

# ON DUAL DIFFERENTIABLE EXACT PENALTY FUNCTIONS FOR EQUALITY CONSTRAINED OPTIMIZATION

Masao Fukushima  
*Kyoto University*

Eiki Yamakawa  
*Kyoto University*

Hisashi Mine  
*Kansai University*

(Received November 12, 1984; Revised June 12, 1985)

*Abstract* This paper is concerned with a differentiable exact penalty function derived by modifying the Wolfe dual of an equality constrained problem. It may be considered that this penalty function belongs to a class of general augmented Lagrangians on which other differentiable exact penalty functions are based. It is shown that this penalty function possesses an attractive property which may be enable us to use Newton like method effectively. Some numerical results are also reported.

## 1. Introduction

Much attention has been paid on differentiable exact penalty functions in constrained optimization [1,2,3,4,5,7,17]. In general, such penalty functions are constructed by performing some modification to augmented Lagrangians. Recently, Han and Mangasarian [12] have introduced a class of differentiable exact penalty functions by using the Wolfe dual of inequality constrained problems. Differentiable exact penalty functions are quite attractive from a computational viewpoint, though their expressions are somewhat involved. In fact, they overcome some disadvantages of nonsmooth exact penalty functions [11,16] when used in connection with sequential quadratic programming methods [9].

In this paper, we consider dual exact penalty functions of Han-Mangasarian type for equality constrained problems. We explicitly state conditions under which a Karush-Kuhn-Tucker (KKT) pair of the original problem

affords an isolated local maximum of the dual exact penalty function. Also, we show that this exact penalty function and those considered in [1,2,3,4,17] belong to the same class of general augmented Lagrangians containing parameters. That is, we point out that a KKT pair may actually correspond to a local minimum [1,2,4], a saddle point [3,17] or a local maximum of the general augmented Lagrangian according to the manner in which the parameters are chosen. In addition, we mention that, in a neighborhood of a solution, we can conveniently approximate Newton directions for unconstrained maximization of the dual exact penalty function by those for solving Karush-Kuhn-Tucker equations.

## 2. Local Duality Results

We consider the following equality constrained problem

$$(1) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array}$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}^1$ ,  $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , and  $m < n$ . We assume that  $f$  and  $h$  are twice continuously differentiable on  $\mathbb{R}^n$ . The Lagrangian  $L: \mathbb{R}^{n+m} \rightarrow \mathbb{R}^1$  for problem (1) is defined by

$$L(x, \mu) = f(x) + \mu^T h(x),$$

where  $\mu$  is the vector of Lagrange multipliers. The gradient and the Hessian of  $L$  with respect to  $x$  and  $\mu$  are given by

$$\begin{aligned} \nabla L(x, \mu) &= \begin{bmatrix} \nabla_x L(x, \mu) \\ \nabla_\mu L(x, \mu) \end{bmatrix} = \begin{bmatrix} \nabla_x L(x, \mu) \\ h(x) \end{bmatrix}, \\ \nabla^2 L(x, \mu) &= \begin{bmatrix} \nabla_{xx}^2 L(x, \mu) & \nabla_{x\mu}^2 L(x, \mu) \\ \nabla_{\mu x}^2 L(x, \mu) & \nabla_{\mu\mu}^2 L(x, \mu) \end{bmatrix} = \begin{bmatrix} \nabla_{xx}^2 L(x, \mu) & \nabla h(x) \\ \nabla h(x)^T & 0 \end{bmatrix}, \end{aligned}$$

respectively, where  $\nabla h(x)$  denotes the  $n \times m$  matrix whose columns are the gradients  $\nabla h_1(x), \dots, \nabla h_m(x)$ . If  $x^*$  is a local solution of (1), then, under a regularity assumption, there exist multipliers  $\mu^*$  such that  $(x^*, \mu^*)$  satisfies the first order optimality conditions [6,8]:

$$(2) \quad \nabla L(x, \mu) = 0.$$

Equations (2) are referred to as *Karush-Kuhn-Tucker (KKT) conditions*. Also,  $x^*$  and  $(x^*, \mu^*)$  are called a *KKT point* and a *KKT pair* of (1), respectively.

In the subsequent arguments, we shall often give attention to KKT pairs

satisfying the second order sufficient conditions [6,8] :

$$(3) \quad y^T \nabla_{xx}^2 L(x^*, \mu^*) y > 0, \quad 0 \neq \forall y \in \{ y \mid \nabla h(x^*)^T y = 0, \quad y \in R^n \}$$

and the regularity condition :

$$(4) \quad \nabla h_i(x^*), \quad i = 1, \dots, m, \quad \text{are linearly independent.}$$

For problem (1), we can formally define the Wolfe dual [10,14] as follows:

$$(5) \quad \begin{array}{ll} \text{maximize} & L(x, \mu) \\ \text{subject to} & \nabla_x L(x, \mu) = 0. \end{array}$$

It is well known that, if  $f$  is convex and  $h$  is affine, then global duality holds for problems (1) and (5) [14, Chapter 8]. For general cases, Fujiwara, Han and Mangasarian [10] have recently established the following local duality results: For a KKT pair  $(x^*, \mu^*)$  of (1), if  $\nabla_{xx}^2 L(x^*, \mu^*)^{-1}$  is positive definite on the subspace spanned by  $\nabla h_i(x^*)$ ,  $i = 1, \dots, m$ , then  $(x^*, \mu^*)$  satisfies the second order sufficient conditions for the dual problem (5) [10, Theorem 2.2]. In general, however, this assumption does not appear to be fulfilled for nonconvex programming problems, as illustrated in the next example.

Example. Consider the following nonconvex programming problem

$$(6) \quad \begin{array}{ll} \text{minimize} & f(x_1, x_2) = \frac{1}{2} (x_1^2 - x_2^2) \\ \text{subject to} & h(x_1, x_2) = x_2 = 0. \end{array}$$

The Lagrangian  $L$  for problem (6) is defined by

$$L(x_1, x_2, \mu) = \frac{1}{2} (x_1^2 - x_2^2) + \mu x_2.$$

It is clear that  $x_1^* = x_2^* = 0$  is the unique optimal solution of (6) and  $\mu^* = 0$  is the corresponding Lagrange multipliers. But we have

$$w^T \nabla_{xx}^2 L(x_1^*, x_2^*, \mu^*)^{-1} w = -z^2 < 0, \quad \text{for } \forall w = \nabla h(x_1^*, x_2^*) z, \quad 0 \neq z \in R^1.$$

In fact, for problem (6), the Wolfe dual is defined by

$$(7) \quad \begin{array}{ll} \text{maximize} & \frac{1}{2} (x_1^2 - x_2^2) + \mu x_2 \\ \text{subject to} & x_1 = 0, \quad -x_2 + \mu = 0. \end{array}$$

Since problem (7) is equivalent to the problem

$$\text{maximize } \frac{1}{2} x_2^2, \quad (x_1, x_2) \in R^2$$

and its optimal value is  $+\infty$ , there is a duality gap between problems (6) and (7).  $\square$

Now we construct a modified Wolfe dual of problem (1), for which the local duality holds under more natural assumptions. Let us consider the following problem

$$(8) \quad \begin{aligned} & \text{minimize} && f(x) + \frac{t}{2} \|h(x)\|^2 \\ & \text{subject to} && h(x) = 0, \end{aligned}$$

where  $t$  is a nonnegative parameter. It is obvious that problem (8) is equivalent to problem (1) for any  $t$ . If  $(x^*, \mu^*)$  is a KKT pair of problem (8), then  $(x^*, \mu^*)$  is also a KKT pair of (1), and vice versa [13, p.321]. The Lagrangian  $L_t : R^{n+m} \rightarrow R^1$  for problem (8) is given by

$$L_t(x, \mu) = f(x) + \frac{t}{2} \|h(x)\|^2 + \mu^T h(x) = L(x, \mu) + \frac{t}{2} \|h(x)\|^2.$$

This function is well known as the *augmented Lagrangian* [2,3,9,13], which is frequently used as a means of resolving duality gap in nonconvex programming. Then we can write the Wolfe dual for problem (8) as follows:

$$(9) \quad \begin{aligned} & \text{maximize} && L_t(x, \mu) \\ & \text{subject to} && \nabla_x L_t(x, \mu) = 0. \end{aligned}$$

Because problem (8) is equivalent to (1), we can regard problem (9) as a dual problem of (1) for any  $t \geq 0$ .

We now proceed to establish local duality results for problems (1) and (9). First of all, we shall begin with the following simple theorem.

**Theorem 1.** (a) If  $(x^*, \mu^*)$  is a KKT pair of (1), then  $(x^*, \mu^*)$  is a KKT point of (9). (b) Let  $(x^*, \mu^*)$  be a KKT point of (9) and assume that  $\nabla_{xx}^2 L_t(x^*, \mu^*)$  is nonsingular. Then  $(x^*, \mu^*)$  is a KKT pair of (1).  $\square$

The proof of this theorem is similar to that of [12, Theorem 1], and hence is omitted here.

The next local duality theorem will play an important role in validating the dual exact penalty function approach to be developed in the subsequent section.

**Theorem 2.** Let  $(x^*, \mu^*)$  is a KKT pair of (1) satisfying the second order sufficient conditions (3) and the regularity condition (4). Then there exists a  $t^* \geq 0$  such that, for any  $t \geq t^*$ ,  $(x^*, \mu^*)$  is an isolated local maximum of problem (9) and  $f(x^*) = L_t(x^*, \mu^*)$ .

Proof: By Theorem 1 - (a),  $(x^*, \mu^*)$  is a KKT point of (9). By (3), it follows that there exists a  $t^* \geq 0$  such that, for any  $t \geq t^*$ ,  $\nabla_{xx}^2 L_t(x^*, \mu^*)$  is positive definite [13, p.321]. Hence  $\nabla_{xx}^2 L_t(x^*, \mu^*)^{-1}$  exists and is also positive definite. Therefore, by (4), we have

$$w = \nabla h(x^*)z, \quad 0 \neq z \in R^m \quad \rightarrow \quad w^T \nabla_{xx}^2 L_t(x^*, \mu^*)^{-1} w > 0.$$

This is equivalent to the standard second order sufficient conditions for problem (9) [10, Theorem 2.2]. Thus  $(x^*, \mu^*)$  is an isolated local maximum of (9) for any  $t \geq t^*$ . The last assertion of the theorem follows immediately from the fact that  $h(x^*) = 0$ .  $\square$

### 3. Dual Differentiable Exact Penalty Functions

For the equality constrained problem (1), our penalty function is defined by constructing an exterior penalty function [6] for problem (9) as follows

$$(10) \quad \theta(x, \mu; t, \gamma) = L_t(x, \mu) - \frac{\gamma}{2} \|\nabla_x L_t(x, \mu)\|^2,$$

where  $\gamma$  is a positive parameter. The penalty problem is then

$$(11) \quad \text{maximize } \theta(x, \mu; t, \gamma), \quad (x, \mu) \in R^{n+m}.$$

Note that Han-Mangasarian penalty function [12] is a special case ( $t = 0$ ) of (10). As in [12], we can write the gradient of the penalty function (10) with respect to  $x$  and  $\mu$  as

$$(12) \quad \nabla \theta(x, \mu; t, \gamma) = \begin{pmatrix} \{I_n - \gamma \nabla_{xx}^2 L_t(x, \mu)\} \nabla_x L_t(x, \mu) \\ h(x) - \gamma \nabla h(x)^T \nabla_x L_t(x, \mu) \end{pmatrix}.$$

Moreover, if  $(x^*, \mu^*)$  is a KKT pair of (1), then the Hessian of (10), evaluated at  $(x^*, \mu^*)$ , may be given as

$$(13) \quad \nabla^2 \theta(x^*, \mu^*; t, \gamma) = \begin{pmatrix} \nabla_{xx}^2 L_t(x^*, \mu^*) \{I_n - \gamma \nabla_{xx}^2 L_t(x^*, \mu^*)\} & \{I_n - \gamma \nabla_{xx}^2 L_t(x^*, \mu^*)\} \nabla h(x^*) \\ \nabla h(x^*)^T \{I_n - \gamma \nabla_{xx}^2 L_t(x^*, \mu^*)\} & -\gamma \nabla h(x^*)^T \nabla h(x^*) \end{pmatrix},$$

where  $I_n$  is the unit matrix of dimension  $n$ . In the subsequent discussion, it will be convenient to rewrite (10), (12) and (13) respectively as follows:

$$\theta(x, \mu; t, \gamma) = L(x, \mu) + \frac{1}{2} \nabla L(x, \mu)^T K[x; t, \gamma] \nabla L(x, \mu),$$

$$(14) \quad \nabla\theta(x, \mu; t, \gamma) = \{I_{n+m} + \frac{1}{2} \nabla^2 L(x, \mu) K[x; t, \gamma] + \frac{1}{2} \nabla [K[x; t, \gamma] \nabla L(x, \mu)]\} \nabla L(x, \mu),$$

$$(15) \quad \nabla^2\theta(x^*, \mu^*; t, \gamma) = \{I_{n+m} + \nabla^2 L(x^*, \mu^*) K[x^*; t, \gamma]\} \nabla^2 L(x^*, \mu^*),$$

where

$$K[x; t, \gamma] = \begin{pmatrix} -\gamma I_n & 0 \\ 0 & t I_m \end{pmatrix} - \gamma \begin{pmatrix} 0 & t \nabla h(x) \\ t \nabla h(x)^T & t^2 \nabla h(x)^T \nabla h(x) \end{pmatrix}.$$

Note that these expressions are closely related to those of Bertsekas [1,2], although the definition of matrix  $K$  is quite different.

The next theorem, which corresponds to [12, Theorem 1], shows the relationship between the first order optimality conditions for problems (1) and (11).

**Theorem 3.** (a) If  $(x^*, \mu^*)$  is a KKT pair of (1), then  $\nabla\theta(x^*, \mu^*; t, \gamma) = 0$  for any  $t$  and  $\gamma$ . (b) Let  $\nabla\theta(x^*, \mu^*; t, \gamma) = 0$  and assume that  $\gamma \neq 0$  and  $\frac{1}{\gamma}$  is not an eigenvalue of  $\nabla_{xx}^2 L_t(x^*, \mu^*)$ . Then  $(x^*, \mu^*)$  is a KKT pair of (1).

**Proof:** Part (a) is obvious from (14). Let  $\nabla\theta(x^*, \mu^*; t, \gamma) = 0$ . Then, it follows from (12) that

$$\begin{aligned} \{I_n - \gamma \nabla_{xx}^2 L_t(x^*, \mu^*)\} \nabla_x L_t(x^*, \mu^*) &= 0, \\ h(x^*) - \gamma \nabla h(x^*)^T \nabla_x L_t(x^*, \mu^*) &= 0. \end{aligned}$$

Since

$$\nabla_x L_t(x^*, \mu^*) = \nabla_x L(x^*, \mu^*) + t \nabla h(x^*) h(x^*),$$

it is easy to prove (b) under the hypothesis of the theorem. □

In the next theorem, we establish conditions under which a KKT pair of (1) affords an isolated local maximum of the penalty problem (11). This theorem may be considered as an extension of Han-Mangasarian [12, Theorem 2] in which similar results are proved by assuming the positive definiteness of  $\nabla_{xx}^2 L(x^*, \mu^*)$ .

**Theorem 4.** Let  $(x^*, \mu^*)$  be a KKT pair of (1) satisfying the second order sufficient conditions (3) and the regularity condition (4). Then there exist a  $t^* \geq 0$  and, for each  $t \geq t^*$ , a scalar  $\gamma^*(t) > 0$  such that  $\nabla^2\theta(x^*, \mu^*; t, \gamma)$  is negative definite and  $(x^*, \mu^*)$  is an isolated local maximum of (11) for any  $\gamma > \gamma^*(t)$ .

Proof: By Theorem 3 - (a),  $(x^*, \mu^*)$  satisfies the first order optimality conditions for problem (11). Note that, under conditions (3) and (4), there exists a  $t^* \geq 0$  such that  $\nabla_{xx}^2 L_t(x^*, \mu^*)$  is positive definite for any  $t \geq t^*$ . Let  $\gamma^*(t) > 0$  denote the minimum eigenvalue of  $\nabla_{xx}^2 L_t(x^*, \mu^*)$ . Then, we can show in a way similar to [12, Theorem 2] that  $\nabla^2 \theta(x^*, \mu^*; t, \gamma)$  is negative definite for each  $\gamma > \gamma^*(t)$ . This completes the proof.  $\square$

Unfortunately, the converse of Theorem 4 does not necessarily hold as depicted in the following example.

Example. Consider the problem

$$(16) \quad \begin{aligned} & \text{minimize} && f(x_1, x_2) = x_1^2 + (x_2 - 1)^2 \\ & \text{subject to} && h(x_1, x_2) = x_1^2(x_1 + 2) - x_2 = 0. \end{aligned}$$

It is easy to verify that

$$(x_1^*, x_2^*, \mu^*) = (0, 0, -2)$$

is a KKT pair which corresponds to a local maximum of (16). However, by Theorem 3 - (a), it satisfies the first order optimality conditions for the penalty problem given by (11). Moreover, since

$$\nabla_{xx}^2 L_t(x_1^*, x_2^*, \mu^*) = \begin{bmatrix} -6 & 0 \\ 0 & t+2 \end{bmatrix}, \quad \nabla h(x_1^*, x_2^*) = \begin{bmatrix} 0 \\ -1 \end{bmatrix},$$

and

$$\nabla^2 \theta(x_1^*, x_2^*; t, \gamma) = \begin{bmatrix} -6 - 36\gamma & 0 & 0 \\ 0 & (t+2) - \gamma(t+2)^2 & \gamma(t+2) - 1 \\ 0 & \gamma(t+2) - 1 & -\gamma \end{bmatrix},$$

it can be shown that, for each  $t > 0$  and sufficiently large  $\gamma$ ,  $\nabla^2 \theta(x_1^*, x_2^*, \mu^*; t, \gamma)$  is negative definite. Therefore, for such  $t$  and  $\gamma$ ,  $(x_1^*, x_2^*, \mu^*) = (0, 0, -2)$  is an isolated local maximum of the penalty problem.  $\square$

This example shows that indefinite  $\nabla_{xx}^2 L_t(x^*, \mu^*)$  may also lead to negative definite  $\nabla^2 \theta(x^*, \mu^*; t, \gamma)$ . In other words, for any  $t \geq 0$  and sufficiently large  $\gamma$ , every KKT pair of (1) is a local maximum of  $\theta$ , even if it does not correspond to the local minimum of (1). Thus in order to establish the converse of Theorem 4, we need a rather restrictive hypothesis that  $\nabla_{xx}^2 L_t$  is positive definite at a local maximum of problem (11).

Theorem 5. Let  $(x^*, \mu^*)$  be a local maximum of (11) and assume that

$$(i) \quad \nabla_{xx}^2 L_t(x^*, \mu^*) \text{ is positive definite;}$$

(ii)  $\gamma \neq 0$  and  $\frac{1}{\gamma}$  is not an eigenvalue of  $\nabla_{xx}^2 L_t(x^*, \mu^*)$ .

Then  $x^*$  is an isolated local minimum of (1) with corresponding multipliers  $\mu^*$ .

Proof: By Theorem 3 - (b),  $(x^*, \mu^*)$  is a KKT pair of (1), so that  $h(x^*) = 0$  and

$$\nabla_{xx}^2 L_t(x^*, \mu^*) = \nabla_{xx}^2 L(x^*, \mu^*) + t \nabla h(x^*) \nabla h(x^*)^T.$$

Then, it is straightforward to show that the second order sufficient conditions (3) for problem (1) hold under assumption (i).  $\square$

In summary, any local minimum of problem (1) satisfying the second order sufficient conditions affords an isolated local maximum of the penalty problem (11). Conversely, every local maximum of (11) is a KKT pair of (1), but if  $\nabla_{xx}^2 L_t(x^*, \mu^*)$  is not positive definite,  $x^*$  may not be a local minimum of (1). This fact appears to delimit the use of the dual differentiable exact penalty functions for nonconvex problems, as compared with the differentiable exact penalty functions such as [1,2,3,4,5,17] for which sufficiency of penalty problems is guaranteed under less restrictive assumptions.

#### 4. Relation with Other Penalty Functions

Various differentiable exact penalty functions introduced by Di Pillo - Grippo [4], Bertsekas [1,2] and Boggs - Tolle [3] are based on augmented Lagrangian functions similar to  $\theta(x, \mu; t, \gamma)$ . Bertsekas [1, Proposition 2.1 (b)] shows that, for every  $\gamma < 0$ , a KKT pair  $(x^*, \mu^*)$  of (1) is a strict local *minimum* of  $\theta$  for sufficiently large  $t$ . On the other hand, Boggs and Tolle [3, Corollary 2.1] show that, for sufficiently large  $t$ , there exists a scalar  $\gamma^*(t) > 0$  such that  $(x^*, \mu^*)$  is a *saddle point* of  $\theta$  for any  $\gamma \in (0, \gamma^*(t)]$ . Moreover, from this property, they construct a differentiable exact penalty function which relates to the primal variables  $x$  only. It is also noted that Zhadan [17] considers, in a rather general context, an approach similar to [3] and proposes a method of finding a saddle point of the augmented Lagrangian directly.

In the previous section, we have demonstrated that, for sufficiently large  $t$ , there exists a scalar  $\gamma^*(t) > 0$  such that  $(x^*, \mu^*)$  is a strict local *maximum* of  $\theta$  for any  $\gamma > \gamma^*(t)$ . Consequently, we may observe that the difference of these three types of penalty functions merely corresponds to that of the choice of penalty parameters.



## 5. Newton Directions

In this section, we briefly mention that Newton directions obtained from the Karush-Kuhn-Tucker equations for the original problem asymptotically approach those which appear in unconstrained maximization of the penalty function  $\theta$ . This property has also been observed by Bertsekas [1,2] and Boggs-Tolle [3] for their penalty functions.

The Newton's method for finding a point which satisfies the Karush-Kuhn-Tucker conditions (2) for problem (1) is given by the iteration

$$(17) \quad d^{(k)} = - [\nabla^2 L(x^{(k)}, \mu^{(k)})]^{-1} \nabla L(x^{(k)}, \mu^{(k)}) ,$$

$$(18) \quad z^{(k+1)} = z^{(k)} + d^{(k)} ,$$

where  $z^{(k)} = (x^{(k)}, \mu^{(k)})$ . If  $(x^*, \mu^*)$  is a KKT pair of (1) satisfying the second order sufficient conditions (3), then  $\nabla^2 L(x^*, \mu^*)$  is nonsingular [13, p.231]. Therefore, there exists a neighborhood of  $(x^*, \mu^*)$  in which the Newton iteration (17) - (18) is well defined and converges to  $(x^*, \mu^*)$  quadratically [15, p.312].

Now define

$$P(x, \mu; t, \gamma) = I + \frac{1}{2} \nabla^2 L(x, \mu) K[x; t, \gamma] + \frac{1}{2} \nabla [K[x; t, \gamma] \nabla L(x, \mu)] .$$

Then, from (14), the gradient of  $\theta$  may be written as

$$(19) \quad \nabla \theta(x, \mu; t, \gamma) = P(x, \mu; t, \gamma) \nabla L(x, \mu) .$$

For a KKT pair  $(x^*, \mu^*)$  of (1), we can easily verify that

$$P(x^*, \mu^*; t, \gamma) = I + \frac{1}{2} \nabla^2 L(x^*, \mu^*) K[x^*; t, \gamma] .$$

Hence, from (15), the Hessian of  $\theta$  can be evaluated at  $(x^*, \mu^*)$  as

$$(20) \quad \nabla^2 \theta(x^*, \mu^*; t, \gamma) = P(x^*, \mu^*; t, \gamma) \nabla^2 L(x^*, \mu^*) .$$

By Theorem 4, if  $(x^*, \mu^*)$  is a KKT pair of (1) satisfying the second order sufficient conditions (3), then there exist a  $t^* \geq 0$  and, for each  $t \geq t^*$ , a scalar  $\gamma^*(t) > 0$  such that  $\nabla^2 \theta(x^*, \mu^*; t, \gamma)$  is negative definite for any  $\gamma > \gamma^*(t)$ . Hence, by (20),  $P(x^*, \mu^*; t, \gamma)$  is also nonsingular for  $t \geq t^*$  and  $\gamma > \gamma^*(t)$ , because  $\nabla^2 L(x^*, \mu^*)$  is nonsingular. By continuity, it then follows that, for  $t \geq t^*$  and  $\gamma > \gamma^*(t)$ , there exists a neighborhood of  $(x^*, \mu^*)$  in which both  $\nabla^2 L(x, \mu)$  and  $P(x, \mu; t, \gamma)$  are nonsingular. Since, in this neighborhood, the matrix

$$(21) \quad Q(x, \mu; t, \gamma) = [\nabla^2 L(x, \mu)]^{-1} [P(x, \mu; t, \gamma)]^{-1}$$

is well defined for  $t \geq t^*$  and  $\gamma > \gamma^*(t)$ , we have

$$(22) \quad d^{(k)} = - Q(x^{(k)}, \mu^{(k)}; t, \gamma) \nabla \theta(x^{(k)}, \mu^{(k)}; t, \gamma)$$

from (17), (19) and (21). Note that (20) and (21) imply

$$(23) \quad Q(x^*, \mu^*; t, \gamma) = [\nabla^2 \theta(x^*, \mu^*; t, \gamma)]^{-1}.$$

Consequently, we may conclude from (22) and (23) that the direction  $d^{(k)}$  asymptotically approaches the Newton direction

$$\tilde{d}^{(k)} = - [\nabla^2 \theta(x^{(k)}, \mu^{(k)}; t, \gamma)]^{-1} \nabla \theta(x^{(k)}, \mu^{(k)}; t, \gamma)$$

for solving the penalty problem (11).

This fact suggests that using the directions  $d^{(k)}$  in unconstrained maximization of  $\theta$  may still result in a very rapid convergence to a solution. Note that the directions  $d^{(k)}$  are much more tractable than the pure Newton directions  $\tilde{d}^{(k)}$  for penalty problem (11) because evaluation of the Hessian  $\nabla^2 \theta$  requires the third derivatives of  $L$ . Of course, the directions  $d^{(k)}$  may well approximate the directions  $\tilde{d}^{(k)}$  only in a neighborhood of the solution. By incorporating the steepest descent method, Bertsekas [1,2] proposes a globally convergent algorithm for minimizing the differentiable exact penalty function of Di Pillo - Grippo type. Since every prerequisite for Bertsekas' algorithm remains in force in the present case, it is also possible to construct a similar algorithm for the penalty problem (11), which is globally and quadratically convergent. That is, Lagrange Newton direction (17) is used to obtain the next iterates  $z^{(k+1)}$  as  $z^{(k+1)} = z^{(k)} + \alpha^{(k)} d^{(k)}$  if

$$d^{(k)T} \nabla \theta(x^{(k)}, \mu^{(k)}; t, \gamma) \geq \delta \|\nabla \theta(x^{(k)}, \mu^{(k)}; t, \gamma)\|^3,$$

where  $\delta > 0$  is a very small parameter. Otherwise, the search direction is chosen as the steepest ascent direction  $\nabla \theta(x^{(k)}, \mu^{(k)}; t, \gamma)$ . The step size  $\alpha^{(k)}$  is determined by the standard Armijo's step size rule using  $\theta$  as a criterion function. Some numerical results are shown in the next section.

As in Bertsekas [1,2] and Di Pillo - Grippo [4], we can consider slightly generalized penalty functions in which the penalty term is replaced by  $\|M(x) \nabla_{x,t} L(x, \mu)\|^2$  where  $M(x)$  is an appropriate weight matrix. We should like to point out that we may also extend the preceding discussions to such generalized penalty functions.

## 6. Numerical Results

The method described in the previous section has been applied to several problems. We present here the numerical results for the following problems.

## Problem 1.

$$\begin{aligned} &\text{minimize} && x_1 \\ &\text{subject to} && x_1^2 + x_2^2 = 1 \end{aligned}$$

Optimal solution :  $x^* = (-1, 0)$

Optimal Lagrange multiplier :  $\mu^* = 0.5$

## Problem 2.

$$\begin{aligned} &\text{minimize} && x_1(x_1 - 1)^2 + x_2^4 \\ &\text{subject to} && x_1 - 1 + \cos x_2 = 0 \end{aligned}$$

Optimal solution :  $x^* = (0, 0)$

Optimal Lagrange multiplier :  $\mu^* = -1$

## Problem 3.

$$\begin{aligned} &\text{minimize} && (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^4 + (x_5 - 1)^6 \\ &\text{subject to} && x_1^2 x_4 + \sin(x_4 - x_5) - 2\sqrt{2} = 0 \\ &&& x_2 + x_3^4 x_4^2 - 8 - \sqrt{2} = 0 \end{aligned}$$

Optimal solution :  $x^* = (1.1661, 1.1821, 1.3802, 1.5060, 0.6109)$

Optimal Lagrange multipliers :  $\mu^* = (-0.08553, -0.03187)$

## Problem 4.

$$\begin{aligned} &\text{minimize} && (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^4 + (x_4 - x_5)^4 \\ &\text{subject to} && x_1 + x_2^2 + x_3^3 - 2 - 3\sqrt{2} = 0 \\ &&& x_2 - x_3^2 + x_4 + 2 - 2\sqrt{2} = 0 \\ &&& x_1 x_5 - 2 = 0 \end{aligned}$$

Optimal solution :  $x^* = (1.1911, 1.3626, 1.4728, 1.6350, 1.6791)$

Optimal Lagrange multipliers :  $\mu^* = (-0.03882, -0.01674, -0.0002873)$

Problem 1 with the equality constraint being replaced by the inequality has been solved by Bertsekas [1] using a primal differentiable exact penalty function. Problem 2 is an example where  $\nabla_{xx}^2 L(x^*, \mu^*)$  is not positive definite. Problems 3 and 4 are the Examples 3 and 4, respectively, of Di Pillo and Grippo [4].

For Problem 1, we set the penalty parameter  $\gamma = 8$  and tested the method using several starting points for  $t = 0$  and  $t = 2$ . For Problems 2, 3 and

4, we selected the same starting point

$$\begin{aligned} x_i^{(0)} &= 2, & i &= 1, \dots, n, \\ \mu_j^{(0)} &= 0, & j &= 1, \dots, m, \end{aligned}$$

as in [4], and carried out the experiments for several values of  $t$  and  $\gamma$ . The computational results are shown in Tables 1-4. In the Tables, GDR, NDR and TOTAL represent the number of steepest ascent iterations, the number of Lagrange Newton iterations and the total number of iterations, respectively, and the symbol \* indicates that convergence to the solution was not obtained within 30 iterations.

Note that, in Table 1, the starting point  $(-0.8, -0.6)$  is on the constraint boundary of Problem 1. In such a case, the unit step size usually increases the value of nonsmooth exact penalty functions when the penalty constant is chosen very large, and convergence occurs at a very slow rate [9]. However, Table 1 shows that the differentiable exact penalty functions do not suffer from such a difficulty. On the other hand, it has been observed that, when a starting point is chosen near the global maximum  $x^* = (1, 0)$  or an initial estimate of Lagrange multiplier is set at a negative value, the generated sequence sometimes tends toward the global maximum or jams at a nonoptimal point.

As shown in Table 2, when  $t = 0$ , the method fails to converge to the solution of Problem 2. This is because  $\nabla_{xx}^2 L(x^*, \mu^*)$  is not positive definite, and hence there is a duality gap between the original problem and its Wolfe dual (see Section 2). Thus, dual differentiable exact penalty functions of Han-Mangasarian type [12], which correspond to the case  $t = 0$ , are only suitable to problems with positive definite  $\nabla_{xx}^2 L(x^*, \mu^*)$ . By virtue of Theorem 4, this difficulty could be overcome by selecting the value of  $t$  large enough. In fact, the results shown in Table 2 indicate that the value of  $t = 2$  is sufficient to obtain convergence.

Tables 3 and 4 reveal that all tested values of  $t$  (including  $t = 0$ ) resulted in convergence to the solution for Problems 3 and 4. Moreover, it is seen that the number of iterations required tends to increase as the value of  $t$  becomes large. The reason for this seems to be that, for those problems where Lagrangian itself is locally convex with respect to  $x$  at  $(x^*, \mu^*)$ , adding the term  $\frac{t}{2} \|h(x)\|^2$  usually makes conditions of the problems worse. Nevertheless, rapid convergence has been obtained for each problem in comparison with the results reported by Di Pillo and Grippo [4] who applied the conjugate gradient method to minimize their exact penalty function directly.

Table 1. Results for Problem 1

Starting point $(x_1^{(0)}, x_2^{(0)}, \mu^{(0)})$	Case (i) $t = 0, \gamma = 8$			Case (ii) $t = 2, \gamma = 8$		
	GDR	NDR	TOTAL	GDR	NDR	TOTAL
(-1.1, -0.1, -0.5)	0	4	4	0	4	4
(-0.8, -0.6, 0)	1	4	5	1	10	11
(-0.2, -0.2, 10)	1	7	8	1	12	13
(-5, -5, -0.5)	1	7	8	25	5	30
(10, 10, 10)	*	*	*	6	13	19

Table 2. Results for Problem 2

$t$	$\gamma$	GDR	NDR	TOTAL
0	16	*	*	*
2	16	4	5	9
2	4	3	6	9
4	4	7	5	12

Table 3. Results for Problem 3

$t$	$\gamma$	GDR	NDR	TOTAL
0	16	0	8	8
0.25	16	2	7	9
0.25	32	3	7	10
1	16	3	12	15

Table 4. Results for Problem 4

$t$	$\gamma$	GDR	NDR	TOTAL
0	4	0	3	3
2	4	1	4	5
2	16	2	5	7
4	4	2	5	7

## 7. Conclusion

We have considered the dual differentiable exact penalty functions in the context of general augmented Lagrangians on which several existing exact penalty functions are based. In particular, we have pointed out that various types of smooth exact penalty functions own some significant properties in common and differ only in the manner of choosing penalty parameters of the augmented Lagrangian.

## 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 (1982), 221-252.
- [2] Bertsekas, D.P. : *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982.
- [3] Boggs, P.T. and Tolle, J.W. : Augmented Lagrangians Which are Quadratic in the Multiplier. *Journal of Optimization Theory and Applications*, Vol. 31 (1980), 17-26.
- [4] Di Pillo, G. and Grippo, L. : A New Class of Augmented Lagrangians in Nonlinear Programming. *SIAM Journal on Control and Optimization*, Vol. 17 (1979), 618-628.
- [5] Di Pillo, G. and Grippo, L. : A Continuously Differentiable Exact Penalty Function for Nonlinear Programming Problems with Inequality Constraints. *SIAM Journal on Control and Optimization*, Vol. 23 (1985), 72-84.
- [6] Fiacco, A.V. and McCormick, G.P. : *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley & Sons, New York, 1968.
- [7] Fletcher, R. : A Class of Methods for Nonlinear Programming with Termination and Convergence Properties. *Integer and Nonlinear Programming*, (ed. J. Abadie). North-Holland, Amsterdam, 1970, 157-175.
- [8] Fletcher, R. : *Practical Methods of Optimization, Volume 2: Constrained Optimization*. John Wiley & Sons, Chichester, 1981.
- [9] Fletcher, R. : Penalty Functions. *Mathematical Programming; The State of the Art, Bonn, 1982*, (eds. A. Bachem, M. Grötschel and B. Korte). Springer-Verlag, Berlin, 1983, 87-114.
- [10] Fujiwara, O., Han, S.-P. and Mangasarian, O.L. : Local Duality of Nonlinear Programs. *SIAM Journal on Control and Optimization*, Vol. 22 (1984), 162-169.

- [11] Han, S.-P. and Mangasarian, O.L. : Exact Penalty Functions in Nonlinear Programming. *Mathematical Programming*, Vol. 17 (1979), 251-269.
- [12] Han, S.-P. and Mangasarian, O.L. : A Dual Differentiable Exact Penalty Function. *Mathematical Programming*, Vol. 25 (1983), 293-306.
- [13] Luenberger, D.G. : *Introduction to Linear and Nonlinear Programming*. Addison-Wesley, Reading, Massachusetts, 1973.
- [14] Mangasarian, O.L. : *Nonlinear Programming*. McGraw-Hill, New York, 1969.
- [15] Ortega, J.M. and Rheinboldt, W.C. : *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.
- [16] Zangwill, W.I. : Nonlinear Programming via Penalty Functions. *Management Science*, Vol. 13 (1967), 344-358.
- [17] Zhadan, V.G. : Modified Lagrange Functions in Nonlinear Programming. *USSR Computational Mathematics and Mathematical Physics* (English Translation), Vol. 22, No. 2 (1982), 43-57.

Masao FUKUSHIMA : Department of  
Applied Mathematics and Physics,  
Faculty of Engineering, Kyoto  
University, Kyoto, 606 Japan