraint j
m_j = bending moment at location j
M = vector of components m_j
M_p = generalized mean of order p
N_p = norm of order p
N_∞ = uniform or infinite norm
r = tradeoff coefficient
S = search vector
t = step size; parameter
w = merit function
x_i = design variable i
X = vector of components x_i

I. Introduction

A BASIC problem of mathematical programming is the unconstrained minimization of a continuous nonlinear merit function $w(X)$ where X is an I-dimensional design vector of components x_i. However, for functions with discontinuous first derivatives in the design space most of the computationally efficient algorithms of the univariate type are inadequate (Ref. 1, pp. 40-51). Optimization methods in which the objective function is gradually reduced by successive one-dimensional minimizations along specified straight lines in the I-dimensional design space are called univariate type algorithms. A large class of such functions are those defined as the maximum component $f_j(X)$ of a J-dimensional vector $F(X)$

$$w(X) = \max_j [f_j(X)] \qquad j \epsilon J \tag{1}$$

where the $f_j(X)$ are differentiable functions of X in the design space. In effect, the classical univariate type algorithms generate at any given design stage $X^{(k)}$ a search vector $S^{(k)}$ along which a better design point $X^{(k+1)}$ is sought

$$X^{(k+1)} = X^{(k)} + t^{(k)} S^{(k)}$$

where k is the iteration index and t the step size. If we consider a function of the type in Eq. (1) in a two-dimensional design space (Fig. 1), we notice that the isocontours exhibit discontinuities in the gradient at the passage lines between one active component to another. Since all the design points exterior to the isocontour (shaded region) correspond to higher merit function values than the one on the contour, only those search vectors inside the angle formed by two adjacent isocomponent lines will have a descent direction. Failure to generate such a vector will induce the algorithm to conclude that it has reached a local minimum. This situation is fundamentally different from a differentiable merit function, for which all search vectors exhibit a descent direction, except at the minimum point.

This phenomenon of "jamming," or premature ending of the search at a gradient discontinuity, is even more critical for problems with more than two variables. In this case the descent directions are inside a hypercone and their number is negligible when compared to all the other possible search directions of the I-dimensional space. Thus, the discontinuity lines behave as traps for the univariate algorithms. The

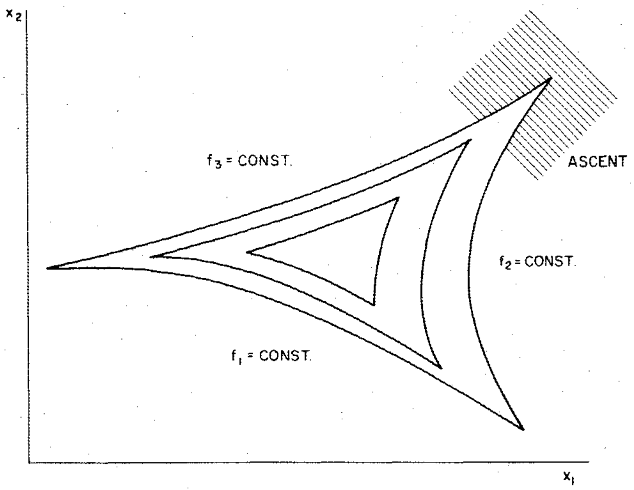

Fig. 1 Nondifferentiable merit function.

Received Jan. 25, 1977; revision received July 15, 1977.
Index category: Structural Design.
*Department of Aeronautical Engineering.
†Dean, School of Engineering, Tel-Aviv University.

present paper proposes a general method for solving this class of problems by systematically minimizing one or more substitute functions having continuous gradients, using classical univariate algorithms. This technique, the Norm Optimization Method (NOM), is shown to converge to the minimum of the original function. The development of the NOM was motivated by the many structural design formulations in which functions of the type in Eq. (1) occur, as will be emphasized in the following section.

The NOM is applied to the minimization of the maximum moment of a continuous beam where the design variables are the coordinates of a given set of simple supports and the method is shown to give excellent results.

In the last section it is indicated that the NOM could be used to solve general nonlinear optimization problems with inequality constraints. This is done by defining a merit function of the type in Eq. (1) whose minimum is identical to the minimum of the original constrained function. When used in this context, the method bears resemblance to typical exterior penalty functions of the SUMT,[2] although it differs in its intrinsic characteristics and solution method. As such, it suffers from the typical shortcoming of all the penalty function methods, that is, the proper a priori evaluation of some "tradeoff" coefficient between the constraints and the objective function. The method is tested on some simple classical inequality constrained structural optimization examples and is shown to give encouraging results. It should be noted that the NOM is a general mathematical programming technique that transforms optimization problems with nondifferentiable functions into equivalent problems involving differentiable functions. The iterative technique to yield the minimum will still require a substantial amount of function and constraint evaluations. It is therefore understood that if the NOM is to be applied to practical structural synthesis problems it will probably be used in conjunction with suitable approximate analysis techniques.[3]

The question of which analysis technique to use is, however, not central to the present paper, since our purpose is primarily to provide a means of overcoming the difficulties caused by gradient discontinuities.

II. Gradient Discontinuities in Structural Design

The structural design motivation underlying the development of the NOM arose during the study of continuous beams of optimum geometry. A continuous beam is a typical example of a structure composed of two distinct types of load-carrying components: a flexible bending member (the actual beam), and a series of rigid load-carrying components (the supports). The "beam" is determined by cross-sectional variables, whereas the support locations are given by geometric design variables. We assume that the beam is of uniform cross-sectional and material properties. Since the beam cross section is determined mainly by the value of the largest bending moment, in absolute value, the design problem is to minimize

$$\max_s \{ |m(X,s)| \} \qquad s \in S \tag{2}$$

where s is the beam coordinate, S the beam span, X an I-dimensional gometry design vector of components x_i (location of support i), and $m(X,s)$ the bending moment at s. If the beam is subjected to a series of concentrated forces, the maximum moment can occur either at a support location or under an applied load. The design problem is thus to minimize

$$\max_j \{ |m_j(X)| \} \qquad j \in J \tag{3}$$

where j identifies the supports and the applied loads and the m_j are the bending moments at the applied loads and at the support locations. This objective function, which is of the type in Eq. (1), has gradient discontinuities, as can be seen in the following one-dimensional design problem.

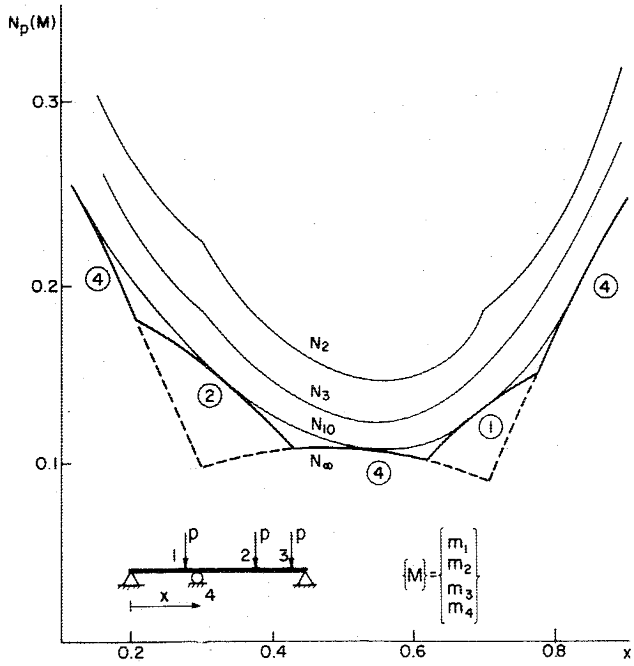

Fig. 2 Example of attenuation of gradient discontinuities.

A simply supported beam is subjected to three concentrated forces of equal magnitude at 0.3, 0.7, and 0.9 of the span (Fig. 2). It is required to locate an additional simple support in order to minimize the maximum bending moment anywhere on the beam. Let the m_j ($j \in 4$) be the bending moments at the applied forces and at the intermediate support. The problem is thus to minimize the objective function

$$\max_j \{ |m_j(x)| \} \qquad j \in 4$$

where x is the coordinate of the intermediate support. The merit function (heavy line in Fig. 2) is composed of differentiable branches, each corresponding to a controlling bending moment in a specific design domain. Slope discontinuities occur at the passages $m_4 - m_2 - m_4 - m_1 \ldots$. Similar gradient discontinuities appear in a multidimensional design problem at the passages from one active bending moment domain to an adjacent domain. As was shown in Sec. I, these gradient discontinuity lines can cause premature ending of classical univariate search algorithms due to a jamming phenomenon.

It is instructive to note that functions of the type in Eq. (1) appear in almost every structural optimization problem as behavioral constraints. In effect, optimum design of structures represented as an assemblage of discrete elements for which one stress is sufficient to describe the behavior of each element, such as axial force members or shear panels, present stress constraints of the type

$$\sigma_i(X) \leq \bar{\sigma}_i \tag{4}$$

where σ_i and $\bar{\sigma}_i$ are, respectively, the stress and allowable stress in element i. If the structure is subjected to a parametric loading $P(t)$, where P is a loading vector and t a parameter defined over a given interval T, the stress constraint (4) becomes a parametric constraint

$$\sigma_i(X,t) \leq \bar{\sigma}_i \qquad t \in T \tag{5}$$

If the structure is subjected to a discrete number J of loadings, the stress constraint can be considered as a "discrete parametric" constraint

$$\sigma_{ij}(X) \leq \bar{\sigma}_i \qquad j \in J \tag{6}$$

where σ_{ij} is the stress in element i due to the jth loading condition. The continuous set of constraints (5) can be replaced by a single constraint

$$\max_t \{\sigma_i(X,t)\} \leq \bar{\sigma}_i \qquad t \in T \tag{7}$$

and the discrete set of constraints (6) can be replaced by the single constraint

$$\max_j \{\sigma_{ij}(X)\} \leq \bar{\sigma}_i \qquad j \in J \tag{8}$$

Transformation (7) was used in the optimization of an aircraft wing using a continuous flutter constraint.[4] The reduction of a set of stress constraints of a given element for various loading conditions to a single constraint is penalized by the fact that constraint (8) has discontinuous gradients in the design space even if the set of constraints (6) is differentiable. The obvious reason for this is that the left-hand side of inequality (8) is an expression of the type in Eq. (1).

The NOM which was initially developed to deal with objective functions of Eq. (1) can also be generalized to deal with nondifferentiable constraints as in Eq. (8).

III. Norm of a Vector

We will recall in this section some properties of the norm of a vector[5] upon which the NOM is based. We have chosen to restate them because these aspects of the norm have not, to our knowledge, been used in structural optimization.

Uniform Norm

The p norm (norm of order p) of a J-dimensional vector X is given by

$$N_p(X) = \Big(\sum_j |x_j|^p\Big)^{1/p} \qquad p \geq 1 \qquad j \in J \tag{9}$$

where x_j are the components of the vector X. Note that for $p=2$ we obtain the classical or Euclidean norm. The uniform norm, also called maximum norm, is defined by

$$N_\infty(X) = \lim_{p \to \infty} N_p(X)$$

and it is easy to show that

$$N_\infty(X) = \max_j[|x_j|] \tag{10}$$

Jensen's Inequality

An inequality attributed to Jensen[6] formalizes the property that the norm is a monotone decreasing function of its order

$$N_s(X) < N_p(X) \qquad \text{for } 1 \leq p < s \tag{11}$$

unless all but one x_i are zero. The importance of this inequality is that it holds also in the limit as $s \to \infty$. Thus, Eq. (11) may be written as

$$N_\infty(X) = \max_j [|x_j|] < N_p(X) \tag{12}$$

that is, any p norm is an upper bound of the uniform norm.

Generalized Mean

The ordinary mean of order p of an n-dimensional vector $\bar{X}$ of components $|x_j|$ is given by

$$M_p(\bar{X}) = \Big(\frac{1}{n}\sum_j |x_j|^p\Big)^{1/p}$$

As in the case of the norm we define

$$M_\infty(\bar{X}) = \lim_{p \to \infty} M_p(\bar{X})$$

and we also have

$$M_\infty(\bar{X}) = N_\infty(X) = \max_j[|x_j|]$$

The equivalent of Jensen's inequality for the norm is in the case of the mean

$$M_p(\bar{X}) < M_s(\bar{X}) \qquad \text{for } p < s$$

and

$$M_p(\bar{X}) < M_\infty(\bar{X}) = N_\infty(X) \tag{13}$$

or any p mean of a vector $\bar{X}$ is a lower bound of the uniform norm of the corresponding vector X.

Upper and Lower Bounds

Combining inequalities (12) and (13) we obtain

$$\Big(\frac{1}{n}\Big)^{1/p} < \frac{N_\infty(X)}{N_p(X)} < 1 \tag{14}$$

Fig. 3 gives the variation of the ratio N_∞/N_p as a function of the order p of the norm and the dimension n of the vector. It shows how the uniform norm is bounded from above by a p-norm and from below by a p-mean.

IV. Minimum of Uniform Norm

Let $f_j(X)$ be the components of a J-dimensional vector function $F(X)$. We will assume that the components f_j are nonnegative and differentiable functions of X. It is required to determine the solution vector $X=X_\infty^*$ such that the maximum component of F is minimum. Mathematically the problem is stated as follows:

$$\text{Find } X=X_\infty^* \text{ such that } N_\infty[F(X)] \to \text{minimum} \tag{15}$$

This merit function has gradient discontinuities along the boundaries of active component zones. The proposed method consists of minimizing $N_p(F)$ instead of $N_\infty(F)$. The substitute problem is now

$$\text{Find } X=X_p^* \text{ such that } N_p[F(X)] \to \text{minimum} \tag{16}$$

Since N_p is an upper bound of N_∞, X_p^* is only an approximation of the solution vector X_∞^* of the original problem (15). But as the upper bound N_p tends to N_∞ as p is increased [Eq. (10)], the solution X_p^* will also tend to X_∞^*. The advantage of solving problem (16) in place of problem (15) is obvious. Since the components f_j are differentiable functions of X, it is clear from the definition of the norm [Eq. (9)] that $N_p(F)$ is also a differentiable function of X. Problem (16) can therefore be solved by an univariate type algorithm. Often approximation (16) is acceptable, but it can be improved, if necessary, by minimizing subsequently higher order norms

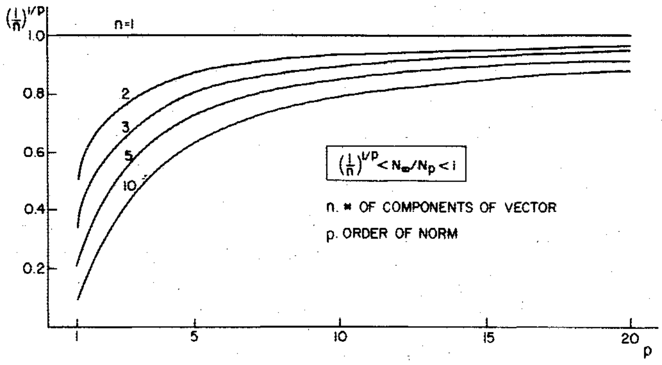

Fig. 3 Ratio N_∞/N_p as a function of the order p and the dimension n.

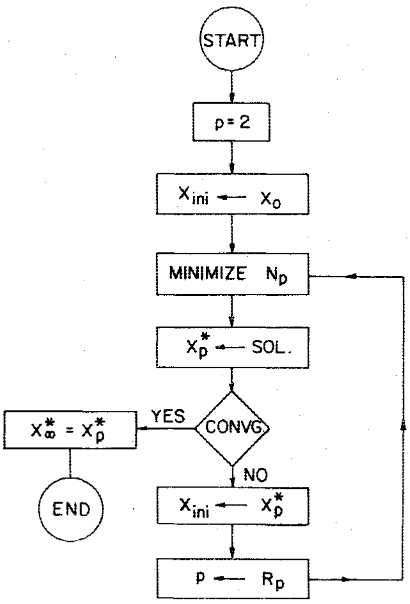

Fig. 4 Logical flow chart of the NOM.

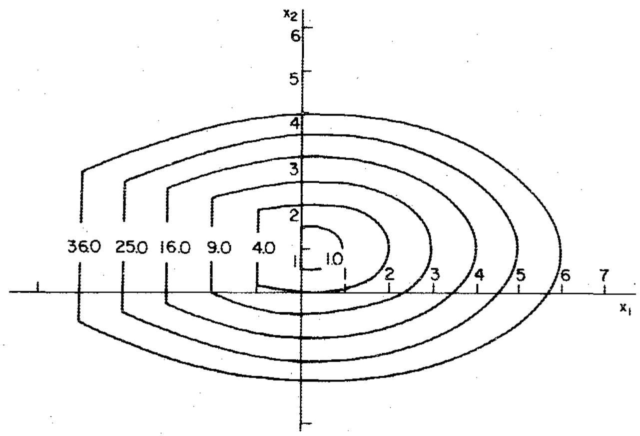

Fig. 5 Typical nondifferentiable merit function (Ref. 1, p. 41).

(similar to SUMT). We have found it good practice to minimize sequentially the series N_2, N_4, N_8, N_{16},..., where the solution of one iteration is used as the initial design of the next iteration (Fig. 4). The procedure is continued until convergence. The accuracy of the result can always be checked using Eq. (14). The concept of the NOM is best visualized by the following simple examples.

As an introductory case which emphasizes the power of the NOM, let us consider the problem given in Ref. 1 (p. 41) as typically ill-conditioned due to gradient discontinuities. Minimize $N_\infty(F)$, where $F(x_1,x_2)$ is a two-dimensional vector of components $f_1=(x_1-1)^2$ and $f_2=x_1^2+4(x_2-1)^2$. The minimum is $N_\infty^*=0.25$ for $x_1^*=0.5$ and $x_2^*=1.0$ (Fig. 5). In this particular case we know that the design X_1^* that minimizes $N_1(F)$ is already identical to the design X_∞^* that minimizes $N_\infty(F)$. The solution can be obtained closed form by solving the equations $\partial N_1/\partial x_1=\partial N_1/\partial x_2=0$, which reduce to a system of two linear equations in x_1 and x_2.

The sequential minimizations of N_p for increasing values of the order p are illustrated in the two following problems.

One Dimensional Case

Consider the minimization of the function $N_\infty(F)$ where F is a two-dimensional vector of components: $f_1=x$ and $f_2=2(1-x)$, the minimum of which is at the gradient discontinuity point $x=2/3$. Such a problem is to be solved by linear programming, but it is instructive to see how it can be transformed into the minimization of N_p. In Fig. 6 we show how the broken line $N_\infty(x)$ is approximated by smooth curves $N_p(x)$ for values of p varying between 1 and 20. One has to bear in mind that we seek mainly the value x_∞^* which minimizes N_∞, and we thus have to compare the minima x_p^* of the norms N_p with the final solution x_∞^*. As can be expected, the error $e_p=|x_\infty^*-x_p^*|$ between the original problem N_∞ and the transformed problem N_p tends to zero as p increases. One is inclined to pick higher values of p in order to reduce e_p. The penalty in doing so is that the functions N_p become more and more difficult to optimize as p is increased, since they approach the ill-conditioned function N_∞. There is therefore a tradeoff to be performed between low order norms (easy to optimize but for which e_p can still be substantial) and higher order norms (more difficult to optimize but small e_p).

Two Dimensional Case

Minimize $N_\infty(F)$ where F is a two-dimensional vector of components $f_1=x_1^2+x_2^2$, $f_2=(x_1-1)^2+x_2^2$. The isocontours of the function N_∞ are formed by two circular arcs that intersect along the line $x_1=0.5$ (Fig. 7). Because of symmetry, any norm N_p will have its minimum at X^*. The example is nevertheless instructive, for it shows how the two-dimensional broken isocontours of N_∞ are approximated by the smooth isocontours of N_p. The advantage of replacing N_∞ by N_2 or N_5 in order to be optimized by univariate algorithms is self-evident. Here again we note that N_{20}, for instance, coincides with N_∞ except for the sharp corners about $x_1=0.5$. We have experienced this feature in all the problems solved and as a rule it was never necessary to go beyond N_{32}.

V. Nondifferentiable Components

We have seen previously that if the components of the vector function F are differentiable functions of X, then $N_p(F)$ is also differentiable. It may occur, however, that some components have discontinuous gradients in the design space, and it is instructive to see how this affects the method. The derivative of $N_p(F)$ with respect to variable x_i is

$$\frac{\partial N_p}{\partial x_i}=\sum_j\left(\frac{f_j}{N_p}\right)^{p-1}\frac{\partial f_j}{\partial x_i} \quad i\epsilon I,\ j\epsilon J \tag{17}$$

Equation (17) can be rewritten

$$\frac{\partial N_p}{\partial x_i}=\sum_j a_{pj}\frac{\partial f_j}{\partial x_i} \quad i\epsilon I,\ j\epsilon J \tag{18}$$

where a_{pj} is an attenuation factor

$$a_{pj}=\left(\frac{f_j}{N_p}\right)^{p-1}\leq 1,\quad p>1$$

If some partial derivative $\partial f_j/\partial x_i$ has a finite discontinuity in the design space, its influence on the derivative of the norm is attenuated by the coefficient a_{pj}, that is, the partial derivative $\partial N_p/\partial x_i$ is "less" discontinuous. If the discontinuity occurs in a noncritical component ($f_j<f$max) it vanishes for the norm N_p as p is increased, since a_{pj} tends to zero as $p\rightarrow\infty$. In two instances only does the attenuation factor not reduce the norm derivative discontinuity:

1) The discontinuity occurs in the governing component and at the minimum of N_∞. In this case a_{pj} tends to one as p is increased. But such a coincidence is not to be expected very often in practical problems.

2) $p=1$: In this case $a_{pj}=1$ and, as a rule, N_1 has to be avoided when nondifferentiable components are present.

The beneficial influence of the attenuation factor a_{pj} is visualized in the beam problem of Fig. 2. We have shown the norms $N_p(M)$, $p=2,3,10$, and ∞, as a function of the design variable x. M is the vector of components m_j which are the moments at the applied leads ($j=1,2,3$) and the moment at the design support ($j=4$). We notice that the component m_4

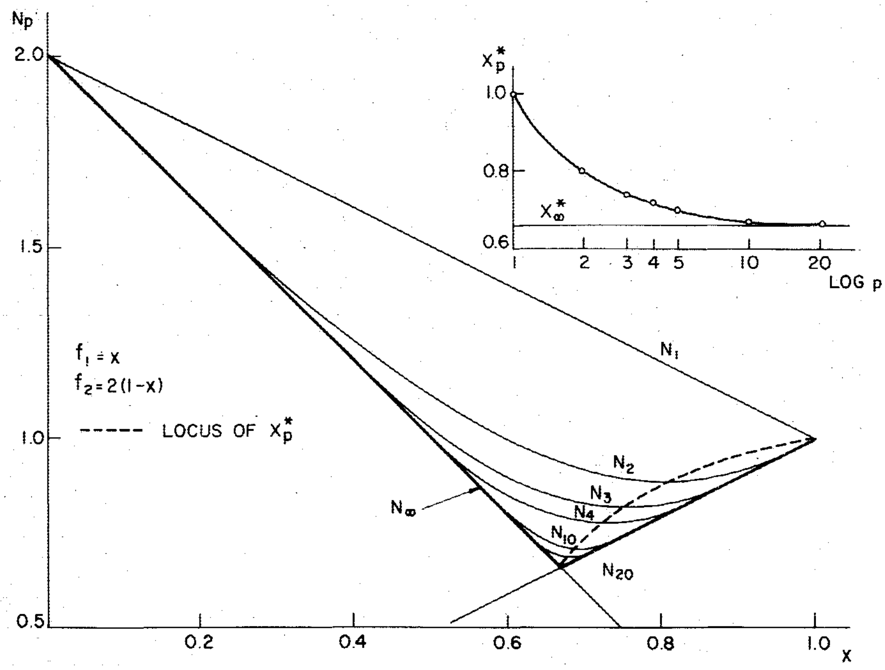

Fig. 6 Approximation of N_∞ by N_p in a one-dimensional example.

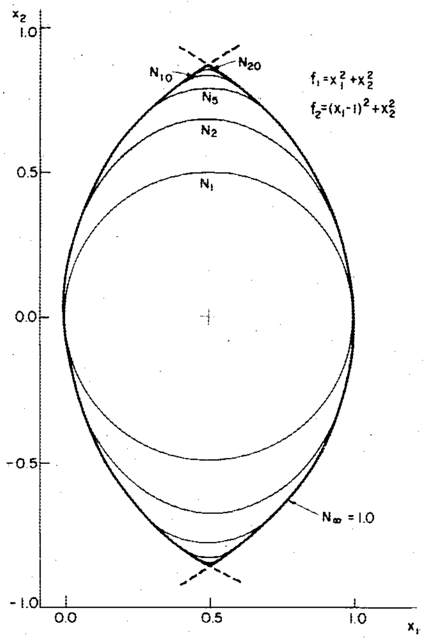

Fig. 7 Approximation of N_∞ by N_p in a two-dimensional example.

(dashed line) has slope discontinuities at $x=0.3$ and $x=0.7$ (locations of the first two forces). These discontinuities are already reduced substantially for the norm N_2 and have almost disappeared from the norm N_3. On the other hand the slope discontinuities of N_∞ which are due to the passage from one active component to another do not appear in the norms of finite order for the reasons that were previously outlined.

VI. Continuous Beams of Optimum Geometry

A few examples will illustrate the use of the NOM as applied to minimization of the maximum moment of a continuous beam where the variables are the coordinates of a preselected set of simple supports. The loading is defined by concentrated forces, and the support locations are the design variables. The maximum moment occurs either at a design support or at an external load location. The procedure consists in generating a moment vector M whose components are the moments at the supports and at the external loads. The norm $N_p(M)$ is minimized sequentially, starting with the order $p=2$. The solution vector of this minimization is used as the initial design for the optimization of N_4, and so on for the norm sequence given in Fig. 4 with $R=2$.[7]

We note that, for the specific case of the maximum moment problem, the major improvement is obtained by the minimization of N_2, while only minor decreases are achieved when using higher order norms.

Two Design Supports

Consider a simply supported beam of uniform cross section and material properties. The locations of two additional simple supports are given by the components x_1 and x_2 of the geometry design vector X. The dimensions of the beam cross section are governed mainly by the value of the largest bending moment (in absolute value) at all stations, and the optimization problem can be stated as:

$$\text{Find } X=X^*_\infty \text{ such that } N_\infty[M(X)] \rightarrow \text{minimum}$$

$$\text{subject to } 0 \le x_1 \le x_2 \le 1$$

where M is a 5-component vector containing the bending moments at the design supports and at the applied loads (Fig. 8). The inequality constraints are introduced in order to eliminate multiple, structurally identical minima. The isocontours of N_∞ are given in Fig. 9. The feasible domain is partitioned into regions in which a specific component of M is active. The optimum design is the point where m_1, m_2, and m_4 are simultaneously critical. Along the boundaries of the active zones the isocontours present sharp gradient discontinuities. Passages 1-3 and 2-5 are especially vulnerable, since along them most of the search directions are ascent (shaded area). The minimization technique is a regular conjugate-direction algorithm in conjunction with a quadratic interpolation for the univariate search. Starting from the initial design ($x_1=0.7$, $x_2=0.9$, $N_\infty=0.193$), the iterations terminated at a discontinuity line ($x_1=0.478$, $x_2=0.897$, $N_\infty=0.078$) since the algorithm did not generate a search vector outside the shaded region. With the use of the NOM the optimum was obtained by minimizing $N_2(M)$ instead of the original function ($x_1^*=0.315$, $x_2^*=0.685$, $N_2^*=0.083$, $N_\infty^*=0.046$). In this particular case the design was not improved by subsequent minimization of higher order norms. As was previously indicated, this type of problem has

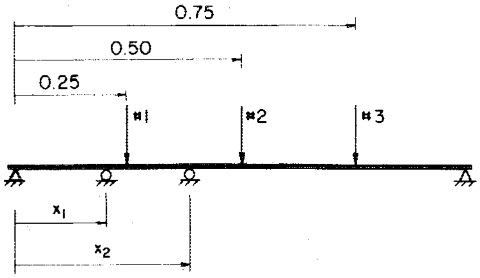

Fig. 8 Design of a continuous beam.

discontinuous component derivatives. The minimization of $N_1(M)$ was stopped even earlier than N_∞ due to a gradient discontinuity of a noncritical component of M.

Three-Design Supports

The beam geometry is defined by two end supports and three intermediate design supports whose coordinates are the components x_1, x_2, and x_3 of the design vector X. Two different loadings are considered: four equidistant forces of equal magnitude (Fig. 10), and four equidistant forces of linearly varying magnitude (Fig. 11). The successive objective functions $N_p(M)$, where the m_i are the bending moments at the applied loads and at the moving supports ($J=7$), were minimized by a conjugate-directions-type algorithm. As will be noticed, the minimization of N_2 reduces the maximum moment almost to its optimum value, while only minor decreases result from subsequent minimizations of higher order norms. For comparison we give the value of the "end value" of N_∞, when this function was minimized from the same initial design point using the same algorithm. The direct minimization of N_∞ gave a nonoptimum result, since the program was jammed at a gradient discontinuity line.

VII. NOM in Constrained Optimization

In this section we indicate a procedure by which the NOM can be used to solve the general nonlinear optimization problem with inequality constraints

$$\text{minimize } w(X)$$

$$\text{subject to } g_j(X) \leq 0 \quad j \in J \tag{19}$$

where X is a vector whose components are the design variables $x_i (i \in I)$. The method has been tested on some simple structural optimization problems and has given interesting although not conclusive results. It is nevertheless of interest to present the method even in the early stage of its development.

The basic approach in the use of the NOM in constrained optimization is to replace the original problem (19) by a transformed problem

$$\text{minimize } N_\infty[F(X)] \tag{20}$$

where $F(X)$ is some vector function. This transformed problem is then solved by the NOM. Obviously, the success of the method depends on the judicious choice of the functions $F(X)$ such that $N_\infty(F)$ and problem (19) have the same minimum.

Consider, for instance, the simple case

$$\text{minimize} \quad w(x) = x^2$$

$$\text{subject to} \quad g(x) = 1 - x \leq 0 \tag{21}$$

the solution of which is $x=1$ (Fig. 12a). In Fig. 12b we show the merit function w and a series of curves

$$h(x,r) = w(x) + rg(x) \quad r>0$$

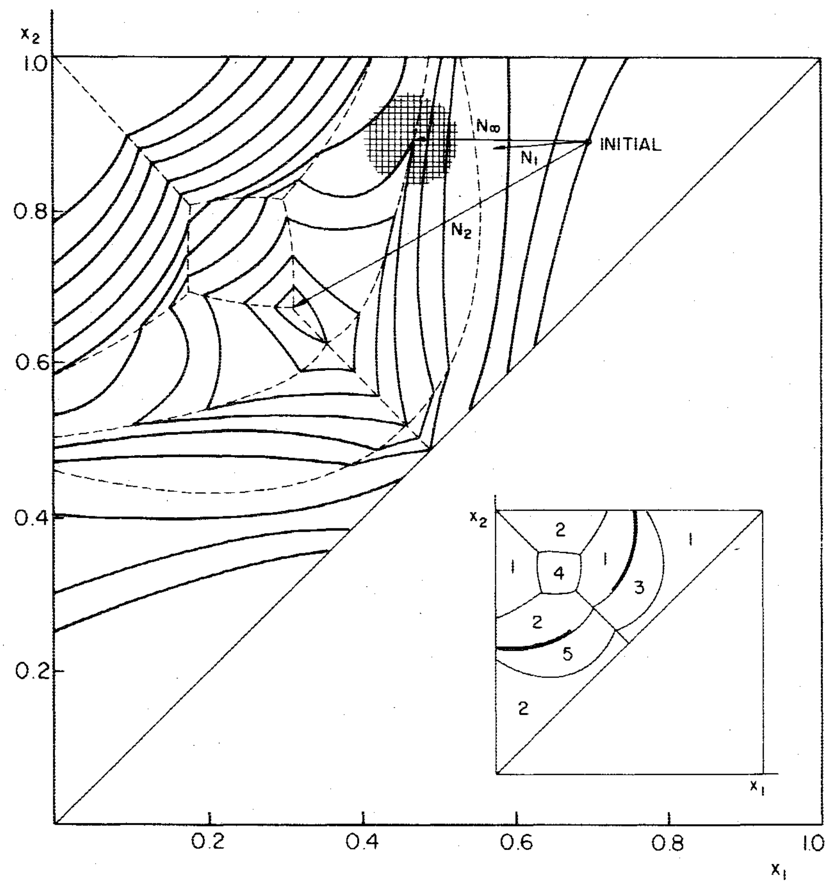

Fig. 9 Design space of maximum moment problem.

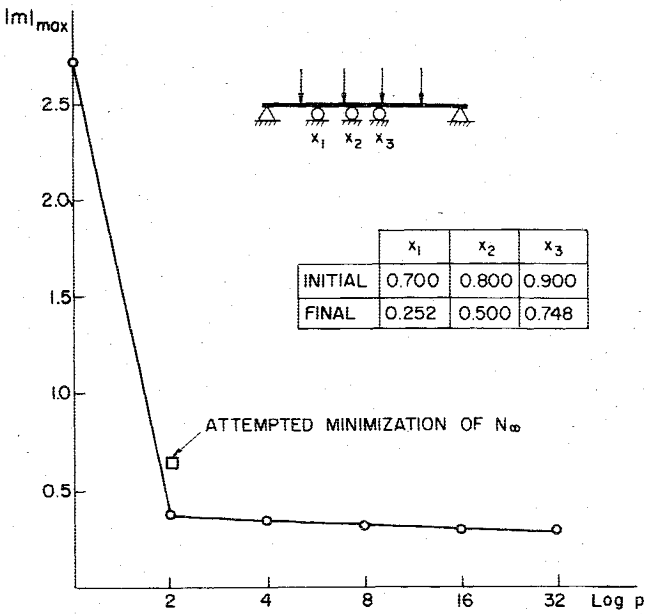

Fig. 10 Uniform loading; convergence of the NOM.

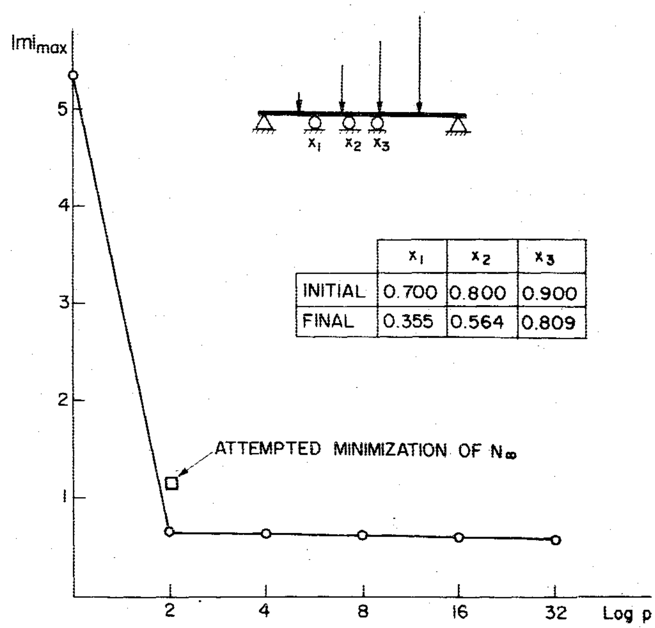

Fig. 11 Triangular loading; convergence of the NOM.

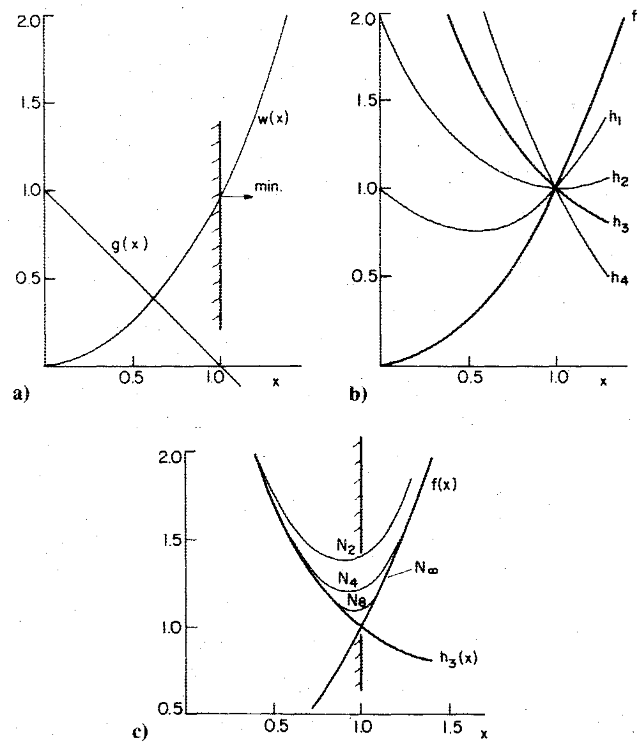

Fig. 12 The NOM in nonlinear constrained optimization.

We notice that, starting from $r=3$ and onwards, the function $N_\infty(F)$, where F is a vector of components

$$f_1 = h(x,r) \quad \text{and} \quad f_2 = w(x)$$

has the same minimum as problem (21). If we chose to minimize N_∞ for $r=3$ we obtain the function shown in Fig. 12c and its successive approximations $N_2, N_4, N_8, \ldots$.

In the general case one generates $(J+1)$-dimensional vector F of components

$$f_j(X) = \langle w(X) + rg_j(X) \rangle \quad j \epsilon J$$

$$f_{J+1}(X) = w(X) \tag{22}$$

where

$\langle\ \rangle$ is the bracket operator

$\langle a \rangle = a$ for $a \geq 0$

$\langle a \rangle = 0$ for $a < 0$.

It is assumed that $w(X)$ is positive in the feasible domain.

A correct choice of the coefficient r is naturally of cardinal importance, as can be seen in Table 1. If r is taken too small (case 1 of Table 1) the two problems (20) and (21) do not have the same solution. Moreover, the minimum of N_∞ lies in the nonfeasible region. This would correspond to a value $r=1$ for the example of Fig. 12. For increasing values of r (case 2) the minima of problems (20) and (21) coincide but the intermediate minima of lower order norms N_p are nonfeasible, so that the iterations cannot be interrupted at an early stage for significant results. This corresponds to a value $r=3$ of Fig. 12 for which the minima of N_2, N_4, and N_8 are in the nonfeasible region $x<1$. In this case the norm behaves as an exterior penalty function. Still higher values of r (case 3) would give "correct" estimates such that the successive minima of N_p are feasible and the process can be interrupted at will. This is similar to an interior penalty function approach. On the other hand, if r is too large (case 4), the functions N_p may become ill-conditioned, requiring a prohibitive number of function evaluations to yield their minima.

It should be noted that the use of the NOM for a constrained optimization problem bears some similarity to the well-known penalty function approach.[2] With the NOM one optimizes sequentially a series of functions

$$N^p(X,p,r) = w^p(X) + \sum_j \langle w(X) + rg_j(X) \rangle^p$$

as compared to a typical exterior penalty function (Ref. 1, p. 127).

$$P(X,r,p) = w(X) + r\sum_j \langle g_j(X) \rangle^p$$

Table 1 Characteristic values of r

Case	r	Diagnosis
1	Too small	N_∞^* in nonfeasible region
2	Small	N_p^* in nonfeasible region
3	Correct	N_p^* in feasible region
4	Too large	N_p ill-conditioned

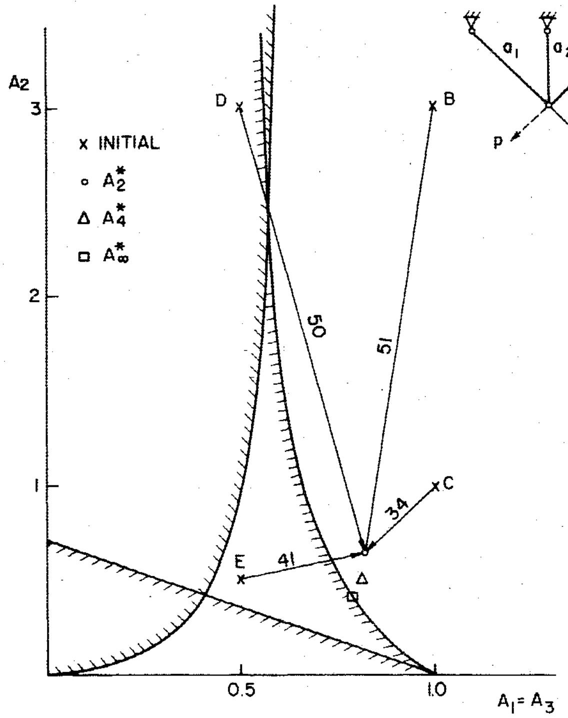

Fig. 13 Three-bar truss minimization by the NOM.

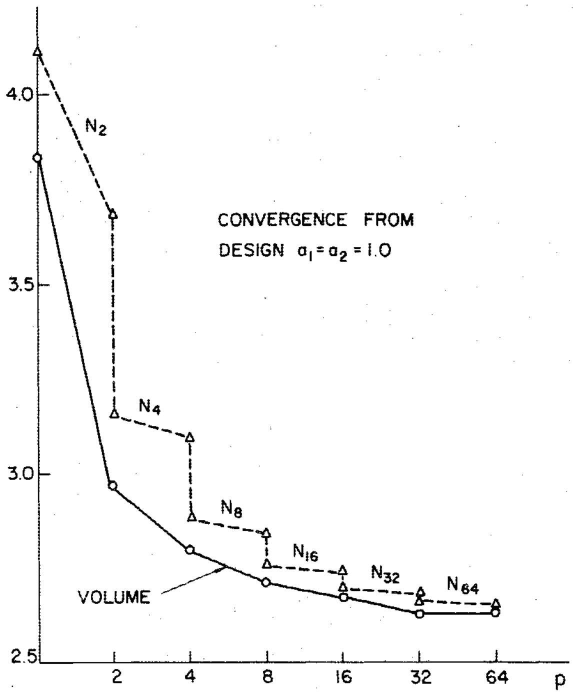

Fig. 14 Convergence of three-bar truss problem.

Except for the definition of the penalty term, the main difference between the NOM and the classical penalty function method is that in the former the coefficient r is constant and the value of the exponent p is sequentially increased, whereas in the latter the value of the coefficient r is increased, the exponent p being kept constant.

As such, the NOM is subjected to the basic difficulty common to all penalty function methods, that is, the correct estimate of the tradeoff coefficient r. Further investigations are therefore required if the NOM is to become competitive with established methods of constrained optimization.

It has become a tradition in structural optimization to test a new method of mathematical programming on the three-bar truss.[8] A symmetric truss of given material properties is subjected to two symmetric loading conditions (Fig. 13). The design variables are the cross sections of the exterior members a_l and the cross section of the interior member a_2. The problem is

$$\text{Find } A = A^* \text{ such that } w(A) \rightarrow \text{minimum}$$

$$\text{subject to} \quad \sigma_l \leq \sigma_i \leq \sigma_u \qquad i = 1,2,3$$

$$a_j \geq 0 \qquad j = 1,2 \tag{23}$$

where w is the volume of the truss, A the design vector of components a_l and a_2, σ_i the stress in member i, and σ_l and σ_u the compressive and tensile allowable stresses of the material. The numerical constants are $P = 20$ kips, $h = 100$ in., $\sigma_l = -15$ ksi and $\sigma_u = 20$ ksi. The stress inequalities were multiplied by the factor 1.0 E-4 in order for all the constraints to be of the same order of magnitude. Problem (23) is replaced by problem (20) and Eq. (22) for $r = 4$. In Fig. 13 we show the locations of the minimum of N_2, N_4, and N_∞. The minimization of N_2 was performed from four different initial designs, two feasible points (B and C) and two nonfeasible points (D and E). The number of function evaluations required to reach A_2^* appears on the arrow. This relatively large number of function evaluations is probably to be attributed to poor termination criteria of the unconstrained minimization routine which was used. It is instructive to notice that the initial design does not have to be feasible, and that the first iteration already generates a feasible intermediate solution. The following iteration minimizes N_4 starting from A_2^* and yields the next approximation A_4^*. Fig. 14 gives the convergence to the final solution due to the successive minimizations, and detailed results are shown in Table 2. It should be remembered that the minimization is performed on N_p and the decrease of the volume is only an additional result of the routine. For instance, at $p = 4$, N_p was reduced from 3.16 to 3.10, which did result in a volume reduction from 2.97 to 2.80. For all practical purposes, the minimum volume $w = 2.64$ was obtained after the minimization of N_{32}.

An alternative classical problem of structural synthesis is the design for minimum volume of a two-bar truss under stress and Euler buckling constraints.[1] The design variables are the mean diameter d of the tubular cross sections and the height h of the structure. In Fig. 15 we compare the convergence of the exterior penalty function method used in Ref. 1 with a solution obtained using the NOM approach. The numerical constants are: Young's modulus $E = 3.0$ E4 ksi, maximum allowable compressive stress $\sigma_y = 100$ ksi, tube thickness $t = 0.1$ in., $b = 30$ in., and $F = 33$ kips. The initial design is $d = 6$ in. and $h = 30$ in. As a preparatory step, the constraints are multiplied by a scaling factor such that the $g_i(X)$ are of the order of the initial volume of the truss.

In this case we took $r = 10$ and the NOM iterations were all within the feasible region, as compared to the penalty function results which converged to the minimum through the nonfeasible domain.

These two examples illustrating an extension of the NOM are obviously not sufficient to be conclusive. Nevertheless

Table 2 Three-bar truss; numerical results

$k = 4$	Initial			Final			
p	a_l	a_2	w	a_l	a_2	w	N_p
2	1.000	1.000	3.83	0.821	0.650	2.97	3.68
4	0.821	0.650	2.97	0.819	0.479	2.80	3.10
8	0.819	0.479	2.80	0.811	0.421	2.72	2.86
16	0.811	0.421	2.72	0.799	0.417	2.68	2.74
32	0.799	0.417	2.68	0.793	0.416	2.64	2.69
64	0.793	0.416	2.64	0.790	0.413	2.64	2.67

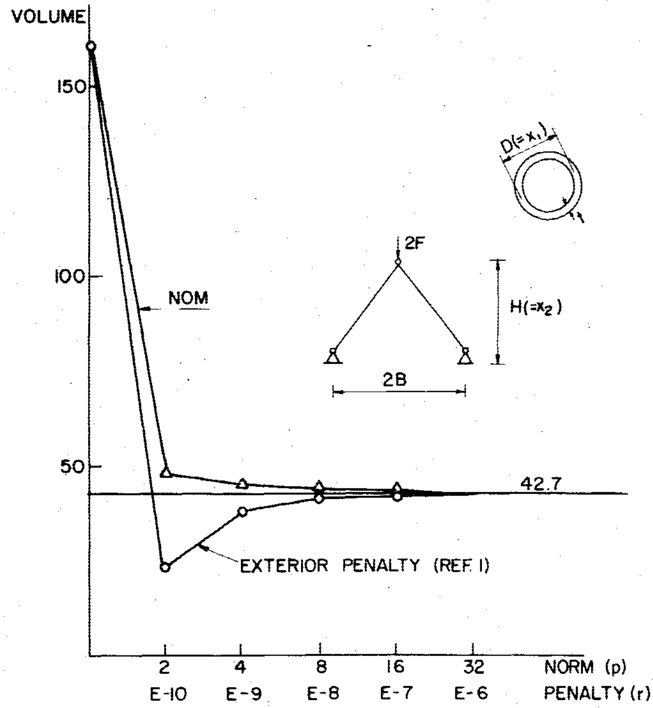

Fig. 15 Comparison of the NOM with an exterior penalty function.

they indicate a possible additional use of the NOM in nonlinear constrained optimization problems. When applied to practical structural optimization problems it will naturally have to be used in conjunction with approximate analysis techniques,[3] since the algorithms that minimize N_p still require a prohibitive number of function and constraints evaluations.

VIII. Conclusions

We have presented the NOM as a general technique to minimize unconstrained merit functions that have gradient discontinuities in the design space of the type $N_\infty(F)$. We have also shown how the NOM can be used to solve general optimization problems with inequality constraints. In this case, the method bears resemblance to the exterior penalty function method, but it differs mainly by the way the transformed functions are varied at every iteration step.

For the solution of the first class of problems the NOM was shown to be very powerful, and, in fact, circumvents the difficulties presented by gradient discontinuities. As for the second class of problems, it is, in our opinion, too early to give the method an absolute rating or a decisive appraisal when compared to the other available transformation methods. However, for the constrained problems that were solved, it gave highly satisfactory results.

Future investigations should concentrate on methods to define proper weighting factors r and also on the use of the NOM in conjunction with appproximate analysis techniques for structural synthesis problems.

Acknowledgement

This paper is based upon a thesis submitted by M. B. Fuchs in partial fulfillment for a Sc.D. at the Technion, I.I.T. under the supervision of A. Betser, A. Libai, and M. A. Brull.

References

[1] Fox, R. L., *Optimization Methods in Engineering Design*, Addison-Wesley, Reading, Mass., 1971.

[2] Fiacco, A. V. and McCormick, G. P., *Nonlinear Programming, Sequential Unconstrained Minimization Techniques*, Wiley, N.Y., 1968.

[3] Schmit, L. A. Jr. and Farshi, B., "Some Approximation Concepts for Structural Synthesis," *AIAA Journal*, Vol. 12, May 1974, pp. 692-699.

[4] Haftka, R. T., "Parametric Constraints with Application to Optimization for Flutter Using a Continuous Flutter Constraint," *AIAA Journal*, Vol. 13, April 1975, pp. 471-475.

[5] Hardy, G. H., Littlewood, J. E., and Polya, G., *Inequalities*, Cambridge University Press, 1934.

[6] Jensen, J.L.W.V., "Sur les Fonctions Convexes et les Inegalites entre les Valeurs Moyennes," *Acta Mathematica*, Vol. 30, 1906.

[7] Fuchs, M., "Substitute Function Methods in Structural Optimization and their Application to Continuous Beams," Sc.D. Thesis, Dept. of Aeronautical Engineering, Technion I.I.T., Haifa, Israel, May 1976.

[8] Schmit, L. A. Jr., "Structural Design by Systematic Synthesis," *Proceedings of the Second National Conference on Electronic Computation*, Structural Division, ASCE, Pittsburgh, Pa., Sept. 1960.