JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 50, No. 3, SEPTEMBER 1986

On Differentiable Exact Penalty Functions

C. VINANTE[1] AND S. PINTOS[2]

Communicated by R. A. Tapia

Abstract. In this work, we study a differentiable exact penalty function for solving twice continuously differentiable inequality constrained optimization problems. Under certain assumptions on the parameters of the penalty function, we show the equivalence of the stationary points of this function and the Kuhn-Tucker points of the restricted problem as well as their extreme points. Numerical experiments are presented that corroborate the theory, and a rule is given for choosing the parameters of the penalty function.

Key Words. Constrained optimization, nonlinear programming, differential exact penalty functions, computational methods, augmented Lagrangian functions.

1. Introduction

Considerable attention has been given in recent years to devising methods for solving nonlinear programming problems via unconstrained minimization techniques. One class of methods which has emerged as very promising is the exact penalty function methods, which avoid the sequence of unconstrained minimization problems characteristic of the augmented Lagrangian methods. In Refs. 1-4, it is shown that, under suitable assumptions, it is possible to define a continuously differentiable function $V(x, \lambda, \epsilon, \alpha)$, whose unconstrained minima yields the solution of the constrained problem and its associated multipliers. Moreover, this function can also be used as a line search function for various direction finding algorithms. One of such possibilities was studied in Ref. 5 in relation to a recursive quadratic programming algorithm for equality constrained problems. For the equality constrained problem

$$\min f(x), \qquad \text{subject to} \quad g(x) = 0,$$

[1] Associate Professor, Posgrado de Ingenieria, Universidad del Zulia, Maracaibo, Venezuela.
[2] Assistant Professor, Department of Mathematics, Facultad de Ingenieria, Universidad del Zulia, Maracaibo, Venezuela.

0022-3239/86/0900-0479$05.00/0

where $f: R^n \to R$, $g: R^n \to R^m$, DiPillo and Grippo (Ref. 2) [see also Boggs and Tolle (Ref. 4)] studied the exact penalty function

$$S(x, \lambda, \epsilon) = f(x) + \lambda^T g(x) + (1/\epsilon)\|g(x)\|^2 + \|M(\nabla f + \nabla g \lambda)\|^2 \tag{1}$$

for two cases:

(i) $M = \sqrt{\alpha} I$, where I is the identity matrix of appropriate dimensions, so that if x is regular point, $M\nabla g$ is of rank m;

(ii) $M = \sqrt{\alpha}|\nabla g|^T$, so that $M(x)\nabla g$ is invertible.

Bertsekas (Ref. 1) studied the same functions in relation to enlarging the region of convergence of the Newton method and also showed that there is a close relation between these functions and the class of penalty functions of Fletcher. See also Tapia (Ref. 6).

Extensions to inequality constrained problems were studied by DiPillo and Grippo (Ref. 3) by converting inequality constraints to equality constraints using squared slack variables and a special choice of the matrix $M(x)$, similar to case (ii) for the equality constrained case. In this work, we study the exact differentiable penalty function with $M = \sqrt{\alpha} I$ for the inequality constrained problem. For this choice of the matrix M, $M\nabla g$ is no longer invertible as in Ref. 3; therefore, some special conditions have to be imposed on the parameters ϵ and α of the penalty function to guarantee the equivalence of its stationary points and the Kuhn-Tucker (K-T) points of the constrained problem, as well as their local minima. This choice of the matrix M is very convenient to implement the algorithms. Some preliminary numerical experiments are also presented with a standard set of problems. The results obtained are in agreement with those of Ref. 3 and the rule suggested by Ref. 1 for choosing the relation ϵ/α.

2. Exact Differentiable Penalty Function

The problem under consideration is the following nonlinear programming problem:

(P) $\min f(x)$,

subject to $g(x) \leq 0$,

where $f: R^n \to R$ and $g: R^n \to R^m$. It is assumed, unless otherwise stated, that the functions f and g are twice continuously differentiable on R^n.

We denote by $L(x, \lambda)$ the Lagrangian function

An equivalent equality constrained problem is given by

(B) $\min f(x)$,

subject to $g_j(x)+y_j^2=0$,

involving the additional vector of squared slack variables y. Now, as proposed in Ref. 1 and Ref. 2, the search for the solution of problem (B) can be converted into the unconstrained minimization with respect to x, y, λ of the augmented function

$$S(x, y, \epsilon, \alpha)=L(x,\lambda)+\sum_1^m \lambda_j y_j^2+(1/\epsilon)\sum_1^m (g_j+y_j^2)^2 + \alpha\left[\|\nabla_x L\|^2+4\sum_1^m (\lambda_j y_j)^2\right]. \tag{3}$$

This minimization can be carried out by minimizing first with respect to y and subsequently the resulting function with respect to (x, λ). A straightforward calculation shows that

$$V(x,\lambda,\epsilon,\alpha)=\min_y S=L(x,\lambda)+(1/\epsilon)\|\nabla_x L\|^2-(1/\epsilon)\|d\|^2, \tag{4}$$

where the components of the vector d are given by

$$d_j=-\min[0, g_j(x)+(\epsilon\lambda_j/2)(1+4\alpha\lambda_j)]. \tag{5}$$

It can be easily proved that, under the conditions of program (P), the function V is continuously differentiable with respect to x and λ with gradients

$$\nabla_x V(x,\lambda,\epsilon,\alpha)=(I+2\alpha\nabla_{xx}^2 L)\nabla_x L+(2/\epsilon)\nabla_x g(g+d), \tag{6}$$

$$\nabla_\lambda V(x,\lambda,\epsilon,\alpha)=(g+d)+2\alpha[\nabla_x g]^T\nabla_x L+8\alpha\Lambda d, \tag{7}$$

where

$$\Lambda=\mathrm{diag}[\lambda_1,\lambda_2,\ldots,\lambda_m].$$

3. Relation between K–T Points and Stationary Points of V

In this section, we will establish the relations between the stationary points of the function V and the Kuhn–Tucker points of problem (P).

Theorem 3.1. Let $(\bar{x}, \bar{\lambda})$ be a point satisfying the K–T necessary conditions for problem (P). Then, $(\bar{x}, \bar{\lambda})$ is a stationary point for the function $V(x, \lambda, \epsilon, \alpha)$.

Proof. At the point $(\bar{x}, \bar{\lambda})$, we have

$$\nabla_x L(\bar{x}, \bar{\lambda}) = 0, \qquad \Lambda g(\bar{x}) = 0, \tag{8}$$

and by (5) and (8) we obtain

$$g(\bar{x}) + d(\bar{x}, \bar{\lambda}, \epsilon, \alpha) = 0, \qquad \Lambda d(\bar{x}, \bar{\lambda}, \epsilon, \alpha) = 0. \tag{9}$$

Therefore, (6) and (7) with (8) and (9) yield

$$\nabla_x V(\bar{x}, \bar{\lambda}, \epsilon, \alpha) = 0,$$

$$\nabla_\lambda V(\bar{x}, \bar{\lambda}, \epsilon, \alpha) = 0,$$

which proves the theorem. □

To prove the converse of Theorem 3.1 with $M = \sqrt{\alpha}\, I$, we state first the following lemmas.

Lemma 3.1. Let Q be an $n \times n$ positive-definite matrix, $y = Qx$ with $\|x\| = 1$, and $M = \min\{x^T Q x\colon \|x\| = 1\}$. Then,

$$\max\{|y_i|\colon i = 1, 2, \ldots, n\} \geq k,$$

where $k = M/\sqrt{n}$.

Proof. Assume that $|y_i| < k$, $\forall i$. Then, for x such that $\|x\| = 1$, we have

$$x^T Q x = x^T y \leq \sum_1^m |x_i||y_i| < k \sum_1^m |x_i| \leq M.$$

Therefore, a contradiction arises. □

Lemma 3.2. Let Q be an $n \times n$ positive-definite matrix and $x(\epsilon)$ a vector whose norm is an infinitesimal of order k for $\epsilon \to 0$. Then, the vector $y = Qx(\epsilon)$ has at least one component y_i whose absolute value is also an infinitesimal of order k for $\epsilon \to 0$.

Proof. Applying the previous lemma to $x(\epsilon)/\|x\|$, there exists i such that $|y_i|/\|x\| \geq k$. Then,

$$|y_i| \geq k\|x(\epsilon)\|; \tag{10}$$

and, from properties of the norm,

$$\|y\| \leq \|Q\|\, \|x(\epsilon)\|. \tag{11}$$

From (10) and (11), the result follows. □

Lemma 3.3. Let $X \times T$ be a compact subset of $R^n \times R^m$ and assume that:

(i) the matrix $(I + 2\alpha \nabla^2_{xx} L)$ is positive definite on $X \times T$;

(ii) every feasible point in X is regular;

(iii) ϵ is infinitesimal of higher order than $\alpha(\epsilon)$, that is, $\lim_{\epsilon \to 0}(\epsilon/\alpha) = 0$.

Then, there exists $\bar{\epsilon} \geq 0$, such that, for $\epsilon \in (0, \bar{\epsilon}]$, every stationary point of $V(x, \lambda, \epsilon, \alpha)$ satisfies the relation

$$g_j + d_j = 0, \qquad j = 1, 2, \ldots, m,$$

that is,

$$g_j + \min[0, g_j + (\epsilon\lambda_j/2)(1 + 4\alpha\lambda_j)] = 0. \tag{12}$$

Proof. We will proceed by contradiction, assuming that the lemma is false, that is, there exists a decreasing sequence $\{\epsilon_k\}$ tending to zero, where $\{x_k, \lambda_k\}$ are stationary points of $V(x, \lambda, \epsilon_k, \alpha_k)$ and do not satisfy (12). Since $X \times T$ is compact, there is a subsequence of $\{x_k, \lambda_k\}$ converging to some $(\bar{x}, \bar{\lambda}) \in X \times T$. To simplify the notation, we will use the same nomenclature $\{x_k, \lambda_k\}$ for this converging subsequence. At each stationary point (x_k, λ_k) of the function V, we have from (6) and (7)

$$\nabla_x L = -(2/\epsilon_k)[I + 2\alpha_k \nabla^2_{xx} L]^{-1} \nabla_x g(g + d), \tag{13}$$

$$g + d = -2\alpha_k[(\nabla_x g)^T \mid 2\sqrt{D}][(\nabla_x L)^T \mid (2\sqrt{D}\lambda)^T]^T, \tag{14}$$

where

$$D = \mathrm{diag}[d_1, d_2, \ldots, d_m].$$

It can be easily shown that

$$2\sqrt{D}\lambda = -(2/\epsilon_k)2D(g + d) - 8\alpha_k\sqrt{D}\Lambda\lambda. \tag{15}$$

Substituting (15) and (13) in (14) yields

$$g + d = 4(\alpha_k/\epsilon_k)Q(g + d) + 32\alpha_k^2 D\Lambda\lambda, \tag{16}$$

with

$$Q = [(\nabla g)^T (I + 2\alpha_k \nabla^2_{xx} L)^{-1} \nabla g + 4D], \tag{17}$$

and this matrix, see Ref. 7, is positive definite in a neighborhood of $(\bar{x}, \bar{\lambda})$, that is, for small values of α_k. Since $D\Lambda\lambda = \Lambda^2 d$, (16) can be expressed as

$$[I - 4(\alpha_k/\epsilon_k)Q - 32\alpha_k^2\Lambda^2](g + d) = -32\alpha_k^2\Lambda^2 g, \tag{18}$$

and this expression is valid for every (x_k, λ_k). From (5), it follows that the components of the vector $g+d$ can be expressed as

$$g_j + d_j = \max[g_1(x_k), -(\epsilon_k\lambda_{jk}/2)(1+4\alpha_k\lambda_{jk})]. \tag{20}$$

As $k\to\infty$, the matrix that multiplies $g+d$ in (18) tends to 0. Then, there is a $\bar{k}$ sufficiently large and a corresponding pair $(\bar{\epsilon}, \bar{\alpha})$ such that, for $k \geq \bar{k}$, the points (x_k, λ_k) belong to the neighborhood of $(\bar{x}, \bar{\lambda})$ for which the matrix $-(\epsilon_k/\alpha_k)I + Q + 8\alpha_k\epsilon_k\Lambda^2$ is positive definite. To establish the contradiction, we need to prove that $g+d=0$ on the points of the subsequence for $k \geq \bar{k}$. The following cases will be considered.

(a) If

$$g_j(x_k) \geq (\epsilon_k\lambda_{jk}/2)(1+4\alpha_k\lambda_{jk}), \qquad \forall j, \tag{21}$$

then $g+d=g$ and, from (19),

$$[-(\epsilon_k/4\alpha_k)I + Q(x_k, \lambda_k, \epsilon_k, \alpha_k)](g+d)=0.$$

Since $|-(\epsilon_k/4\alpha_k)I + Q|$ is positive definite, $g+d=0$.

(b) If (21) is not true for every j, define the set

$$J = \{j: g_j(x_k) < (\epsilon_k\lambda_{jk}/2)(1+4\alpha_k\lambda_{jk}); \bar{\lambda}_j \neq 0\}.$$

If J is not empty, i.e., there is at least one sequence $\{\lambda_{jk}\}$ not tending to zero, the infinitesimal order of the vector $g+d$ is the same as the order of ϵ. Since Q is positive definite, by Lemma 3.2 the infinitesimal order of some elements of the left-hand side vector in (19) is of the same order as ϵ, but the corresponding elements on the right-hand side of (19) are of the form $8\alpha_k\epsilon_k\lambda_{jk}^2 g_j(x_k)$, which are of higher infinitesimal order. Therefore, a contradiction arises and $g+d=0$ for $k \geq \bar{k}$.

(c) If (21) is not true for every j and if J is empty, we partition the vector $g+d$ into two sets, set Y where the maximum in (20) is g_j, and set Z where the maximum is $-(\epsilon_k\lambda_{jk}/2)(1+4\alpha_k\lambda_{jk})$.

Considering the case where the infinitesimal order of $g+d$ is given by an element of the set Y, let it be $g_j(x_k)$. By Lemma 3.1, there is at least one component of the left-hand side vector; let it be the ith component, with infinitesimal order equal to $g_j(x_k)$. If $i \in Y$, the corresponding component of the right-hand side vector is $8\alpha_k\epsilon_k\lambda_{jk}^2 g_i(x_k)$, an infinitesimal of higher order than $g_j(x_k)$, since $|g_i(x_k)| \leq |g_j(x_k)|$. If, on the other hand $i \in Z$, then $g_j(x_k)$ is of lower infinitesimal order then $|\epsilon_k\lambda_{ik}|$ and $8\alpha_k(\epsilon_k\lambda_{ik})\lambda_{ik}^2 g_i(x_k)$; in both cases, a contradiction arises; therefore, $g+d=0$.

Following the same line of reasoning, based on comparing the infinitesimal order of the components of the vectors in (20), see Ref. 7, we can conclude that the vector $g+d$ is zero on the points of the convergent subsequence, for $k \geq \bar{k}$, which contradicts the initial assumption, and the lemma is proved. □

Based on Lemma 3.3, we prove now the following theorem.

Theorem 3.2. Let $X \times T$ be a compact subset of $R^n \times R^m$; and let the conditions of Lemma 3.3 be satisfied. Then, there exists $\bar{\epsilon} \geq 0$ and a corresponding $\bar{\alpha}(\bar{\epsilon})$, such that, for $\epsilon \in (0, \bar{\epsilon}]$, every stationary point $(\bar{x}, \bar{\lambda})$ of $V(x, \lambda, \epsilon, \alpha)$ is a K-T point of problem (P).

Proof. At $(\bar{x}, \bar{\lambda})$,

$$\nabla_x V(\bar{x}, \bar{\lambda}) = 0, \qquad g(\bar{x}) + d(\bar{x}, \bar{\lambda}, \epsilon, \alpha) = 0.$$

Then, (6) yields

$$\nabla_x L(\bar{x}, \bar{\lambda}) = 0; \tag{22}$$

and from (7) we obtain $\Lambda d = 0$, which implies that

$$\Lambda g = 0. \tag{23}$$

Since by definition $d \geq 0$, we have

$$g(\bar{x}) \leq 0. \tag{24}$$

Finally, for α such that

$$-1/4\alpha < \min(\bar{\lambda}_1, \bar{\lambda}_1, \ldots, \bar{\lambda}_m),$$

we have

$$[1 + 4\alpha\lambda_j] \geq 0. \tag{25}$$

If $\bar{\lambda}_j \neq 0$, then by (23) $g_j(\bar{x}) = 0$, which implies that

$$d_j = -\min\{0, g_j(\bar{x}) + (\epsilon\lambda_j/2)(1 + 4\alpha\lambda_j)\} = 0. \tag{26}$$

From (25) and (26), we have

$$\lambda_j \geq 0. \tag{27}$$

From (22), (23), (24), (27), the proof follows. □

The results of this section show that α should be taken small enough so that $I + 2\alpha\nabla^2_{xx}L$ is positive definite on $X \times T$, and that, for this α, ϵ should be such that the matrix $-(\epsilon/4\alpha)I + (\nabla_x g)^T(I + 2\alpha\nabla^2_{xx}L)^{-1}\nabla_x g + 4D$, where the diagonal matrix D is

$$D(j, j) = -\min[0, g_j(x) + (\epsilon\lambda_j/2)(1 + 4\alpha\lambda_j)],$$

is positive definite on $X \times T$. Under these conditions, the critical points of $V(x, \lambda, \epsilon, \alpha)$ are equivalent to the K-T points for problem (P).

4. Local Optimality Results

In this section, we will establish local optimality results considering second-order derivatives under the assumption that the functions f and g are three times continuously differentiable. With these assumptions, the first three terms in (4) are twice continuously differentiable. For the fourth term, it can be proved (see Ref. 7) that, if $(\bar{x}, \bar{\lambda})$ is a K–T point that satisfies the property of strict complementarity, then there exists a neighborhood of the point in which the function $\|d\|^2$ is twice continuously differentiable.

Establishing the following convention to distinguish between active and nonactive constraints:

$$I_A(\bar{x}) \triangleq \{j: g_j(\bar{x}) = 0\},$$

$$I_B(\bar{x}) \triangleq \{j: g_j(\bar{x}) \neq 0\},$$

and defining the diagonal matrices

$$E(j,j) = 0, \qquad j \in I_A,$$

$$E(j,j) = 1, \qquad j \in I_B,$$

and $G(j,j) = g_j(x)$, we have

$$\nabla_x d = -\nabla_x g E,$$

$$\nabla_\lambda d = -(\epsilon/2)(I + 8\alpha\Lambda)E,$$

and

$$\nabla^2_{xx} V = (\nabla^2_{xx} L)(I + 2\alpha \nabla^2_{xx} L) + (2/\epsilon)(\nabla_x g)(I - E)(\nabla_x g)^T, \tag{28}$$

$$\nabla^2_{\lambda x} V = (\nabla_x g)^T (I + 2\alpha \nabla^2_{xx} L) - E(\nabla_x g)^T, \tag{29}$$

$$\nabla^2_{\lambda\lambda} V = 2\alpha(\nabla_x g)^T \nabla_x g - (\epsilon/2)E - 8\alpha G. \tag{30}$$

We now prove the following theorems.

Theorem 4.1. Let f and g be three times continuously differentiable; and let (x^*, λ^*) be a regular K–T point for problem (P) where strict complementarity is satisfied and

$$x^T(\nabla^2_{xx} L(x^*, \lambda^*)x > 0, \qquad \forall x > 0, \text{ with } \nabla g_j(x^*)^T x = 0, j \in I_A(x^*). \tag{31}$$

Then, there exists $\bar{\epsilon}$ and the corresponding $\bar{\alpha}$ such that, for $\epsilon \in (0, \bar{\epsilon}]$, the function $V(x, \lambda, \epsilon, \alpha)$ has a strict local minimum at (x^*, λ^*).

Proof. See Ref. 1, Proposition 2.1. □

We will prove now the converse of Theorem 4.1.

Theorem 4.2. With the assumptions of Theorem 3.2 and the condition of strict complementarity for the K–T points, there exists $\bar{\epsilon}$ and the corresponding α such that, for $\epsilon \in (0, \bar{\epsilon}]$, every local minimum (x^*, λ^*) of $V(x, \lambda, \epsilon, \alpha)$ that satisfies the second-order sufficient conditions is a local minimum of problem (P) satisfying the second-order sufficient conditions

$$x^T[\nabla^2_{xx} L(x^*, \lambda^*)]x > 0, \qquad \forall x \neq 0,$$
$$\text{with } (\nabla_x g_j(x^*))^T x = 0, j \in I_A(x^*). \tag{32}$$

Proof. Let $\bar{\epsilon}$ be defined by the conditions of Theorem 3.2. Then, if $V(x, \lambda, \epsilon, \alpha)$ has a local minimum at (x^*, λ^*), this point is also a K–T for problem (P); and, since strict complementarity holds, we have only to show that $\nabla^2_{xx} L(x^*, \lambda^*)$ is positive on the tangent hyperplane. At (x^*, λ^*), we have

$$(x, \lambda)^T[\nabla^2 V(x^*, \lambda^*)](x, \lambda) > 0, \qquad \forall (x, \lambda) \neq 0. \tag{33}$$

Let x be an element of the tangent hyperplane, and let λ be such that $\lambda_j = 0$ for $j \in I_B(x^*)$. Expanding (33) and defining the vectors

$$g_a(x^*) = (g_j(x^*)), \qquad j \in I_A(x^*), \tag{34}$$
$$\lambda_b = (\lambda_j), \qquad j \in I_B(x^*), \tag{35a}$$
$$\lambda_a = (\lambda_j), \qquad j \in I_A(x^*), \tag{35b}$$

we have

$$(x, \lambda)^T[\nabla^2 V(x^*, \lambda^*)](x, \lambda)$$
$$= (2/\epsilon)\|(\nabla_x g_a)^T x\|^2 + x^T \nabla^2_{xx} L x + 2\alpha \|\nabla^2_{xx} L x + \nabla_x g_a \lambda\|$$
$$+ 2\lambda_a(\nabla_x g_a)^T x - 8\alpha\lambda^T G\lambda - (\epsilon/2)\|\lambda_b\|^2 > 0. \tag{36}$$

By the conditions of the theorem, the first, fourth, fifth, and sixth terms of the right-hand side are zero. Therefore,

$$(x, \lambda)^T[\nabla^2 V(x^*, \lambda^*)](x, \lambda)$$
$$= 2\alpha \|\nabla^2_{xx} L(x^*, \lambda^*)x + \nabla g_a(x^*, \lambda^*)\lambda_a\|^2$$
$$+ x^T[\nabla^2_{xx} L(x^*, \lambda^*)]x > 0. \tag{37}$$

Since the norm in the first term on the right is continuous, it is bounded on the compact set and this term tends to zero with α. Then, $x^T[\nabla^2_{xx} L(x^*, \lambda^*)]x$ cannot be negative. Furthermore, it cannot be zero, since this would imply that either $\nabla^2_{xx} L(x^*, \lambda^*)x = 0$ or that x and $\nabla^2_{xx} L(x^*, \lambda^*)$ are orthogonal. In the first case, taking $\lambda_a = 0$ leads to a contradiction. The second case implies that $\nabla^2_{xx} L(x^*, \lambda^*)x$ belongs to the linear combination of the gradients of the active constraints, $g_j(x), j \in I_A(x^*)$. Then, by choosing appropriate values for the vector λ_a, we can have

$$\|\nabla^2_{xx}L(x^*, \lambda^*)x + \nabla g_a \lambda_a\|^2 = 0 \quad \text{and} \quad (x, \lambda)^T[\nabla^2 V(x^*, \lambda^*)](x, \lambda) = 0;$$

a contradiction arises, and the theorem is proved. □

It can also be proved (see Ref. 7) that, if in an interior K-T point problem (P) has a global minimum, the function $V(x, \lambda, \epsilon, \alpha)$ has also a global minimum. The converse result also holds with the conditions of Theorems 3.2.

5. Numerical Experiments

In this section, we present the computational experiments done on four problems studied by Ref. 3. These experiments were performed primarily to study the algorithm in relation to the parameters ϵ and α. The problems were solved by means of a quasi-Newton algorithm using the DFP updating formula and the stopping rule

$$\|(x, \lambda)^{k+1} - (x, \lambda)^k\| \leq x_{\text{tol}} \quad \text{and} \quad \|f(x, \lambda)^{k+1} - f(x, \lambda)^k\| \leq f_{\text{tol}};$$

however, in a more complete numerical study, the use of more efficient unconstrained minimization procedures and a more modern and less scale-dependent stopping rule should be considered.

For each example, we give the optimal values, the starting point, and the error (ERROR) in the optimal value of the function. In Tables 1-4, correspondingly to each problem, we specify the number of linear searches (LS) and the number of function evaluations (FE). The cases where the algorithm failed to achieve the optimum were three: the function $V(x, \lambda, \epsilon, \alpha)$ was unbounded below (N1); the method did not converge (N2); or the method converged to a different point (N3).

As the tables show, for each value of α, there is a range of values of ϵ for which the method failed to converge. For large values of ϵ, we are probably violating the conditions of Theorem 3.2, while for small values of ϵ this behavior can be attributed to the ill-conditioning associated. The results also show that, in most of the cases, there exist convergence for $0.01 \leq \epsilon/\alpha \leq 1$ and that the best relation is $\epsilon/\alpha = 1$, as suggested in Ref. 1. In Ref. 3, similar pairs of optimal values for ϵ and α are reported for the first three problems; but, for the problem of eleven variables (Wong's problem), a relation $\epsilon/\alpha \simeq 10{,}000$ was obtained. This limited set of results suggests the feasibility of this exact penalty function for solving inequality constrained problems. The possibility of its use as a line search function in recursive quadratic programming algorithms is presently under investigation.

Table 1. Results for Example 5.1.

α	ϵ	LS	FE	ERR
1	1	N3	—	—
	0.1	11	84	1E−7
	0.01	11	75	2E−7
	0.001	15	107	2E−7
	0.0001	11	263	3E−4
	0.00001	16	2085	1E−7
0.1	1	N1	—	—
	0.1	21	189	3E−6
	0.01	10	74	1E−6
	0.001	10	96	2E−6
	0.0001	16	177	1E−7
	0.00001	21	3117	1E−7
	0.000001	N2	—	—
0.01	0.1	N1	—	—
	0.01	11	127	2E−4
	0.001	11	134	1E−4
	0.0001	13	415	1E−5
	0.00001	N2	—	—
0.001	0.01	N1	—	—
	0.001	11	124	1E−5
	0.0001	11	340	2E−4
	0.00001	11	969	1E−3
	0.000001	N2	—	—
0.0001	0.01	N1	—	—
	0.001	6	67	2E−3
	0.0001	11	194	1E−3
	0.00001	11	949	2E−3

Example 5.1. (Fiacco and McCormick, Ref. 3)

Minimize $f(X) = (1/3)(x_1+1)^3 + x_2$,

subject to $x_1 \geq 1$,

$x_2 \geq 0$.

Optimal values: $X^* = (1, 0)$, $\lambda^* = (4, 1)$, $f(X^*) = 8/3$.

Starting point: $X = (1.125, 0.125)$, $\lambda = (0, 0)$.

Table 2. Results for Example 5.2.

α	ϵ	LS	FE	ERR
1	1	N1	—	—
	0.1	25	166	1E−6
	0.01	25	294	1E−4
	0.001	34	731	4E−5
	0.0001	56	2797	1E−7
	0.00001	N2	—	—
0.1	1	N1	—	—
	0.1	18	140	4E−6
	0.01	33	220	5E−6
	0.001	45	432	2E−6
	0.0001	81	1671	1E−6
	0.00001	N2	—	—
0.01	0.1	N1	—	—
	0.01	41	328	2E−4
	0.001	49	592	4E−4
	0.0001	81	3511	4E−5
	0.00001	N2	—	—
0.001	0.1	N1	—	—
	0.01	33	274	6E−4
	0.001	49	616	7E−4
	0.0001	73	2504	3E−4
	0.00001	N2		
0.0001	0.01	N1	—	—
	0.001	73	1064	3E−4
	0.0001	74	2071	1E−3
	0.00001	N2	—	—

Example 5.2 (Rosen and Suzuki, Ref. 3)

Minimize $f(X) = -5(x_1 + x_2) + 7(x_4 - 3x_3) + x_1^2 + x_2^2 + 2x_3^2 + x_4^2,$

subject to $(x_i^2) + x_1 - x_2 + x_3 - x_4 \leq 8,$

$$x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 \leq 10,$$

$$2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 \leq 5.$$

Optimal values: $X^* = (0, 1, 2, -1)$, $\lambda^* = (1, 0, 2)$, $f(X^*) = -44$.

Starting point: $X = (0, 0, 0, 0)$, $\lambda = (0, 0, 0)$.

Table 3. Results for Example 5.3.

α	ϵ	LS	FE	ERR
1	1	49	423	6E−7
	0.1	15	109	3E−7
	0.01	25	181	4E−8
	0.001	N3	—	—
0.1	1	N1	—	—
	0.1	14	140	1E−8
	0.01	15	127	1E−8
	0.001	8	115	8E−6
	0.0001	15	287	1E−8
	0.00001	N2	—	—
0.01	0.01	7	203	2E−3
	0.001	17	196	2E−5
	0.0001	25	2730	3E−3
	0.00001	N2	—	—
0.001	0.1	N3	—	—
	0.01	9	82	1E−3
	0.001	17	151	8E−5
	0.0001	16	230	3E−4
	0.00001	33	3398	2E−4
	0.000001	N2	—	—
0.0001	0.01	N1	—	—
	0.001	N3	—	—
	0.0001	20	3986	6E−3
	0.00001	N2	—	—

Example 5.3 (Beale, Ref. 3)

Minimize $f(X) = -8x_1 - 6x_2 - 4x_3 + 2x_1^2 + 2x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 9,$

subject to $x_1 + 2x_3x_1 + 2x_3 < 3,$

$$x_1 \geq 0,$$
$$x_2 \geq 0,$$
$$x_3 \geq 0.$$

Optimal values: $X^* = (4/3, 7/9, 4/9),$

$\lambda^* = (2/9, 0, 0, 0), f(X^*) = 1/9.$

Starting point: $X = (0.5, 0.5, 0.5), \lambda = (0, 0, 0, 0).$

Convergence parameters: $x = 0.001$, $f = 0.01$

Table 4. Results for Example 5.4.

α	ϵ	LS	FE	ERR
1	0.01	61	886	6E−5
0.1	10	N1	—	—
	1	49	346	6E−5
	0.1	65	538	6E−5
	0.01	100	1635	6E−5
	0.001	140	2577	3E−4
	0.0001	N2	—	—
0.01	1	N1	—	—
	0.01	11	127	2E−4
	0.001	11	134	1E−4
	0.0001	13	415	1E−5
	0.00001	N2	—	—
0.001	0.01	N1	—	—
	0.001	11	124	1E−5
	0.0001	11	340	2E−4
	0.00001	11	969	1E−3
	0.000001	N2	—	—
0.0001	0.01	N1	—	—
	0.001	6	67	2E−3
	0.0001	11	194	1E−3
	0.00001	11	949	2E−3

Example 5.4 (Wong, Ref. 3)

Minimize $f(X) = (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7,$

subject to

$$2x_1^2 + 3x_2^4 + x_3 + 4x_4^2 + 5x_5 \le 127,$$

$$7x_1 + 3x_2 + 10x_3^2 + x_4 - x_5 \le 282,$$

$$23x_1 + x_2^2 + 6x_6^2 - 8x_7 \le 192,$$

$$4x_1^2 + x_2^2 - 3x_1x_2 + 2x_3^2 + 5x_6 - 11x_7 \le 0.$$

Optimal values: $X^* = (2.33050, 1.95137, -0.47754, 4.36573, -0.62448, 1.03813, 1.59423)$

$\lambda^* = (1.13972, 0, 0, 0.36861), f(X^*) = 680.630.$

Starting point: $X = (1, 2, 0, 4, 1, 1, 1), \lambda = (0, 0, 0, 0).$

Convergence parameters: $x_{tol} = 0.0000001, f_{tol} = 0.001.$

$\lambda^* = (2/9, 0, 0, 0), f(X^*) = 1/9$

References

1. BERTSEKAS, D. P., *Enlarging the Region of Convergence of Newton's Method for Constrained Optimization*, Journal of Optimization Theory and Applications, Vol. 36, pp. 221–251, 1982.
2. DIPILLO, G., and GRIPPO, L., *A New Class of Augmented Lagrangians in Nonlinear Programming*, SIAM Journal on Control and Optimization, Vol. 17, pp. 618–628, 1979.
3. DIPILLO, G., and GRIPPO, L., *A New Augmented Lagrangian Function for Inequality Constraints in Nonlinear Programming Problems*, Journal of Optimization Theory and Applications, Vol. 36, pp. 495–519, 1982.
4. BOGGS, P. T., and TOLLE, J. W., *Augmented Lagrangians Which Are Quadratic in the Multipliers*, Journal of Optimization Theory and Applications, Vol. 31, pp. 17–26, 1980.
5. BARTHOLOMEW-BIGGS, M. C., *The Recursive Quadratic Programming Approach to Constrained Optimization*, Paper Presented at the 2nd IFAC Workshop on Control Applications of Nonlinear Programming and Optimization, Oberpfaffenhofen, Germany, 1980.
6. TAPIA, R. A., *Diagonalized Multiplier Methods and Quasi-Newton Methods for Constrained Optimization*, Journal of Optimization Theory and Applications, Vol. 22, pp. 135–194, 1977.
7. VINANTE, C. D., and PINTOS, S., *On Exact Differentiable Penalty Functions*, Facultad de Ingenieria, Zulia University, Maracaibo, Venezuela, Report No. 12-83, 1983.