Journal of Computational and Applied Mathematics 23 (1988) 155–168
North-Holland

A parameter-free multiplier method for constrained minimization problems

J.A. SNYMAN
Department of Applied Mathematics, University of Pretoria, Pretoria 0002, Republic of South Africa

Received 3 September 1986
Revised 7 January 1988

Abstract: This paper is concerned with the development of a parameter-free method, closely related to penalty function and multiplier methods, for solving constrained minimization problems. The method is developed via the quadratic programming model with equality constraints. The study starts with an investigation into the convergence properties of a so-called "primal-dual differential trajectory", defined by directions given by the direction of steepest descent with respect to the variables x of the problem, and the direction of steepest ascent with respect to the Lagrangian multipliers λ, associated with the Lagrangian function. It is shown that the trajectory converges to a stationary point (x^*, λ^*) corresponding to the solution of the equality constrained problem. Subsequently numerical procedures are proposed by means of which practical trajectories may be computed and the convergence of these trajectories are analyzed. A computational algorithm is presented and its application is illustrated by means of simple but representative examples. The extension of the method to inequality constrained problems is discussed and a non-rigorous argument, based on the Kuhn–Tucker necessary conditions for a constrained minimum, is put forward on which a practical procedure for determining the solution is based. The application of the method to inequality constrained problems is illustrated by its application to a couple of simple problems.

Keywords: Constrained minimization, multiplier methods, penalty function methods, augmented Langrangian.

1. Introduction

In an important class of methods the approach to solving constrained minimization problems is by transforming the constrained problem into a sequence of unconstrained problems which may be solved by standard methods. The most important methods in this class are the penalty function methods and so-called multiplier methods. Reviews of these two methods are given by Avriel [1] and by Bertsekas [2]. Both cases involve the solution of a sequence of unconstrained minimizations in which parameters are adjusted from one minimization to another. Exact penalty function methods, performing a single unconstrained optimization, have also been proposed [1].

In this paper we propose a method which has much in common with the multiplier methods but which performs a single unconstrained minimization and requires no parameter adjustments. The method is developed via the quadratic programming model with equality constraints. Extension of the method to deal with inequalities is also considered.

2. Penalty and multiplier methods

Because of their relevance to the current work we briefly describe the penalty and multiplier methods for the equality constrained problem

$$\text{minimize } F(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^n, \\ \text{subject to } h_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \ldots, m \tag{1}$$

where F, h_1, $h_2, \ldots, h_m$ are real valued functions on $\mathbb{R}^n$. The most common penalty function method uses a quadratic penalty and consists of sequential minimizations of the form

$$\text{minimize } P(\boldsymbol{x}, c_k) = F(\boldsymbol{x}) + c_k \| \boldsymbol{h}(\boldsymbol{x}) \|^2, \quad k = 0, 1, 2, \ldots \tag{2}$$

where $\boldsymbol{h}(\boldsymbol{x}) = [h_1(\boldsymbol{x}), \ldots, h_m(\boldsymbol{x})]^{\mathrm{T}}$ and the euclidean norm is used and the positive scalar sequence $\{c_k\}$ is such that $c_k \leqslant c_{k+1}$ for all k and $c_k \to \infty$. It can be shown that, under relatively mild assumptions on the nature of F and $\boldsymbol{h}$, the sequence of corresponding optimal solutions $\{\boldsymbol{x}_k^*\}$ of (2) converges to the optimal solution $\boldsymbol{x}^*$ of (1). Penalty methods are simple to implement, are applicable to a broad class of problems and take advantage of the very powerful minimization methods which have been developed for solving the unconstrained problem (2). On the other hand penalty methods are hampered by slow convergence and numerical instabilities associated with ill-conditioning of the Hessian matrix of the penalty function when the values of c_k become large.

The necessity for methods in which the parameters $\{c_k\}$ would need to assume moderate values only, prompted the development of a new class of methods known as multiplier methods. The first method, representative of this class of methods, was independently put forward by Hestenes [3] and by Powell [4]. We present the formulation due to Hestenes. In this method the penalty term is added not to the objective function F but rather to the Lagrangian function associated with (1). Consequently, Hestenes considers the sequence of minimizations of the form

$$\text{minimize } L(\boldsymbol{x}, \lambda_k, c_k) = F(\boldsymbol{x}) + \lambda_k^{\mathrm{T}} \boldsymbol{h}(\boldsymbol{x}) + \tfrac{1}{2} c_k \| \boldsymbol{h}(\boldsymbol{x}) \|^2, \quad k = 0, 1, 2, \ldots \tag{3}$$

where $\{c_k\}$ again is a sequence of positive penalty parameters and where the multiplier sequence $\{\lambda_k\}$ is updated by the formula

$$\lambda_{k+1} = \lambda_k + c_k \boldsymbol{h}(\boldsymbol{x}_k^*). \tag{4}$$

Here $\boldsymbol{x}_k^*$ is the optimal solution corresponding to (3) and the initial multiplier vector λ_0 is selected a priori. The sequence $\{c_k\}$ may be either preselected or generated during the computation according to some scheme.

The important aspect of this method is that it can be shown that under mild conditions, convergence may occur to the solution $\boldsymbol{x}^*$ of (1) without the need to increase c_k to infinity. In this case convergence is induced not merely by ever increasing values of the penalty parameters but also by the multiplier iteration (4). In this way the ill-conditioning, associated with the penalty methods, can be avoided and consequently multiplier algorithms have emerged as attractive methods for constrained optimization.

3. The primal-dual trajectory

It is well known that the constrained problem (1) may be considered via the unconstrained Lagrangian function

$$L(\boldsymbol{x}, \lambda) = F(\boldsymbol{x}) + \lambda^T \boldsymbol{h}(\boldsymbol{x}), \tag{5}$$

where $\lambda \in \mathbb{R}^m$. If F, h_i, $i = 1, 2, \ldots, m$, are continuously differentiable functions and if x^* is a constrained local minimum of problem (1) then there exists a vector of multipliers $\lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*)^T$ such that

$$\nabla_x L(x^*, \lambda^*) = \mathbf{0} \quad \text{and} \quad \nabla_\lambda L(x^*, \lambda^*) = \mathbf{0}. \tag{6}$$

One may therefore in principle obtain candidates for x^* by applying the above necessary conditions for stationarity and solving for x^* and the Lagrange multipliers λ^*. However, since equations (6) are in general difficult to solve, direct methods such as the penalty and multiplier methods mentioned in the previous section have been developed. These methods, as has been pointed out, require the solution of a sequence of subproblems as well as the selection of suitable parameters. We now investigate the possibility of obtaining the solution (x^*, λ^*) to equations (6) by following a single trajectory and without the need for selecting parameters. When solving the unconstrained minimization problem; minimize $F(x)$, $x \in \mathbb{R}^n$; then the application of the steepest descent method to convex functions leads to the minimum x^*. Essentially the steepest descent method consists of following the gradient path given by the solution of

$$\frac{dx}{dt}(t) = -\nabla F(x), \quad x(0) = x_0, \tag{7}$$

where x_0 is some given initial starting point. For the Lagrangian function the gradient path defined by

$$\begin{aligned} \frac{dx}{dt} &= -\nabla_x L(x, \lambda), \quad x(0) = x_0, \\ \frac{d\lambda}{dt} &= -\nabla_\lambda L(x, \lambda), \quad \lambda(0) = \lambda_0, \end{aligned} \tag{8}$$

will not necessarily converge to (x^*, λ^*) since if it represents a solution to constrained problem (1) it also corresponds to a saddlepoint. The two equations in (8) correspond to directions of steepest descent with respect to both x and λ. One intuitively feels that by switching the sign of the second equation to be positive, i.e. to steepest ascent one may improve the convergence property of the resulting gradient path. Switching the sign may be viewed as attempting to solve the primal and associated dual problem simultaneously. Consequently we consider the convergence property of the 'primal-dual trajectory' defined by

$$\begin{aligned} \frac{dx}{dt} &= -\nabla_x L, \quad x(0) = x_0, \\ \frac{d\lambda}{dt} &= +\nabla_\lambda L, \quad \lambda(0) = \lambda_0. \end{aligned} \tag{9}$$

To simplify matters we consider the quadratic programming problem

$$\begin{aligned} &\text{minimize } F(x) = \tfrac{1}{2}x^T A x + b^T x, \\ &\text{subject to } h(x) = Cx - d = \mathbf{0}, \end{aligned} \tag{10}$$

where A is a $n \times n$ positive definite matrix and C a $m \times n$ matrix, $m < n$. The Lagrangian function is given by

$$L(x, \lambda) = \tfrac{1}{2}x^T A x + b^T x + \lambda^T(Cx - d) \tag{11}$$

and the corresponding primal-dual trajectory is defined by

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -\nabla_x L = -A\boldsymbol{x} - \boldsymbol{b} - C^{\mathrm{T}}\boldsymbol{\lambda}, \qquad \frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t} = +\nabla_\lambda L = C\boldsymbol{x} - \boldsymbol{d}. \tag{12}$$

By equation (6) the necessary conditions for stationarity at $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is

$$-\nabla_x L = -A\boldsymbol{x}^* - \boldsymbol{b} - C^{\mathrm{T}}\boldsymbol{\lambda}^* = \boldsymbol{0} \tag{13}$$

and

$$\nabla_\lambda L = C\boldsymbol{x}^* - \boldsymbol{d} = \boldsymbol{0}.$$

Let $\boldsymbol{X} = \boldsymbol{x} - \boldsymbol{x}^*$ and $\boldsymbol{\Lambda} = \boldsymbol{\lambda} - \boldsymbol{\lambda}^*$ then it follows by subtracting (13) from (12) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{\Lambda}\end{bmatrix} = -\begin{bmatrix}A & C^{\mathrm{T}}\\ -C & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{\Lambda}\end{bmatrix} \tag{14}$$

which may be written as

$$\frac{\mathrm{d}\boldsymbol{Z}}{\mathrm{d}t} = -G\boldsymbol{Z} \tag{15}$$

$$\text{where } \boldsymbol{Z} = \begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{\Lambda}\end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix}A & C^{\mathrm{T}}\\ -C & \boldsymbol{0}\end{bmatrix}. \tag{16}$$

It follows that

$$\begin{aligned}-\boldsymbol{Z}^{\mathrm{T}}G\boldsymbol{Z} &= -\boldsymbol{X}^{\mathrm{T}}A\boldsymbol{X} - \boldsymbol{X}^{\mathrm{T}}C^{\mathrm{T}}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}^{\mathrm{T}}C\boldsymbol{X}\\ &= -\boldsymbol{X}^{\mathrm{T}}A\boldsymbol{X} < 0\end{aligned} \tag{17}$$

for all $\boldsymbol{X} \neq \boldsymbol{0}$, by virtue of the fact that A is positive definite. Thus if $\boldsymbol{Z}(t)$ represents the solution of (15) from some initial point $\boldsymbol{Z}(0) = \boldsymbol{Z}_0$, then

$$\tfrac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(\|\boldsymbol{Z}\|^2) = \boldsymbol{Z}^{\mathrm{T}}\frac{\mathrm{d}\boldsymbol{Z}}{\mathrm{d}t} = -\boldsymbol{Z}^{\mathrm{T}}G\boldsymbol{Z} < 0, \ \boldsymbol{X} \neq \boldsymbol{0}. \tag{18}$$

If $\boldsymbol{X} = \boldsymbol{0}$ and $\boldsymbol{\Lambda} \neq \boldsymbol{0}$ then we momentarily have

$$\mathrm{d}\|\boldsymbol{Z}\|/\mathrm{d}t = 0 \quad \text{with } \mathrm{d}\boldsymbol{\Lambda}/\mathrm{d}t = \boldsymbol{0} \quad \text{and} \quad \mathrm{d}\boldsymbol{X}/\mathrm{d}t = -C^{\mathrm{T}}\boldsymbol{\Lambda} \neq \boldsymbol{0}.$$

The trajectory therefore continues with $\boldsymbol{X}$ changing to $\boldsymbol{X} \neq \boldsymbol{0}$ and consequently, by (18), this results in a further decrease in $\|\boldsymbol{Z}\|$. The overall effect is that a spiralling trajectory is obtained which converges to $\|\boldsymbol{Z}\| = 0$ as t tends to infinity, i.e.

$$\operatorname*{limit}_{t\to\infty} \|\boldsymbol{Z}(t)\| = 0. \tag{19}$$

In the case of the quadratic problem (10) the primal-dual trajectory therefore converges to the unique saddle point $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ corresponding to the solution of the constrained problem.

More generally it can also be shown that, for a general function $F(\boldsymbol{x})$ with general constraints $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, the equilibrium point $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is asymptotically stable provided

$$A = (a_{ij}) = \left(\frac{\partial^2 F}{\partial x_i\, \partial x_j} + \sum_{r=1}^{m} \lambda_r \frac{\partial^2 h_r}{\partial x_i\, \partial x_j}\right) \tag{20}$$

is positive-definite at $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$.

4. The convergence of practical trajectories

From the above exposition it appears that under suitable conditions the solution $\boldsymbol{x}^*$ and the associated multiplier vector $\boldsymbol{\lambda}^*$ may be obtained in a single iteration by following the trajectory given by the solution of initial value problem (9). To do this economically required that a practical numerical procedure be available for computing the trajectory such as, for example, the optimal gradient method which is used for the unconstrained minimization of a function. In the latter case a trajectory is obtained by successive one-dimensional minimizations in search directions given by directions of steepest descent of the function. Unfortunately in our case the directions given by equations (9) do not directly correspond to the minimization of an associated function. We can however show that, under certain conditions, the directions given by (9) correspond to descent directions for the norm of the associated gradient vector. Therefore, since the necessary conditions (6) must apply at the solution $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$, the persuit of a trajectory obtained by using (9) as search directions with associated one-dimensional minimizations of the norm of the gradient vector, should yield the solution $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$.

Let ρ denote a function proportional to the square of the norm of the gradient vector and defined by

$$\rho = \tfrac{1}{2}\| G\boldsymbol{Z} \|^2 = \tfrac{1}{2}(G\boldsymbol{Z})^{\mathrm{T}}(G\boldsymbol{Z}). \tag{21}$$

For the direction given by (9) and therefore by (15) it follows that in the direction of the primal-dual trajectory we have

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = (G\boldsymbol{Z})^{\mathrm{T}}\frac{\mathrm{d}}{\mathrm{d}t}(G\boldsymbol{Z}) = \boldsymbol{Z}^{\mathrm{T}}G^{\mathrm{T}}G\frac{\mathrm{d}\boldsymbol{Z}}{\mathrm{d}t} = -\boldsymbol{Z}^{\mathrm{T}}G^{\mathrm{T}}GG\boldsymbol{Z}. \tag{22}$$

Let $G\boldsymbol{Z} = \left[\begin{smallmatrix}\boldsymbol{p}\\ \boldsymbol{q}\end{smallmatrix}\right]$, where $\boldsymbol{p} = \nabla_x L$ and $\boldsymbol{q} = -\nabla_\lambda L$ then we have by (16) and similarly to (17) that

$$\mathrm{d}\rho/\mathrm{d}t = -\boldsymbol{p}^{\mathrm{T}}A\boldsymbol{p} < 0 \tag{23}$$

for $\boldsymbol{p} \neq \boldsymbol{0}$, since A is positive definite.

Using the directions of equations (9) as search directions for minimizing ρ we have descent to $\rho = 0$, i.e: to $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$, provided of course that $\boldsymbol{p} \neq \boldsymbol{0}$ at each step of the computed trajectory. Unfortunately, as we may expect and as is also borne out by experiment, the latter is not always true and we may have that $\boldsymbol{p}$ approaches $\boldsymbol{0}$ at a point along the trajectory which is far removed from $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ and consequently we have very little or effectively zero descent in ρ.

To prevent the above situation we consider the following auxiliary Lagrangian function which is similar to the augmented Lagrangian (3) of Hestenes [3] except for the absence of the penalty parameter:

$$L_A(\boldsymbol{x}, \boldsymbol{\lambda}) = F(\boldsymbol{x}) + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}) + \tfrac{1}{2}\| \boldsymbol{h}(\boldsymbol{x}) \|^2. \tag{24}$$

The motivation for the use of the above Lagrangian is the following. If the components of $\boldsymbol{p}$, i.e. of $\nabla_x L$, become small, we have by (23) very slow convergence, i.e. the search direction given by (9) gives minimal descent. In particular if $\boldsymbol{p} = \boldsymbol{0}$ the x-components of $\mathrm{d}\boldsymbol{Z}/\mathrm{d}t$ equals zero. In such a case we may wish to choose a new search direction as $-\nabla\rho$, the direction of steepest descent of the more complicated function ρ.

Since

$$\rho = \tfrac{1}{2}(G\boldsymbol{Z})^{\mathrm{T}}G\boldsymbol{Z}$$

and the gradient is invariant under a linear transformation it follows that

$$\rho = \tfrac{1}{2}\left\{ \sum_{i=1}^{n} \left(\frac{\partial L}{\partial x_i}\right)^2 + \sum_{j=1}^{m} h_j^2 \right\}.$$

Thus the components of $\nabla\rho$ are given by

$$\frac{\partial \rho}{\partial x_k} = \sum_{i=1}^{n} \left(\frac{\partial L}{\partial x_i}\right) \frac{\partial^2 L}{\partial x_i \partial x_k} + \sum_{j=1}^{m} h_j \frac{\partial h_j}{\partial x_k}, \quad k = 1, 2, \ldots, n,$$

and

$$\frac{\partial \rho}{\partial \lambda_l} = \sum_{i=1}^{n} \left(\frac{\partial L}{\partial x_i}\right) \frac{\partial^2 L}{\partial x_i\, \partial \lambda_l}, \quad l = 1, 2, \ldots, m.$$

Now if $\boldsymbol{p} = \boldsymbol{0}$ we have that the components of the steepest descent direction are given by

$$-(\nabla\rho)_k = \begin{cases} -\sum_{j=1}^{m} h_j \dfrac{\partial h_j}{\partial x_k}, & k = 1, 2, \ldots, n, \\ 0 & k = n+1, \ldots, n+m. \end{cases}$$

This direction is in fact orthogonal to the primal-dual direction at the point where $\boldsymbol{p} = \boldsymbol{0}$. We also notice that if we consider the auxiliary function L_A given by (24) instead of the, in general, more complicated function ρ that

$$-\nabla_x L_A = -\nabla_x \rho$$

at the point where $\boldsymbol{p} = \boldsymbol{0}$. This gives the direction of steepest descent of ρ with respect to the $\boldsymbol{x}$-components and it therefore seems worthwhile to investigate the convergence of the differential path (9) with L replaced by L_A.

We consider again the quadratic model (10) for which we have

$$L_A = \tfrac{1}{2}\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{\lambda}^{\mathrm{T}}(C\boldsymbol{x} - \boldsymbol{d}) + \tfrac{1}{2}(C\boldsymbol{x} - \boldsymbol{d})^{\mathrm{T}}(C\boldsymbol{x} - \boldsymbol{d}). \tag{25}$$

The corresponding primal-dual path is defined by

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} &= -\nabla_x L_A = -\{A\boldsymbol{x} + \boldsymbol{b} + C^{\mathrm{T}}\boldsymbol{\lambda} + C^{\mathrm{T}}(C\boldsymbol{x} - \boldsymbol{d})\}, \\ \frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t} &= +\nabla_\lambda L_A = C\boldsymbol{x} - \boldsymbol{d}. \end{aligned} \tag{26}$$

Clearly the stationary point $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ of this path is the same as the previous equilibrium point of the trajectory given by (12).

Again let $\boldsymbol{X} = \boldsymbol{x} - \boldsymbol{x}^*$ and $\boldsymbol{\Lambda} = \boldsymbol{\lambda} - \boldsymbol{\lambda}^*$. It then follows from (26) and (13) that

$$\mathrm{d}\boldsymbol{X}/\mathrm{d}t = -(A + C^{\mathrm{T}}C)\boldsymbol{X} - C^{\mathrm{T}}\boldsymbol{\Lambda}, \qquad \mathrm{d}\boldsymbol{\Lambda}/\mathrm{d}t = C\boldsymbol{X}, \tag{27}$$

which may be written as

$$\frac{\mathrm{d}\boldsymbol{Z}}{\mathrm{d}t} = -\begin{bmatrix} A + C^{\mathrm{T}}C & C^{\mathrm{T}} \\ -C & \boldsymbol{0} \end{bmatrix} \boldsymbol{Z} = -G\boldsymbol{Z} - B\boldsymbol{Z} = -D\boldsymbol{Z}, \tag{28}$$

where $B = \begin{bmatrix} C^{\mathrm{T}}C & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$.

Here again we have for trajectory (26) that

$$\tfrac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(\|\boldsymbol{Z}\|^2) = -\boldsymbol{X}^{\mathrm{T}}(A + C^{\mathrm{T}}C)\boldsymbol{X} < 0$$

for all $\boldsymbol{X} \neq \boldsymbol{0}$, since $(A + C^{\mathrm{T}}C)$ is positive definite. Similarly to the argument preceding equation (19) it now follows that the trajectory converges to $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$.

Suppose now we use the directions given by (26) as search directions for minimizing ρ. We then have that

$$\begin{aligned}\frac{\mathrm{d}\rho}{\mathrm{d}t} &= (G\boldsymbol{Z})^{\mathrm{T}}\frac{\mathrm{d}}{\mathrm{d}t}(G\boldsymbol{Z}) = (G\boldsymbol{Z})^{\mathrm{T}}G\frac{\mathrm{d}\boldsymbol{Z}}{\mathrm{d}t}\\ &= -(G\boldsymbol{Z})^{\mathrm{T}}GD\boldsymbol{Z} = -(G\boldsymbol{Z})^{\mathrm{T}}G(G+B)\boldsymbol{Z}\\ &= -\boldsymbol{Z}^{\mathrm{T}}G^{\mathrm{T}}G G\boldsymbol{Z} - \boldsymbol{Z}^{\mathrm{T}}G^{\mathrm{T}}GB\boldsymbol{Z}.\end{aligned} \tag{29}$$

In particular we consider the eventuality $\boldsymbol{p} = 0$. In this case the first term vanishes and it can also easily be shown that the second term reduces to $-\boldsymbol{X}^{\mathrm{T}}C^{\mathrm{T}}CC^{\mathrm{T}}C\boldsymbol{X}$ so that we have

$$\mathrm{d}\rho/\,\mathrm{d}t = -\boldsymbol{X}^{\mathrm{T}}C^{\mathrm{T}}CC^{\mathrm{T}}C\boldsymbol{X} < 0 \tag{30}$$

for $\boldsymbol{X} \neq \boldsymbol{0}$, since $C^{\mathrm{T}}CC^{\mathrm{T}}C$ is positive-definite.

We may now combine search directions (12) and (26), i.e. L and L_A, in a convergent minimization procedure. This is done by alternative applying search directions (12) and (26). If $\boldsymbol{p} = \boldsymbol{0}$, $\boldsymbol{X} \neq \boldsymbol{0}$, then the application of (26) gives descent in ρ by the argument presented immediately above. If $\boldsymbol{p} \neq \boldsymbol{0}$ then direction (12) gives descent by (23) even if no descent was obtained in the previous step by the application of (26). Thus the alternate application of search directions (12) and (26) ensures descent in ρ unless both $\boldsymbol{X} = \boldsymbol{0}$ and $\boldsymbol{p} = \boldsymbol{0}$. If this happens then obviously $\boldsymbol{x} = \boldsymbol{x}^*$ and, since $\boldsymbol{p} = \nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \boldsymbol{0}$, it follows that $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ and convergence is attained.

Much of the theory of the method developed here, as is the case in the development of many other gradient methods, is only valid for the case where $F(\boldsymbol{x})$ is a positive definite quadratic function and $\boldsymbol{h}(\boldsymbol{x})$ represents linear constraints. We may however expect that this method, which essentially is also a gradient method, may be applicable to general nonlinear functions with, in reference to (20), the condition that $F(\boldsymbol{x})$ and $\boldsymbol{h}(\boldsymbol{x})$ be convex functions. The truth of this statement can only be established by much more difficult and sophisticated analysis and in practice by applying it to standard test functions. Also if the constraints $\boldsymbol{h}(\boldsymbol{x})$ are nonlinear the stationary point $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ may no longer be unique. Allowance must then be made for the possibility of convergence, depending on the starting point $(\boldsymbol{x}_0, \boldsymbol{\lambda}_0)$, to more than one stationary point corresponding to different constrained local minima.

5. Computational algorithm

A formal algorithm representing the method developed in the previous sections may now be stated as follows.

Given a small number $\epsilon > 0$ and a starting point $\boldsymbol{Z}_0 = (\boldsymbol{x}_0, \boldsymbol{\lambda}_0)$, set $k \leftarrow 0$ and perform the following steps.

Step 1. If k equals zero or is even then

1.1. set $\boldsymbol{g}_k \leftarrow (-\nabla_x L, +\nabla_\lambda L)$;

1.2. else set $\boldsymbol{g}_k \leftarrow (-\nabla_x L_A, +\nabla_\lambda L_A)$.

Step 2. Perform a one-dimensional minimization of ρ in the search direction $\boldsymbol{g}_k$, i.e. determine α_k such that

$$\rho(\boldsymbol{Z}_k + \alpha_k \boldsymbol{g}_k) = \min \rho(\boldsymbol{Z}_k + \alpha \boldsymbol{g}_k), \quad \alpha \in \mathbb{R}.$$

Step 3. Set $\boldsymbol{Z}_{k-1} \leftarrow \boldsymbol{Z}_k + \alpha_k \boldsymbol{g}_k$.

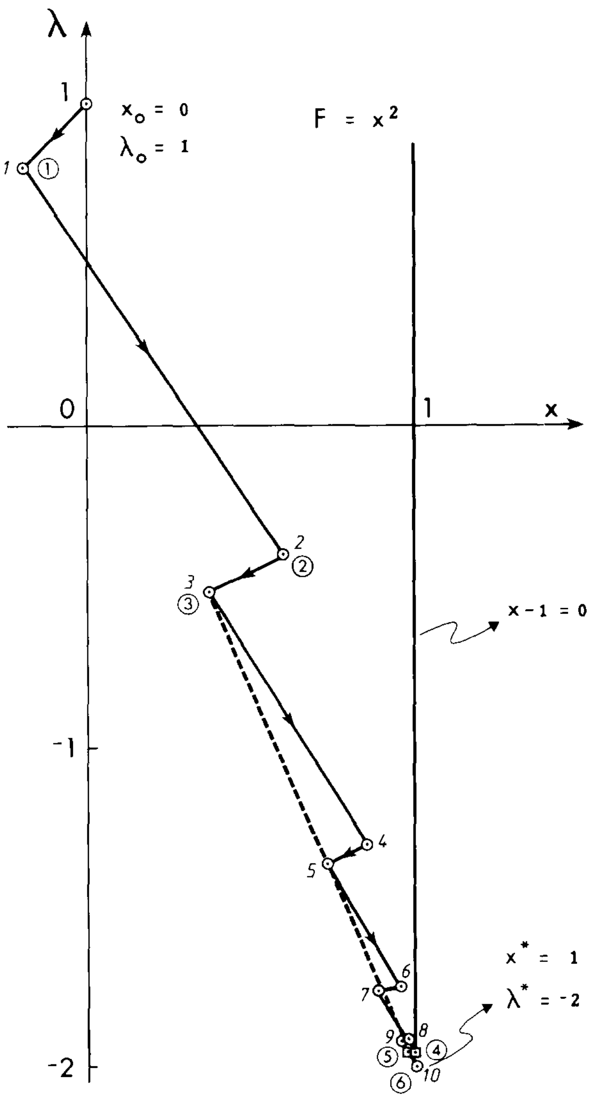

Fig. 1. Minimization trajectories for Example 1 computed by the original (full line) and modified (broken line) algorithms respectively. The encircled numbers denote the minimization steps for the modified algorithm.

Step 4. If $\rho(Z_{k+1}) < \epsilon$, then STOP;
else set $k \leftarrow k+1$ and go to Step 1.

Typically we have, starting at Z_0, that the search direction (1.1) in step 1 yields Z_1, then search direction (1.2) gives Z_2 and again applying (1.1) we obtain Z_3. Since, under the conditions assumed in the derivation of the method, we must have descent from Z_1 to Z_3 for $(x_1, \lambda_1) \neq (x^*, \lambda^*)$, we may seek to improve the efficiency of the algorithm by performing a further minimization along the descent direction $g_3 = Z_3 - Z_1$ to obtain Z_4. Having determined Z_4 in this manner we may set $Z_0 \leftarrow Z_4$ and repeat the complete procedure until convergence is obtained. The effect of this modification is illustrated in Figure 1 which shows the descent trajectories given by both the original and the modified algorithms for the following rather trivial example.

Example 1. Minimize x^2, subject to $x - 1 = 0$.
The starting point is $Z_0 = (x_0, \lambda_0) = (0, 1)$ and the problem has the solution $x^* = 1$, $\lambda^* = -2$. The figure clearly indicates that the modification drastically reduces the number of minimization steps required for convergence. On the basis of this evidence the modified algorithm is accepted

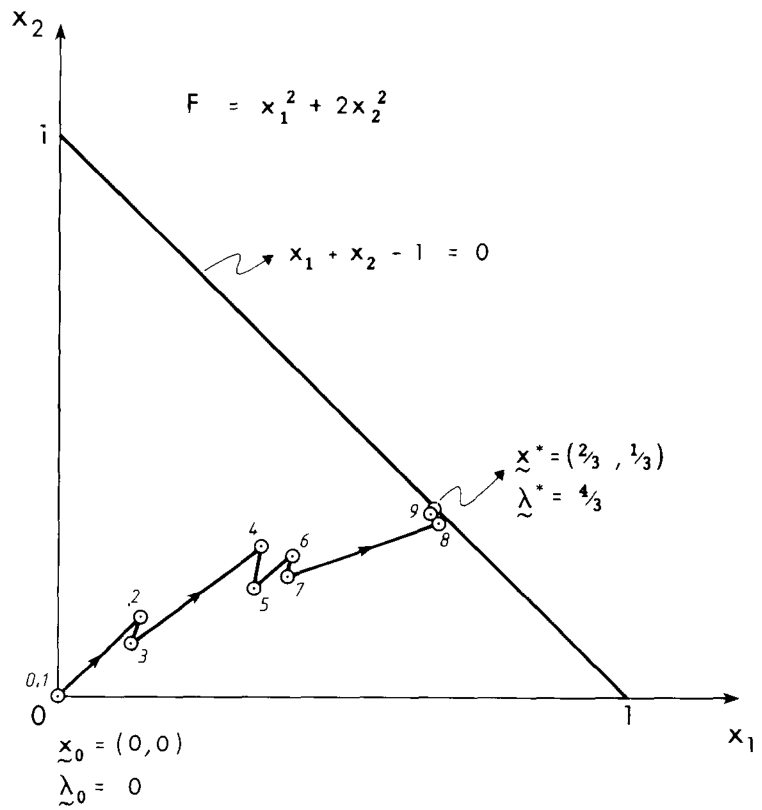

Fig. 2. Minimization trajectory in x-space for the quadratic Example 2.

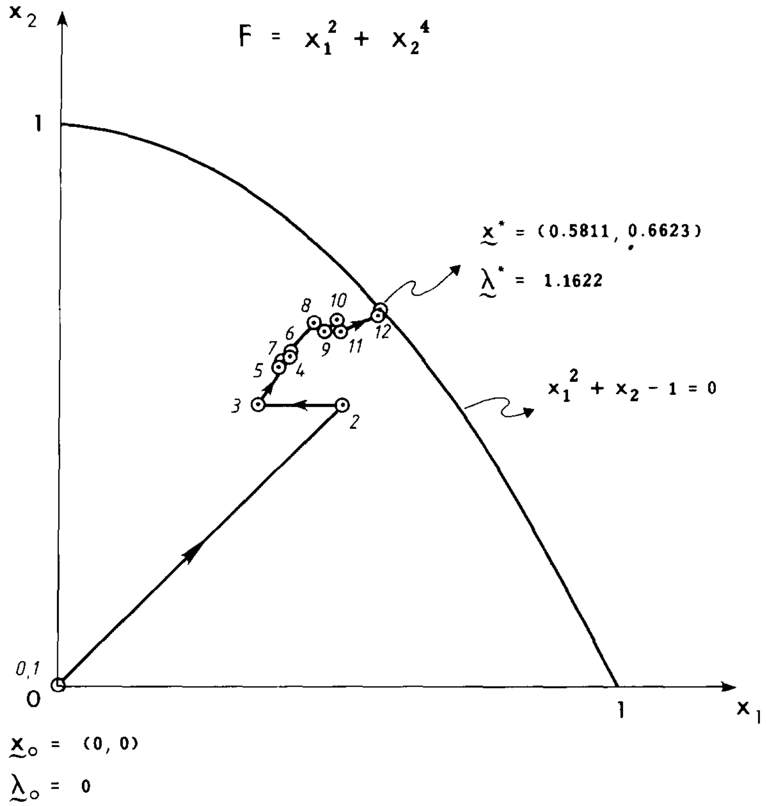

Fig. 3. Minimization trajectory in x-space for the non-quadratic Example 3.

as the standard algorithm and in all further references to the algorithm the standard form will be assumed.

The working of the algorithm is further illustrated in Fig. 2 and 3 which depict the respective trajectories in x-space for the two problems listed below. The first problem is quadratic with a linear constraint in accordance with (10) but the second is nonquadratic with a nonlinear constraint.

Example 2. Minimize $x_1^2 + 2x_2^2$, subject to $x_1 + x_2 - 1 = 0$.
The starting point is $x_0 = (0,0)$ and $\lambda_0 = 0$ and the problem has the solution $x^* = (\frac{2}{3}, \frac{1}{3})$ and $\lambda^* = \frac{4}{3}$.

Example 3. Minimize $x_1^2 + x_2^4$, subject to $1 - x_1^2 - x_2 = 0$.
The starting point is $x_0 = (0, 0)$ and $\lambda_0 = 0$, and the problem has the solution $x^* = (0.5811, 0.6623)$ and $\lambda^* = 1.1622$.

Although these examples are almost trivial in terms of mathematical complexity, the trajectories shown in Figs. 2 and 3 are representative of the general performance of the algorithm when applied to more complicated problems.

6. Treatment of inequality constraints

Inequality constraints may be treated trivially by converting them to equality constraints by using additional variables [2]. Consider the following problem involving one-sided inequality constraints.

$$\text{minimize } F(\boldsymbol{x}),$$
$$\text{subject to } k_j(\boldsymbol{x}) \leqslant 0, \quad j = 1, 2, \ldots, r. \tag{31}$$

This problem is equivalent to the equality constrained problem

$$\text{minimize } F(\boldsymbol{x}),$$
$$\text{subject to } k_j(\boldsymbol{x}) + z_j^2 = 0, \quad j = 1, 2, \ldots, r; \tag{32}$$

where $z_1, z_2, \ldots, z_r$ are additional variables. Thus we may use our parameter-free multiplier method to solve (32) in place of (31). One must however bear in mind that the introduction of the quadratic terms z_j^2 in the constraints may result in the existence of more than one stationary point and therefore, depending on the choice of starting point, convergence to a non-optimal point.

We now present an alternative approach to inequality constraints via a somewhat non-rigorous argument involving the Kuhn–Tucker necessary conditions for a minimum involving inequality constraints. The Kuhn–Tucker theorem [5] states that if $\boldsymbol{x}^*$ is a solution of (31), then there exists a vector $\boldsymbol{\lambda}^*$ such that

$$\nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{0}, \qquad \lambda_j^* k_j(\boldsymbol{x}) = 0 \tag{33}$$

and

$$\lambda_j \geqslant 0, \quad j = 1, 2, \ldots, r;$$

where $L(\boldsymbol{x}, \boldsymbol{\lambda}) = F(\boldsymbol{x}) + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{k}(\boldsymbol{x})$.

We may therefore conclude that if $\lambda_j^* > 0$ then $k_j(\boldsymbol{x})^* = 0$, i.e. it is an active constraint. (If $\lambda_j^* = 0$ and $k_j(\boldsymbol{x}) = 0$ we may effectively take the constraint as not being active since it plays no active role in determining $\boldsymbol{x}^*$). If the active constraints were known beforehand, i.e. if we knew which $\lambda_j^* > 0$, we could determine $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ as we have done for equality constraints before by determining the saddle point of the Lagrangian. This is so since the Kuhn–Tucker stationary conditions can also be stated in terms of the following Saddlepoint Theorem [5]: If the point $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$, $\boldsymbol{\lambda}^* \geqslant \boldsymbol{0}$ is a saddlepoint for the Lagrangian associated with primal problem (31), then $\boldsymbol{x}^*$ solves the primal problem.

The problem which remains is to determine which constraints are active. Obviously if we assumed they were all active and they were not, application of our method would yield a stationary point with $\lambda_j < 0$ for some j. This would, by (33), prove that our assumption was wrong and that some constraints (those corresponding to $\lambda_j^* < 0$?) should be omitted from the active list.

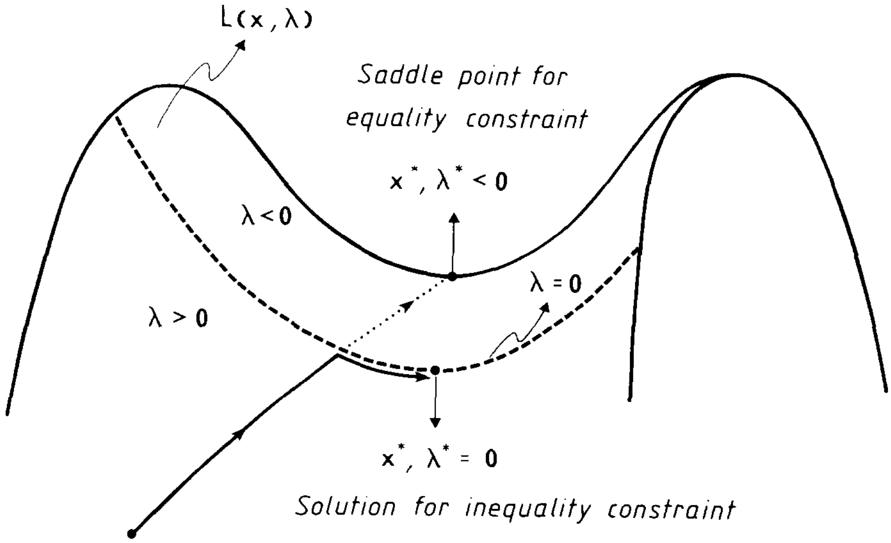

Fig. 4. Schematic representation of the convergence of the primal-dual trajectory on the Lagrangian surface for both an equality and inequality constraint.

Consider the quadratic programming problem:

$$\begin{aligned}&\text{minimize } F(\boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x},\\ &\text{subject to } \boldsymbol{k}(\boldsymbol{x}) = C\boldsymbol{x} - \boldsymbol{d} \leqslant \boldsymbol{0},\end{aligned} \tag{34}$$

and matrices A and C as defined in (10). We now conjecture that if the primal-dual trajectory, defined by (12), is followed with the proviso that λ_j is not allowed to become negative; i.e. if $\lambda_j = 0$ and $\mathrm{d}\lambda_j/\mathrm{d}t < 0$, then set $\lambda_j \leftarrow 0$; then the trajectory will lead to a point $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$, $\boldsymbol{\lambda}^* \geqslant \boldsymbol{0}$, which corresponds to the solution of the primal problem. This proposition is schematically illustrated in Fig. 4.

Since our practical algorithm of Section 5 in effect attempts to approximate trajectory (12) we may apply it to inequality constrained problems with the following modification. Before the calculation of the gradient vector $\boldsymbol{g}$ in Step 1, perform the following test: if $\lambda_j \leqslant 0$ and $k_j(\boldsymbol{x}) \leqslant 0$ then set $\lambda_j \leftarrow 0$ and $k_j \leftarrow 0$, i.e., we effectively drop the constraints from our active list. We do however allow for the subsequent reintroduction of $k_j(\boldsymbol{x})$ should it become positive. The algorithm may deal with equality and inequality constraints simultaneously, but the above test is, of course, only carried out for those λ_j which correspond to inequality constraints.

The application of the algorithm to inequality constrained problems is illustrated in $\boldsymbol{x}$-space in Fig. 5 for the following example.

Example 4. Minimize $F(\boldsymbol{x}) = (x_1 - 1)^2 + (x_2 - 2)^2$ subject to

Case (i) $x_2 + x_1 - 2 \leqslant 0$ and $x_2 - x_1 + 1 \leqslant 0$;

Case (ii) $x_2 + x_1 - 2 \leqslant 0$ and $x_2 - x_1 - 2 \leqslant 0$.

In both cases the starting point is $\boldsymbol{x}_0 = (0, 0)$, $\boldsymbol{\lambda}_0 = (0, 0)$. In Case (i) both constraints become active and the trajectory converges to $\boldsymbol{x}^* = (\frac{1}{2}, \frac{3}{2})$, $\boldsymbol{\lambda}^* = (1, 2)$. In Case (ii) only the first

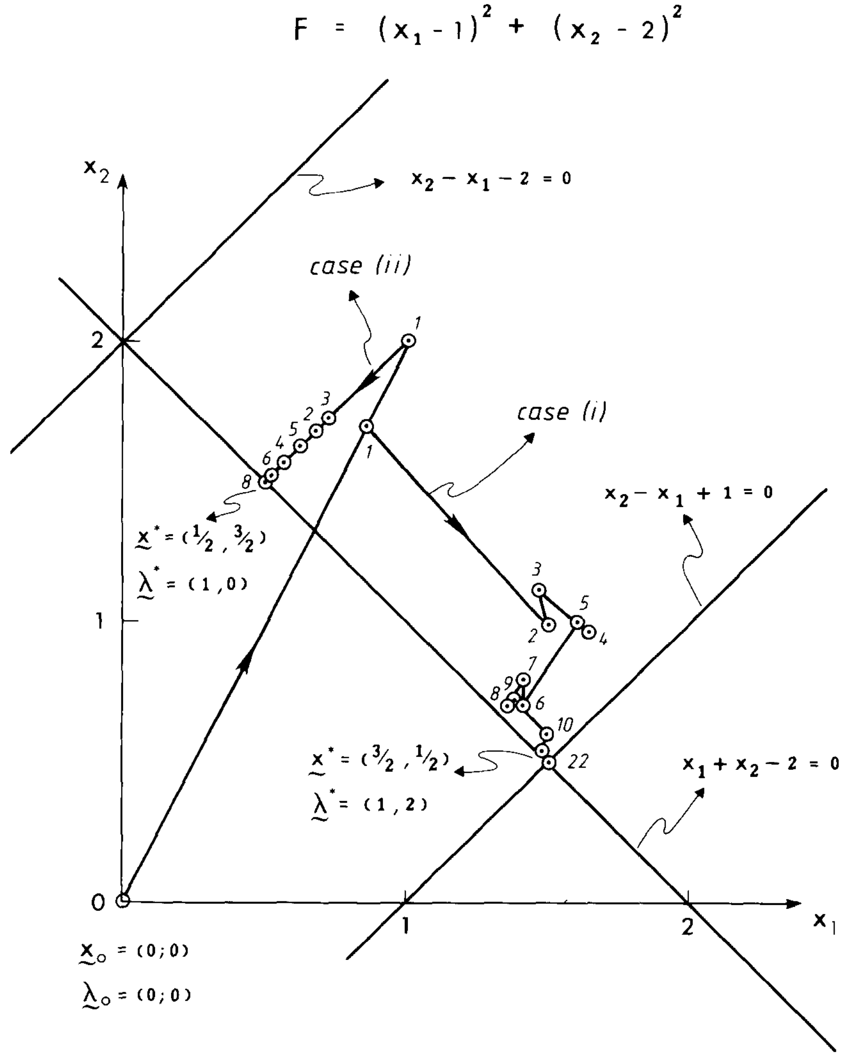

Fig. 5. Minimization trajectories for the inequality constrained problems of Example 4. In Case (i) two constraints are active at the minimum while in Case (ii) a single constraint is active at the solution.

constraint is active and we have convergence to $x^* = (\frac{1}{2}, \frac{3}{2})$, $\lambda^* = (1, 0)$. Again the trajectories depicted in Fig. 5 are representative of the algorithms general performance when applied to more complicated problems.

7. Concluding remark

If the matrix A in problem (10) is negative definite instead of positive definite then the procedure outlined in this paper still guarantees convergence to the solution provided the following modifications are applied. Instead of using search directions (12) we apply

$$dx/dt = +\nabla_x L, \qquad d\lambda/dt = -\nabla_\lambda L, \tag{35}$$

and in the place of the auxiliary Lagrangian function (24) we consider

$$L_A(\boldsymbol{x}, \lambda) = F(\boldsymbol{x}) + \lambda^T \boldsymbol{h}(\boldsymbol{x}) - \tfrac{1}{2} \| \boldsymbol{h}(\boldsymbol{x}) \|^2. \tag{36}$$

Proof of convergence, for the negative definite case with these modifications, follows simply by substituting expressions (35) and (36) in the argument outlined in Sections 2 and 3.

Acknowledgements

The author wishes to acknowledge that the idea of the "primal-dual trajectory", as defined by equation (9), was suggested by a perusal of an article by Solheim and Ali [6]. They computed a trajectory which degenerated into a closed orbit around a saddle point of a potential function $V(x, y)$, by numerically solving the initial-value problem:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\nabla_x V, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = +\nabla_y V, \qquad x(0) = x_0, \quad y(0) = y_0.$$

The author also wishes to thank Miss Alna van der Merwe of the Department of Applied Mathematics of the University of Pretoria for fruitful discussions regarding the convergence of the primal-dual trajectory. Financial assistance by the South African Council for Scientific and Industrial Research is gratefully acknowledged.

References

[1] M. Avriel, *Non-Linear Programming: Analysis and Methods* (Prentice-Hall, Englewood Cliffs, NJ, 1976).

[2] D.P. Bertsekas, Multiplier methods: a survey, *Automatica* **12** (1976) 133–145.

[3] M.R. Hestenes, Multiplier and gradient methods, *J. Opt. Theory Appl.* **4** (1969) 303–320.

[4] M.J.D. Powell, A method for nonlinear constraints in minimization problems, in: R. Fletcher, Ed., *Optimization* (Academic Press, New York, 1969).

[5] D.A. Wismer and R. Chattergy, *Introduction to Nonlinear Optimization* (North-Holland, New York, 1978).

[6] L. Solheim and M.K. Ali, Hamilton's equations with complex time and nonconservative systems, *Phys. Rev. Lett.* (1988) to appear.