Miller, C. A., "Stability of Moving Surfaces in Fluid Systems with Heat and Mass Transport. II. Combined Effects of Transport and Density Difference Between Phases," *AIChE J.*, **19**, 909 (1973).
———., and K. Jain, "Stability of Moving Surfaces in Fluid Systems With Heat and Mass Transport. I. Stability in the Absence of Surface Tension Effects," *Chem. Eng. Sci.*, **28**, 157 (1973).
Saffman, P. G., and G. I. Taylor, "The Penetration of a Fluid into a Porous Medium or Hele-Shaw Cell Containing a More Viscous Liquid," *Proc. Roy. Soc.*, (London), **A245**, 312 (1958).
Sekerka, R. F., "Morphological Stability," *J. Cryst. Growth*, 3-4, 71 (1968).
Shutler, N. D., "Numerical Three-Phase Simulation of the Linear Steam Flood Process," *Soc. Pet. Eng. J.*, 9, 232 (1969).

Manuscript received August 14, 1974; revision received and accepted December 11, 1974.

A New Method for Nonlinearly Constrained Optimization

J. S. NEWELL
and
D. M. HIMMELBLAU
Department of Chemical Engineering
University of Texas
Austin, Texas 78712

A new penalty function method of solving problems involving a nonlinear objective function subject to nonlinear equality and inequality constraints is described. It ameliorates difficulties experienced with the ill-conditioning of the Hessian matrix of classical penalty function methods. Experience based on the solution of 25 test problems indicates the proposed method is as good as, or better, than methods that are now being used.

SCOPE

Chemical engineers frequently encounter realistic optimization problems that involve maximization or minimization of a cost, revenue, or other function subject to certain constraints. The constraints may be either equations, such as material or energy balances, or inequalities, such as capacity limits, or both. Often the objective function and constraints are nonlinear. A surprising number of practical chemical engineering problems have the characteristics just listed, such as the determination of steady state plant operating conditions, scheduling problems, capacity expansion analysis, some types of model building, and constrained regression. Furthermore, one of the most important steps in chemical engineering process design is the specification of the best values of the design variables according to some criterion.

Although the available techniques of optimization via numerical methods have multiplied exponentially in the last ten years with the advent of large-scale computers, most of the techniques prove to be ineffective when applied to solvable practical problems of the type outlined above. We describe here a method of minimizing a nonlinear objective function subject to nonlinear constraints that is efficient, robust, user oriented, and based on sound theoretical principles. The proposed technique is termed the General Augmented Penalty Function (GAPF) method in as much as it is one of a class of nonlinear optimization methods that employ penalties to accommodate the constraints in the problem.

CONCLUSIONS AND SIGNIFICANCE

Although numerous algorithms have been proposed to solve constrained nonlinear optimization problems, surprisingly little is known about their relative merits. It is difficult for an engineer to isolate methods that are best suited for his problems. It has been demonstrated both theoretically and by numerical tests that the GAPF method in most instances alleviates the traditional numerical problems of penalty function methods while retaining their attractive features, such as the use of unconstrained minimization and the absence of a requirement to maintain feasibility as the optimization proceeds.

To demonstrate the performance of the proposed method relative to other algorithms that have been reported in the literature to be effective, the GAPF method was implemented in a computer code and tested on 25 problems of which 15 involved nonlinear constraints. One-half of the problems have their origin in the field of chemical engineering. It was found that the GAPF Code was both robust and quick, and outperformed the traditional penalty function methods. For the problems involving nonlinear equality constraints, the method was distinctly superior to all other methods. Data to support these conclusions is compiled. Because the computer code is easy to execute from the user's viewpoint and because the method essentially requires no arbitrary parameters to be selected by the user, it can be recommended for use by those who have only minimal experience with optimization techniques and "need to get the job done."

Nonlinear programming is the name given to the class of numerical and analytic techniques that solve the problem

minimize $f(x)$ $\quad x \epsilon E^n$

subject to the equality constraints

$$h_i(x) = 0 \quad i = 1, \ldots\ldots, m \tag{1}$$

and the inequality constraints

$$g_i(x) \geq 0 \quad i = m+1, \ldots\ldots, p$$

In Problem (1) the objective function $f(x)$ can be linear or nonlinear; the equality constraints $h_i(x)$ can be linear or nonlinear; and the inequality constraints $g(x)$ can be linear or nonlinear. The objective function can be viewed as a performance index, cost function, response surface, or revenue function, while the constraints can be physical limitations, correlations, mass and energy balances, etc. Problem (1) is often encountered by chemical engineers because a wide variety of practical problems can be represented in the form of Problem (1). Examples include the optimal allocation of resources and distribution of products, the optimal operating policy of a manufacturing process, the construction and operating design of industrial units at minimum cost, and the determination of chemical and thermodynamic equilibrium. Excluded from consideration here are two special cases of Problem (1), namely those problems

1. which involve integer variables and
2. which include differential or integral equations as constraints because both cases require specialized solution techniques.

Methods for solving the Problem (1) usually fall into one of two categories: (a) linearization methods or (b) penalty function methods as described in Himmelblau (1972). Linearization methods in general are more efficient if the constraints are linear or near linear whereas penalty function methods are relatively more efficient when the constraints are highly nonlinear. We are concerned here mainly with the latter type of problems in contrast to the work of Goldfarb (1969), Murtagh and Sargent (1969), Fletcher (1972), and Gill and Murray (1973), and Husen and Eakman (1974) who proposed algorithms to solve linearly constrained problems. However, the method described below is perfectly satisfactory for linearly constrained problems although not quite as fast in execution on a computer.

One widely used technique to solve the general nonlinear programming problem is to employ penalty functions. The essential idea of a penalty function is to transform Problem (1) into a problem in which a single unconstrained function is minimized

$$\left.\begin{array}{l}\text{minimize } f(x) \\ \text{s. t. } g(x) \geq 0 \\ \quad\quad h(x) = 0\end{array}\right\} \text{minimize } P(f, g, h) \tag{1'}$$

Usually a penalty function is minimized sequentially for a series of parameters whose values are altered between unconstrained minimizations so as to force the sequence of x vectors minimizing the penalty function to approach the constrained optimum of Problem (1). Thus, the strategy of penalty function methods converts Problem (1) from a constrained problem to a systematic series of minimizations of unconstrained functions of the same form but with different parameters.

The idea of approximating a constrained optimization problem by an unconstrained problem through the use of a penalty function is a fairly old one. It has been systematically employed in numerical optimization for approximately 15 years and was popularized mainly through the work of Fiacco and McCormick (1966, 1968). Penalty function methods have been widely accepted in practice despite criticism directed at their slow convergence properties and numerical instabilities primarily because they are general and simple relative to other methods for constrained minimization plus the fact that they make full use of the very efficient unconstrained minimization procedures which have been developed in recent years. A general parametric penalty function takes the form

$$P(x, r^{(k)}) = f(x) + \sum_{i=1}^{m} (r^{(k)})^{\alpha_i} H(h_i(x)) + \sum_{i=m+1}^{p} (r^{(k)})^{\alpha_i} G(g_i(x)) \tag{2}$$

where $(r_i^{(k)})^{\alpha_i}$ is a scalar parameter or weighting factor associated with the ith constraint on the kth unconstrained minimization of (2) with the property

$$0 < r_i^{(0)} < r_i^{(1)} < \ldots < r_i^{(k)} < \ldots \tag{3}$$

α_i is an exponent of r_i usually chosen as $+1$ or -1, and H and G are functions of $h_i(x)$ and $g_i(x)$, respectively.

Designating $\hat{x}^{(k)}$ as the vector minimizing (2) on the kth stage, and x^* as the solution of Problem (1), the following properties of parametric penalty functions are desired:

$$\lim_{k \to \infty} P(\hat{x}^{(k)}, r^{(k)}) \to f(x^*) \tag{4}$$

$$\lim_{k \to \infty} (r_i^{(k)})^{\alpha_i} H(h_i(\hat{x}^{(k)}) \to 0 \tag{5}$$

$$\lim_{k \to \infty} (r_i^{(k)})^{\alpha_i} G(g_i(\hat{x}^{(k)}) \to 0 \tag{6}$$

$$\lim_{k \to \infty} r_i^{(k)} \to \infty \tag{7}$$

Usually $H(h_i(x))$ is a quadratic loss penalty $h_i^2(x)$ but $G(g_i(x))$ can be either a barrier function, such as $1/g_i(x)$, or a penalty function.

DIFFICULTIES WITH PENALTY FUNCTIONS

In the past few years Murray (1967), Lootsma (1970), and others have investigated the problem of illconditioning associated with traditional penalty function methods. It has been demonstrated that as the constrained minimum of Problem (1) is approached, the Hessian matrix of a penalty function becomes increasingly more illconditioned, thus causing the successive unconstrained minimizations to become increasingly sensitive to numerical truncation and roundoff errors that often cause failure of the method to solve the NLP problem. Illconditioning of a square symmetric matrix can be measured by a "condition number" defined as the ratio of the absolute value of the largest to the smallest eigenvalue of the matrix. As the condition number increases, the illconditioning of the matrix increases.

The effect of illconditioning can be interpreted and understood by examining a problem of two variables, namely the quadratic function,

$$f(x) = \left(\frac{1}{2}\right) x^T \nabla^2 f(x) x = \left(\frac{1}{2}\right) x^T \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix} x \tag{8}$$

where $a \geq 1$. Since the diagonal elements of $\nabla^2 f(x)$ are

its eigenvalues, the condition number of $\nabla^2 f(x)$ is simply a. As a increases, the contours of $f(x)$ elongate along the x_1 axis. In the limit, as a tends to infinity, the contours of $f(x)$ will tend toward straight lines forming a stationary valley. Effective numerical minimization of $f(x)$ will become limited by the accuracy of the computer as a increases.

The Hessian matrix of a parametric penalty function has a condition number proportional to its weighting factors r. As an example, consider the well known SUMT penalty function of Fiacco and McCormick (1968)

$$P(x, r_S^{(k)}) = f(x) + (r_S^{(k)})^{1/2} \sum_{i=1}^{m} h_i^2(x) + \left(\frac{1}{r_S^{(k)}}\right) \sum_{i=m+1}^{p} \frac{1}{g_i(x)} \quad (9)$$

Suppressing the k superscript, and letting the vector x at the minimizing point for a given r_S, $\hat{x}(r_S)$, be designated just as $\hat{x}$, the gradient of (9) at $\hat{x}$ is

$$\nabla P(\hat{x}(r_S)) = \nabla f(\hat{x}) + 2r_S^{1/2} \sum_{i=1}^{m} h_i(\hat{x}) \nabla h_i(\hat{x}) - \frac{1}{r_S} \sum_{i=m+1}^{p} \frac{\nabla g_i(\hat{x})}{g_i^2(\hat{x})} \quad (10)$$

The terms $2r_S^{1/2}h_i$ and $-1/(r_S g_i^2)$ can be viewed as the nonoptimal Lagrangian multipliers associated with the constraints h_i and g_i respectively. The Hessian matrix of (9) is

$$\nabla^2 P(\hat{x}(r_S)) = \nabla^2 f(\hat{x}) + 2r_S^{1/2} \sum_{i=1}^{m} h_i(\hat{x})(\nabla^2 h_i(\hat{x}) + \nabla h_i(\hat{x}) \nabla^T h_i(\hat{x})) - \frac{1}{r_S} \sum_{i=m+1}^{p} \left(\frac{\nabla^2 g_i(\hat{x})}{g_i^2(\hat{x})} - 2 \frac{\nabla g_i(\hat{x}) \nabla^T g_i(\hat{x})}{g_i^3(\hat{x})} \right) \quad (11)$$

Assume that the first q inequality constraints are active where $0 \leq q < n - m$. Taking the limit of (11) as r_S tends to infinity yields

$$\nabla^2 P(x^*) = \left[\nabla^2 f(x^*) + \sum_{i=1}^{m} w_i^* \nabla^2 h_i(x^*) - \sum_{i=m+1}^{p} u_i^* \nabla^2 g_i(x^*) \right] + \lim_{r_S \to \infty} \left[2r_S^{1/2} \sum_{i=1}^{m} \nabla h_i(\hat{x}) \nabla^T h_i(\hat{x}) + \frac{1}{r_S} \sum_{i=m+1}^{p} \frac{\nabla g_i(\hat{x}) \nabla^T g_i(\hat{x})}{g_i^3(\hat{x})} \right] \quad (12)$$

where

$$w_i^* = \lim_{r_S \to \infty} 2r_S^{1/2} h_i(\hat{x}) \quad \text{and} \quad u_i^* = \lim_{r_S \to \infty} \frac{1}{r_S g_i^2(\hat{x})}$$

and x^* is the optimal x vector at the solution of Problem 1. The first bracketed term can be recognized as the Hessian matrix of the Lagrangian function

$$L(x, u, w) = f(x) + \sum_{i=1}^{m} w_i h_i(x) - \sum_{i=m+1}^{p} u_i g_i(x) \quad (13)$$

evaluated at u^*, w^*, x^* and will be assumed to be of full rank (rank n) and not badly conditioned. The term

$$2r_S^{1/2} \sum_{i=1}^{m} \nabla h_i(\hat{x}) \nabla^T h_i(\hat{x})$$

is at most of rank m with m eigenvalues proportional to $r_S^{1/2}$ and $n - m$ zero eigenvalues. As r_S tends to infinity, the m nonzero eigenvalues will approach infinity. Nonoptimal Lagrangian multipliers can be substituted into the remaining term and a separation based on constraint activity can be made:

$$\frac{1}{r_S} \sum_{i=m+1}^{p} \frac{\nabla g_i(\hat{x}) \nabla^T g_i(\hat{x})}{g_i^3(\hat{x})} = \sum_{i=m+1}^{m+q} \frac{u_i \nabla g_i(\hat{x}) \nabla^T g_i(\hat{x})}{g_i(\hat{x})} + \sum_{i=m+q}^{p} \frac{u_i \nabla g_i(\hat{x}) \nabla^T g_i(\hat{x})}{g_i(\hat{x})} \quad (14)$$

where the first matrix term on the right-hand side of (14) corresponds to the active inequality constraint set and the second to the inactive inequality constraint set.

It can be assumed that the elements of u are finite as r_S approaches infinity. The $(n \times n)$ active inequality matrix term

$$\sum_{i=m+1}^{m+q} u_i \nabla g_i(\hat{x}) \nabla^T g_i(\hat{x}) / g_i(\hat{x})$$

will then be at most of rank q with q eigenvalues proportional to $1/g_i$ and $n - q$ zero eigenvalues. The remaining inactive inequality term

$$\sum_{i=m+q}^{p} u_i \nabla g_i(\hat{x}) \nabla^T g_i(\hat{x}) / g_i(\hat{x})$$

will vanish in the limit due to Kuhn-Tucker optimality condition $u_i^* g_i(x^*) = 0$. The net effect of these contributions is that the $(n \times n)$ Hessian matrix (11) will have $m + q$ eigenvalues proportional to r_S and $n - m - q$ finite eigenvalues independent of r_S. As r_S tends to infinity, the $m + q$ eigenvalues proportional to r_S will approach infinity as will the condition number of $\nabla^2 P(x(r_S))$. By definition (12) will become illconditioned.

AVOIDANCE OF ILLCONDITIONING

Only two practical methods of avoiding the illconditioning described in the section above have been proposed. One general class of methods is known as acceleration techniques, the most common of which is that of extrapolation such as employed in both Fiacco and McCormack's SUMT method and Lootsma's interior-exterior point method. In both methods $\hat{x}(r_S)$ is extrapolated through previous minimizing points $\hat{x}(r_S^{(i)})$ to an estimated constrained minimum of Problem (1), $\tilde{x}^*$, which in turn is used as the starting vector for the next unconstrained minimization of P. The strategy is to begin the next unconstrained minimization of the penalty function as close to the minimum of Problem (1) as possible in an attempt to accelerate the search and decrease the existing numeri-

cal difficulties. In practice, Fiacco and McCormack provide the option of either a first or second-order extrapolation while Lootsma provided up to a seventh-order extrapolation. Observations based on the performance of Lootsma's method in solving test problems indicate that extrapolation can be an appropriate method of decreasing the effect of illconditioning of the Hessian matrix of a penalty function, but not a completely satisfactory one.

The second method, and the one that is the basis of the algorithm proposed here, is the use of a penalty function method termed the *method of multipliers* to solve the nonlinear programing problem (1). Powell (1969) and Hestenes (1969) independently suggested the new type of penalty as did Haarhoff and Buys (1970), a penalty function in which a quadratic loss term was added to the usual Lagrangian function:

$$P(x, r_A, w) = f(x) + \sum_{i=1}^{m} w_i h_i(x) + r_A \sum_{i=1}^{m} h_i^2(x) \quad (15)$$

From another viewpoint, a linear term was added to the quadratic loss penalty function. Hestenes suggested that solving a sequence of unconstrained minimizations of (15), each followed by updating the values of w_i by a simple formula that in effect maximized a dual problem (described in the next section), would solve Problem (1).

Miele et al. (1971a, 1971b, 1972a, 1972b) tried out different modifications of the basic concept in which only approximate minimizations were required for (15). Fletcher (1970a) and Fletcher and Lill (1971) proposed a related technique in which the multipliers w are adjusted continuously as the minimization of (15) is carried out.

Successful computational experience cited in several of the above references encouraged publication of many interesting as well as practical theoretical developments. Buy's in his dissertation (1972) proved local convergence of the dual technique of solution under reasonable assumptions. He also proved that the unconstrained minimization of (15) did not have to be exact. Bertsekas (1973, 1975) demonstrated global convergence of a more general class of penalty functions than (15) for the convex programing problem. Rockafellar (1973a, 1973b) summarized much of the literature and presented a number of new theorems for inequality constraints. Martensson (1973) indicated a number of ways in which the Lagrange multipliers could be computed and Bertsekas (1973b) described global duality theory for nonconvex problems as well as relations for convergence and the rates of convergence for multiplier methods.

It can be concluded that the method of multipliers ameliorates to some extent numerical instability caused by illconditioning of the Hessian matrix of the quadratic loss type penalty function as the weighting factor r_A increases.

Numerical test results reported by O'Doherty and Pierson (1974) for Powell's, Haarhoff and Buy's, and Meile's algorithms in solving selected equality constrained problems showed Powell's method to be superior in both robustness and efficiency. But Powell's method has a less than optimal convergence rate because with respect to the dual problem it represents nothing more than the method of steepest ascent with a fixed step length. Consequently the main contribution of this paper lies in the development of a superior technique to solve the dual problem, and hence the overall Problem (1).

NEW METHOD

Consider the augmented penalty function

$$P(x, \theta^{(k)}, r_A^{(k)}) = f(x) + r_A^{(k)} \sum_{i=1}^{m} (h_i^2(x) - 2h_i(x)\theta^{(k)}) \quad (16)$$

where $\theta_i^{(k)}$ is a parameter associated with the ith equality constraint, and $r_A^{(k)}$ is a scalar positive weighting factor. Rather than use slack variables, inequality constraints are included in the penalty function by transforming them into a moving loss function that maintains continuity of the gradient of P with respect to x

$$h_i(x, \theta_i) = \min \{g_i(x) - \theta_i, 0\} + \theta_i$$

Function (16) is equivalent to Powell's augmented penalty function (1969)

$$P(x, \theta^{(k)}, r^{(k)}) = f(x) + \sum_{i=1}^{m} r_i(h_i(x) - \theta_i)^2 \quad (17)$$

insofar as the minimization of $P(x, \theta, r)$ with respect to x is concerned for $r_A = r_1 = r_2 = \ldots = r_m$, but differs in suppressing the term $r_A \sum_{i=1}^{m} \theta_i^2$ in solving the dual problem. Function (16) is the same as (15) if the nonoptimal Lagrange multipliers are $w_i^{(k)} = 2r_A^{(k)}\theta_i^{(k)}$.

If between unconstrained minimizations of (16), r_A remains unaltered and θ is adjusted in manner such that $2r_A\theta$ converges to w^*, the Lagrange multiplier at the constrained solution x^*, the source of the illconditioning of the Hessian matrix of the penalty function is avoided. If the Hessian matrix of (16) is evaluated at some minimizing vector $\hat{x}(\theta, r_A)$ where θ and r_A are considered continuous for convenience:

$$\nabla_x^2 P(\hat{x}(\theta, r)) = \nabla_x^2 f(\hat{x}) + 2r_A \sum_{i=1}^{m} \theta_i \nabla_x^2 h_i(x)$$

$$+ 2r_A \sum_{i=1}^{m} h_i(\hat{x}) \nabla_x^2 h_i(\hat{x}) + 2r_A \sum_{i=1}^{m} \nabla h_i(\hat{x}) \nabla^T h_i(\hat{x})$$

The first two terms $\nabla_x^2 f + 2\Sigma r_A \theta_i \nabla_x^2 h_i$ can be recognized as the Hessian matrix of the Lagrangian function and will be assumed to have n finite eigenvalues. If r_A is maintained at a moderate value, the term $2r_A \Sigma h_i \nabla_x^2 h_i$ will have finite eigenvalues that will approach zero as x approaches x^*. The term $2r_A \Sigma \nabla h_i \nabla^T h_i$ will be at most of rank m with m eigenvalues proportional to r_A and $n - m$ zero eigenvalues. The net result is that if the magnitude of the elements of r_A are maintained at moderate values and the elements of w are finite, then $\nabla_x^2 P(\hat{x}(\theta, r))$ will have n nonzero finite eigenvalues and will not become illconditioned as the constrained optimum is approached.

If we examine the characteristics of the penalty function (16) as a function of θ, we find it has very nice properties for optimization. A minimum of $P(x, \theta^{(k)}, r^{(k)})$ at $\hat{x}$, denoted by $P(\hat{x}^{(k)}, \theta^{(k)}, r^{(k)})$, occurs when $\nabla_x P(\hat{x}^{(k)}, \theta^{(k)}, r^{(k)}) = 0$ and $v^T \nabla_x^2 P(\hat{x}^{(k)}, \theta^{(k)}, r^{(k)}) v > 0$, where v is a vector (that is, $\nabla_x^2 P$ is positive definite). Newell (1974) has shown, after some lengthy manipulations, that

$$v^T \nabla_\theta^2 P(\hat{x}^{(k)}, \theta, r^{(k)}) v = -4 \nabla_x^T h(\hat{x}^{(k)})$$

$$[\nabla_x^2 P(x, \theta^{(k)}, r^{(k)})]^{-1} \nabla_x h(\hat{x}^{(k)}) \quad (18)$$

so that $P(\hat{x}^{(k)}, \theta, r^{(k)})$ has the desirable property that

$\nabla_\theta^2 P(\hat{x}^{(k)}, \theta)$ is strictly negative definite if

1. $[\nabla_x^2 P(\hat{x}^{(k)}, \theta, r^{(k)}]^{-1}$ exists and is positive definite.

2. $\nabla_x h(\hat{x}^{(k)})$ is of full rank (for example, the gradients of constraints are linearly independent at $\hat{x}(\theta)$). If condition 1 is satisfied but not condition 2, then $\nabla_\theta^2 P(x^{(k)}, \theta, r^{(k)})$ is negative semidefinite. Therefore $P(\hat{x}^{(k)}, \theta, r^{(k)})$ is concave in θ for all vectors $\hat{x}$ minimizing $P(x, \theta, r)$.

Based on the properties of $P(x, \theta, r)$, Newell has shown that the solution of

$$\max_\theta (\min_x P(x, \theta, r)) = \max_\theta P(\hat{x}, \theta, r)$$

satisfies the necessary conditions for a constrained minimum of Problem (1). In other words, it can be demonstrated (a) that the Lagrangian multiplier (here θ) correction between exact unconstrained minimizations of the augmented penalty function should be a step maximizing the augmented penalty function in θ space, and (b) it is more effective to compute the Lagrangian multiplier correction by any efficient unconstrained maximization technique than by any of the methods recommended in the literature. These two concepts form the basis of an iterative primal minimization followed by a maximization of the dual problem, or the Generalized Augmented Penalty Function (GAPF) method.

The GAPF method follows a simple stepwise procedure (r_A is suppressed in the arguments).

1. Let $k = 0$ initially. Given the starting vector $x^{(0)}$, which may be a nonfeasible vector, select $\theta^{(0)}$, $r_A^{(0)}$, and $\eta^{(0)}$, where $\eta^{(0)}$ is the direction matrix. Typical examples are: $r_A^{(0)} = 1$, $\theta^{(0)} = 0$, and $\eta^{(0)} = -I$ (where I is the identity matrix). Set $P(\hat{x}^{(-1)}) = -10^{10}$ and $||h(\hat{x}^{(-1)})|| = 10^{10}$.

2. After the initial stage, let $k = k + 1$. Minimize $P(x, \theta^{(0)})$ with respect to x in search direction $s^{(k)}$ using any unconstrained minimization method to yield $(\hat{x}, \theta^{(0)})$. In the numerical evaluation of the proposed algorithm, Fletcher's unconstrained method (1970b) proved to be one of the best minimization techniques. A new $x^{(k+1)}$ is computed from the various stage $x^{(k)}$ as follows:

$$x^{(k+1)} = x^{(k)} - \gamma^{(k)} \eta^{(k)} \nabla P(x^{(k)}, \theta^{(k)})$$

The detailed formula for computing the direction matrix $\eta^{(k+1)}$ from $\eta^{(k)}$ can be found in the cited reference.

3. Terminate the method if $|h_i(x^{(k)})| < \epsilon_t$, $i = 1, \ldots, m$, where $0 < \epsilon_t << 1$, that is ϵ_t is some preselected convergence tolerance such as 10^{-5}; otherwise continue to step 4.

4. If $P(\hat{x}^{(k)}) > P(\hat{x}^{(k+1)})$ and $||h(\hat{x}^{(k)})|| \leq ||h(\hat{x}^{(k)})||$ continue to step 5. Otherwise let $r^{(k)} = 10r^{(k)}$ and $\theta^{(k)} = 0.1\theta^{(k)}$, and return to step 2 with $\hat{x}^{(k)}$ being the starting vector.

5. Let $\theta^{(k+1)} = \theta^{(k)} + \lambda^{(k)} s^{(k)}$, where $s^{(k)}$ is a vector in the direction of increasing values of $P(\hat{x}^{(k)}, \theta)$ with respect to θ, and $\lambda^{(k)}$ is a positive scalar chosen to maximize or increase $P(\hat{x}^{(k)}, \theta)$ in the direction $s^{(k)}$. Take one step to maximize $P(\hat{x}^{(k)}, \theta)$ with respect to θ yielding $(\hat{x}, \theta^{(k+1)})$. As a specific example, one might use Broyden's rank 1 method (1967) to maximize $P(\hat{x}^{(k)}, \theta)$

$$\theta^{(k+1)} = \theta^{(k)} + \eta^{(k)} h(x^{(k)}, \theta)$$

The formula to compute $\eta^{(k)}$, which is an estimate of $2r_A[\nabla_\theta^2 P(\hat{x}^{(k)}, \theta^{(k)}]^{-1}$, is given in the cited reference.

6. Return to step 2 with the starting x vector for the next stage $(k + 1)$ being $\hat{x}^{(k)}$ and the starting values for θ and r_A being the values from stage 5.

Although Fletcher's and Broyden's methods have been employed in steps 2 and 5, respectively, any unconstrained minimization technique could be used for steps 2 and 5. The method used for maximizing $P(\hat{x}^{(k)}, \theta)$ with respect to θ (that is, the method used for selecting $\lambda^{(k)}$ and $s^{(k)}$ in step 5) would depend on the information generated by the unconstrained minimization method used in step 2. Most variable metric minimization methods would supply a good estimate of $[\nabla_x^2 P(\hat{x}^{(k)}, \theta, r^{(k)}]^{-1}$ (or its inverse) which would allow an explicit calculation of $\nabla_\theta^2 P(\hat{x}^{(k)}, \theta)$ (or its inverse). The availability of $\nabla_\theta^2 P(\hat{x}^{(k)}, \theta)$ itself would allow the use of Newton's method for the unconstrained maximization of $P(\hat{x}^{(k)}, \theta)$ with respect to θ.

So far the role of the positive weighting factor r_A has not been discussed. As r_A increases, the penalty function as a function of x becomes increasingly poorly scaled for the minimization phase of the algorithm. On the other hand, as r_A increases, the penalty function, as a function of θ, becomes increasingly better scaled for the maximization phase. It is recommended that a moderate choice of $r_A^{(0)}$ be made (say $r_A^{(0)} = 1$) and that r_A be increased as the algorithm proceeds only if an unconstrained minimization fails to converge or if the maximization phase is not monotonically convergent with respect to θ.

CONVERGENCE

The convergence of any implementation of the GAPF method is dependent on the convergence properties of the unconstrained minimization and maximization techniques used and the properties of the penalty function itself. Most variable metric and conjugate methods are guaranteed to converge to the optimum of a quadratic function in a finite number of steps if an optimum exists and the Hessian matrix of the function does not become singular. For example, take the case in which $f(x)$ is convex and quadratic and the $h_i(x)$, $i = 1, \ldots, m$, are linear. Then $P(x, \theta, r)$ is convex and quadratic. Furthermore, if $[\nabla_x^2 P(\hat{x}^{(k)}, \theta]^{-1}$ exists, then $P(\hat{x}^{(k)}, \theta, r)$ is concave and quadratic in θ and is strictly concave and quadratic if $\nabla_x h(\hat{x}^{(k)}, \theta)$ is of full rank. Consequently, if Broyden's rank 1 formula with a unit step length were used for the unconstrained minimization phase and Newton's method with a unit step length were used for the maximization phase, in theory convergence to the constrained minimum could be guaranteed after one minimization stage ($n + 1$ function and constraint evaluations), one maximization stage, and one more minimization stage ($n + 1$ additional function and constraint evaluations).

If $h_i(x)$ is nonlinear for any $i = 1, \ldots m$ or if $f(x)$ is not convex, then $P(x, \theta, r)$ is not necessarily convex in x everywhere. Bertsekas (1973, 1975) has delimited the conditions under which convergence can be proved and also the rate of convergence. He has shown (1974) how duality gaps due to lack of convexity are eliminated through the convexification effect introduced by the

penalty function.

For highly nonlinear functions more than one minimum of $P(x, \theta, r)$ with respect to x may occasionally be possible. Consequently, for a given $\theta^{(k)}$ several values of $(\hat{x}^{(k)}, \theta^{(k)})$ might be computed on one stage (k) in which case $P(\hat{x}^{(k)}, \theta^{(k)}, r)$ would not necessarily have a unique value. If the sequential unconstrained minimizations of $P(x, \theta, r)$ with respect to x remained within one particular but shifting convex region, then $P(\hat{x}^{(k)}, \theta)$ can be treated as a unique function. But if an unconstrained minimization jumps to a neighboring convex region, the maximization of $P(\hat{x}^{(k)}, \theta)$ with respect to θ could possibly appear to be diverging, or an apparent failure of a univariant search could occur in the maximization phase.

NUMERICAL RESULTS

To investigate the practicality and validity of the GAPF method presented above, its performance was compared with four other effective algorithms on the 25 test problems in Appendix A of Himmelblau (1972). Twelve of the problems were of chemical engineering origin. The four codes that survived the initial screening (besides the GAPF algorithm) offered a broad spectrum of techniques. They were:

1. The 1972 version of the Generalized Reduced Gradient (GREG) algorithm of Abadie and Guigou (1969). This code minimizes the objective function along directions predicted by a conjugate gradient method but restricted to the subspace of linear and locally linearized active constraints by means of the constrained gradient concept. Infeasibility with respect to nonlinear constraints is periodically corrected by means of repetitive generalized Newton iterations.

2. The Generalized Projection Method for Nonlinear Constraints (GPMNLC) developed by Kreuser and Rosen (1971). A mixed projection-penalty function algorithm, GPMNLC minimizes the exterior point penalty function

$$P(x, r_S^{(k)}) = f(x) + r_S^{(k)} \sum_i h_i^2(x) + r_S^{(k)} \sum_i [\min\{g_i(x), 0\}]^2$$

by means of Goldfarb's projection method (1968). Only nonlinear constraints are included in the penalty function while all linear constraints are accommodated by projection. As in Goldfarb's method, the Davidon-Fletcher-Powell (Fletcher and Powell, 1963) method is used as the basic minimization technique.

3. A penalty function method by Lootsma (1970) that sequentially minimizes the mixed interior-exterior point parametric penalty function

$$P(x, r^{(k)}) = f(x) + r_S^{(k)} \sum_{i=1}^{m} h_i^2(x) + r_S^{(k)} \sum_{i=m+1}^{m+q} [\min\{g_i(x), 0\}]^2 + \frac{1}{r_S} \sum_{i=m+q}^{k} \frac{1}{g_i(x)}$$

On initiating an unconstrained minimization, all violated inequality constraints are placed in the quadratic loss term; the remainder are placed in the barrier function term. The Davidon-Fletcher-Powell method is used for

TABLE 1. NUMBER OF VARIABLES, CONSTRAINTS, AND BOUNDS ON VARIABLES FOR THE TEST PROBLEMS

Prob. no.	No. of variables	No of eq. constraints Linear	No of eq. constraints Nonlinear	No. of ineq. constraints Linear[a]	No. of ineq. constraints Nonlinear[a]	Upper and lower bounds[a]
Category A						
2	2	—	—	—	—	—
8	4	—	—	—	—	8 (0)
17	10	—	—	—	—	20 (0)
21	3	—	—	—	—	6 (0)
Category B						
4	10	3	—	—	—	10 (0)
6	45	16	—	—	—	45 (18)
10	5	—	—	10 (4)	—	5 (0)
19	16	8	—	—	—	32 (5)
23	100	—	—	12 (7)	—	100 (75)
Category C						
1	2	1	—	—	1 (1)	—
3	2	—	—	—	3 (0)	4 (2)
4a	10	—	3	—	—	—
5	3	1	1	—	—	3 (0)
7	3	—	—	—	14 (1)	6 (1)
9	4	—	—	—	1 (0)	2 (0)
11	5	—	—	—	6 (2)	10 (3)
12	5	—	—	1 (1)	37 (2)	10 (1)
13	5	—	—	—	2 (1)	9 (4)
14	6	—	—	—	4 (2)	—
15	6	—	4	—	—	12 (1)
16	9	—	—	—	13 (8)	1 (1)
18	15	—	—	—	5 (5)	15 (6)
20	24	2	12	—	6 (2)	24 (15)
22	6	—	—	—	4 (1)	12 (5)
24	2	—	—	1 (1)	1 (1)	—

[a] The number of active constraints at the solution are shown in parentheses.

the unconstrained minimizations of the penalty function. Between each unconstrained minimization, an initial starting point is predicted for the next minimization by an up to seventh-order extrapolation.

4. The Constrained Optimization via Moving Exterior Truncations (COMET) algorithm by Staha (1973). It is a traditional exterior point nonparametric penalty function algorithm. Fletcher's method (1970b) is employed for the sequential unconstrained minimizations. Before each minimization, an initial starting point is predicted by first-and second-order extrapolation.

To compare and evaluate the performance of the five codes, it was convenient to separate the problems into the following three categories:

Category A: Problems containing no constraints other than upper and lower bounds.
Category B: Problems containing only linear constraints in addition to upper and lower bounds.
Category C: Problems having one or more nonlinear constraints.

Table 1 lists each test problem by category, number of independent variables, number and type of constraints, and number of constraints active at the constrained optimum.

Each of the problems was solved (or the solution attempted) by each of the five codes. Computation time, and the number of objective function, objective function gradient, constraint and constraint gradient evaluations required by a code to solve a problem to the same degree of precision were recorded, namely,

$$||h_i(x^*)|| \leq 5.0 \times 10^{-6}$$

$$g_i(x^*) \geq -5.0 \times 10^{-6}$$

The time or number of equivalent function evaluations (that is, times for function evaluations other than the objective function were evaluated so that the functions could be expressed as being equivalent to the objective function) required by a code to solve a particular problem was divided into least time or lowest number of equivalent function evaluations recorded for that problem and multiplied by 100 to give a scaled performance index. Thus the most efficient code in solving a problem would have a scaled value of 100 whereas a code that failed to solve a problem would have a scaled value of 0 for that problem.

Table 2 presents the average scaled computation times required by each of the five codes to solve the problems in Categories A, B, and C and the whole set of problems. Table 3 lists the scaled equivalent function evaluations for each code. The average scaled equivalent function evaluations required by each code to solve the unconstrained problems in Category A are a direct measure of the efficiency of the unconstrained minimization technique used by each code. COMET was the most efficient followed closely by GAPF. The low efficiencies of GREG and Lootsma indicate relatively inefficient unconstrained searches. The COMET and GAPF codes used Fletcher's unconstrained method but differ somewhat in relative efficiency.

For the linearly constrained problems in Category B, the GPMNLC was superior to those of the three penalty function algorithms as anticipated because it uses a projection method. The problems in Category C all contain one or more nonlinear constraints and in general are more difficult to solve than problems in the other two categories. Based on computation time, GAPF is the most efficient of the codes followed by GREG. Based on function evaluations GREG is first followed by GAPF.

Table 2. Average Values of Scaled Computation Times[a]

	Category A	Category B	Category C	Overall
Lootsma	14.3	11.9	14.5	13.9
COMET	78.4	36.5	40.3	45.7
GAPF	64.7	28.2	67.8	59.4
GPMNLC	52.3	50.3	11.8	26.0
GREG	32.4	52.8	62.3	55.6

[a] Efficiency increases as scaled values increase.

Table 3. Average Values of Scaled Equivalent Function Evaluations

	Category A	Category B	Category C	Overall
Lootsma	25.5	2.4	8.7	10.1
COMET	83.1	9.5	21.0	28.7
GAPF	65.0	9.8	50.8	44.9
GPMNLC	63.8	51.0	7.2	25.0
GREG	34.9	11.9	63.7	48.8

Table 4. Average Values of Scaled Computation Times and Scaled Equivalent Function Evaluations for Problems with Nonlinear Equality Constraints

	Scaled time	Scaled function evaluations
Lootsma	18.9	15.1
COMET	44.1	25.2
GAPF	95.6	68.7
GPMNLC	6.9	6.1
GREG	32.0	41.3

Table 4 reflects the advantage of the GAPF code in solving problems with nonlinear equality constraints, the most difficult type of problems to treat.

It was concluded that for problems with nonlinear constraints, and especially nonlinear equality constraints, the GAPF algorithm was as at least as robust as the best existing algorithms and probably would be more efficient over a broad range of problems.

NOTATION

a = a parameter in Equation (8)
E^n = n dimensional Euclidean space
$f(x)$ = objective function
$g(x)$ = vector of inequality constraints
$g_i(x)$ = ith inequality constraint
$G(g(x))$ = function for inequality constraints
$h(x)$ = vector of equality constraints
$h_i(x)$ = ith equality constraint
$H(h(x))$ = function for equality constraints
I = identity matrix
$L(x, u, w)$ = Lagrange function
m = number of equality constraints
n = number of variables
p = number of inequality plus equality constraints
$P(f, g, h)$ = general penalty function
$P(x, r^{(k)})$ = general parametric penalty function
q = active inequality constraints
r = vector of scalar parameters (weighting factors)
r_i = scalar parameter (weighting factor) associated with ith constraint
r_A = parameter in the method of multipliers

r_S = parameter in SUMT
s = search direction vector
u_i = Lagrange multiplier associated with ith inequality constraint
u = vector of Lagrange multiplier associated with the inequality constraints
v = arbitrary vector
w_i = Lagrange multiplier associated with ith equality constraint
w = vector of Lagrange multipliers associated with the equality constraints
x = vector of variables
x_i = ith variable
$\hat{x}$ = column vector of x denoting the vector minimizing $P(x, r_S^{(k)})$
x^* = column vector of x at the optimal solution of Problem (1)

Greek Letters

α_i = exponent on r_i
γ = step length in a given search direction used in minimization phase
∇ = gradient operator (vector of first partial derivatives with respect to x, or θ as shown by subscript)
∇^2 = Hessian operator (matrix of second partial derivatives with respect to x or θ as shown by subscript)
ϵ = convergence criterion
λ = step length in a given search direction used in maximization phase
η = direction matrix in a conjugate direction minimization algorithm
θ_i = parameter associated with ith constraint in the augmented penalty function
θ = vector of parameters associated with the constraints in the augmented penalty function

Superscripts

(k) = index for kth stage of the optimization
T = transpose
$*$ = at optimal solution
$\sim$ = estimated

LITERATURE CITED

Abadie, J., and J. Carpentier, "Generalization of the Wolfe Reduced Gradient Method to the Case of Nonlinear Constraints," in *Optimization*, R. Fletcher (ed.), p. 37, Academic Press, New York (1969).

Abadie, J., and J. Guigou, "Gradient Reduit Generalise," Note HI 069/02, Electricité de France, 92-Clamart, France (1969).

Bertsekas, D. P., "Combined Primal Dual and Penalty Methods for Constrained Minimization," *SIAM J. Control*, in press (1975).

———., "Convergence Rate of Penalty and Multiplier Methods," in *Proc. of 1973 IEEE Conf. on Decision and Control*, p. 260, San Diego, Calif. (1973).

Broyden, C. G., "Quasi-Newton Methods and Their Application to Function Minimization," *Math. Comp.*, **12**, 368 (1967).

Buys, J. D., "Dual Algorithms for Constrained Optimization," Ph.D. thesis, Univ. Leiden, The Netherlands (1972).

Fiacco, A. V., and G. P. McCormick, "Extension of SUMT for Nonlinear Programming: Equality Constraints and Extrapolation," *Management Science*, **12**, 816 (1966).

———., *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York (1968).

Fletcher, R., "A Class of Methods for Nonlinear Programming with Termination and Convergence Properties," in *Integer and Nonlinear Programming*, J. Abadie (ed.), North Holland, Amsterdam (1970a).

———., "A New Approach to Variable Metric Algorithms," *Computer J.*, **13**, 317 (1970b).

———., "An Algorithm for Solving Linearly Constrained Optimization Problems," *Math. Progr.*, **2**, 133 (1972).

———., and S. A. Lill, "A Class of Methods for Nonlinear Programming, II: Computational Experience," in *Nonlinear Programming*, J. B. Rosen, O. L. Mangasarian, and K. Ritter (ed.), Academic Press, New York (1971).

Fletcher, R., and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization," *Computer J.*, **6**, 163 (1963).

Gill, P.E., and W. Murray, "Quasi-Newton Methods for Linearly Constrained Optimization," Nat. Phys. Lab. Rept. NAC 32 (May, 1973).

Goldfarb, D., "Extension of Davidon's Variable Metric Method to Maximization under Linear Inequality and Equality Constraints," *SIAM J. Appl. Math.*, **17**, 739 (1969).

Haarhoff, P. C., and J. D. Buys, "A New Method for the Optimization of a Nonlinear Function Subject to Nonlinear Constraints," *Computer J.*, **13**, 178 (1970).

Hestenes, M. R., "Multiplier and Gradient Methods," *J. Opt. Theory Appl.*, **4**, 303 (1969).

Himmelblau, D. M., *Applied Nonlinear Programming*, McGraw-Hill, New York (1972).

Husen, G. J., and J. M. Eakman, "Quadratically Convergent Techniques in Linearly Constrained Optimization," *AIChE J.* **20**, 555 (1974).

Kreuser, J. L., and J. B. Rosen, "GPM/GPMNLC Extended Gradient Projection Method Nonlinear Programming Subroutines," Academic Computing Center, Univ. Wisconsin, Madison (1971).

Lootsma, F. A., "Boundary Properties of Penalty Functions for Constrained Minimization," *Phillips Res. Repts.*, Suppl. No. 3 (1970).

Martensson, K., "A New Approach to Constrained Function Optimization," *J. Optimization Theory Applic.*, **12**, 531 (1973).

Miele, A., E. E. Cragg, R. R. Iver, and A. V. Levy, "Use of the Augmented Penalty Function in Mathematical Programming Problems, Part I," *ibid.*, **8**, 115 (1971a).

Miele, A., E. E. Cragg, and A. V. Levy, "Use of the Augmented Penalty Function in Mathematical Programming Problems, Part 2," *ibid.*, **8**, 131 (1971b).

Miele, A., P. E. Moseley, and E. E. Cragg, "A Modification of the Method of Multipliers for Mathematical Programming Problems," in *Techniques of Optimization*, A. V. Balakrishnan (ed.), Academic Press, New York (1972a).

Miele, A., P. E. Moseley, A. V. Levy, and G. M. Coggins, "On the Method of Multipliers for Mathematical Programming Problems," *J. Optimization Theory Applic.*, **10**, 1 (1972b).

Murray, W., "Ill-Conditioning in Barrier and Penalty Functions Arising in Constrained Nonlinear Programming," *Proc. 6th Intern. Symp. on Mathematical Programming*, Princeton, N. J. (1967).

Murtagh, B. A., and R. W. H. Sargent, "A Constrained Minimization Method with Quadratic Convergence," in R. Fletcher (ed.) *Optimization*, p. 215 Academic Press, New York (1969).

Newell, J. S., "An Augmented Penalty Function Method for Solving the General HLP Problem," Ph.D. dissertation, Univ. Texas, Austin (1974).

O'Doherty, R. J., and B. L. Pierson, "A Numerical Study of Multiplier Methods for Constrained Parameter Optimization," *Intern. J. of Systems Science*, **5**, 187 (1974).

Powell, M. J., "A Method for Nonlinear Constraints in Minimization Problems," in R. Fletcher (ed.), p. 283, *Optimization*, Academic Press, New York (1969).

Rockafellar, R. T., "The Multiplier Method of Hestenes and Powell Applied to Convex Programming," *J. Optimization Theory Applic.*, **12**, 555 (1973).

Rockafellar, R. T., "A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Optimization," *Math. Prog.*, to appear.

Staha, R. L., "Constrained Optimization Via Moving Exterior Truncations," Doctoral thesis, Univ. of Texas, Austin (1973).

Manuscript received September 16, 1974; revision received January 13 and accepted January 16, 1975.