ZESZYTY NAUKOWE UNIWERSYTETU JAGIELLOŃSKIEGO
DCLXI UNIVERSITATIS IAGELLONICAE ACTA MATHEMATICA, FASCICULUS XXIV 1984

The Lagrange Multiplier Functions in the Equation Approach to Constrained Optimization

by Nguyen Dinh Hoa

1. Introduction. In the multiplier methods for constrained optimization problems:

$$(1)\qquad \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) = 0 \quad i = 1, 2, \ldots, p \\ & g_i(x) \leqslant 0 \quad i = p+1, \ldots, m \end{cases}$$

one has to solve an infinite number of unconstrained minimizations of the augmented Lagrangian functional: $\min_x L(x, u^k, C)$ with $u^k \to u^*$ where u^* is the optimal Lagrange multiplier. In order to transfer this infinite sequence of unconstrained minimizations into a single problem Fletcher [4], [5] proposed to replace the sequence $\{u^k\}$ by a function $u(x)$. His idea is, generally speaking, to choose $u(x)$ in such a way that the augmented Lagrangian functional $L(x, u(x), C)$ has unconstrained local minimum x^* whenever x^* is a local solution of problem (1). If such a function $u(x)$ is already known, it remains only to apply the well-known algorithms to the unconstrained problem $\min_x L(x, u(x), C)$.

Let us denote by $\partial L(x, u(x), C)$ and $\partial^2 L(x, u(x), C)$, respectively, the first and the second order derivatives of the functional L considered as the functional only of variable x, while $\nabla_x L(x, u(x), C)$ and $\nabla^2_{xx} L(x, u(x), C)$ denote, respectively, the gradient vector and the Hessian matrix of the functional L with respect to the first variable x.

Usually the Lagrange multiplier function $u(x)$ is constructed in such a way that the sufficient conditions for unconstrained minimum of the functional $L(x, u(x), C)$ are satisfied at every strict local solution x^* of problem (1) i.e.: $\partial L(x^*, u(x^*), C) = 0$, $\partial^2 L(x^*, u(x^*), C)$ is positive definite. So it is necessary that $u(x)$ should be twice differentiable in a neighborhood of x^* ([8], [9], [10]). In [9] Martensson investigated this approach to solve the constrained optimization problem and suggested a class of multiplier functions having the properties mentioned above.

Owing to the fact that at every local solution x^* the gradient $\nabla_x L(x^*, u(x^*), C)$ vanishes for all $C \geqslant 0$, we can locate the local solution x^* of problem (1) by solving the nonlinear system of equations:

$$\nabla_x L(x, u(x), C) = 0 .$$

If $u(x)$ is continuously differentiable in a neighborhood of x^* and $\partial \nabla_x L(x^*, u(x^*), C)$ is invertible or, moreover, it is positive definite, then to solve this nonlinear system one can use the efficient methods, for instance, Newton's method or the secant quasi-Newton methods. In the case of the optimization problem with equality constraints this approach has been realized by Tapia [13] with a special Lagrange multiplier function.

In this paper we will discuss this approach to the general optimization problem with equality and inequality constraints. A new definition of the Lagrange multiplier function will be presented. The requirements on the Lagrange multiplier function will be weakened in comparison to that in [9]. It is sufficient that the Lagrange multiplier function $u(x)$ be continuously differentiable and not necessarily twice continuously differentiable. This permits us to construct a special class of Lagrange multiplier functions $w(x)$ which will satisfy the following condition:

(e) *For every* $C \geqslant C_0$ *(*C_0 *is given) a vector* x *solves the nonlinear system* $\nabla_x L(x, w(x), C) = 0$ *if and only if* x *is a critical point of problem* (1).

Hence, using the Lagrange multiplier function of this class we can locate all local solution of problem (1) by solving one nonlinear system of equations. The penalty constant C may be fixed. In order to ensure the high rate of convergence we can use, like in other approaches, Newton's method or the secant methods with C large enough, provided x^* is a nonsingular solution of the problem (1).

In this paper several results on the equation $\nabla_x L(x, u(x), C) = 0$ and on the derivative $\partial\nabla_x L(x, u(x), C)$ will be established for a general multiplier function in Section 2. In Section 3 the approach using the general multiplier function will be proposed and its drawbacks will be discussed. The improvements obtained by using a test function and an algorithm with increasing C will be considered in Section 4. In Section 5 and 6 we will deal with a construction of the Lagrange multiplier function for the optimization problem with equality constraints and with inequality constraints, respectively. The special class of Lagrange multiplier functions satisfying the condition (e) mentioned above will be proposed, too. In Section 7 we will present some algorithms and results on the convergence rate.

II. The Lagrange multiplier function. Consider the general constrained optimization problem with equality and inequality constraints (1). Assume that $f, g_1, \ldots, g_m$ are twice continuously differentiable functionals on R^n, with $m \leqslant n$.

We recall that the notation $g(x) = (g_1(x), \ldots, g_m(x))^T$, $u = (u_v, \ldots, u_m)^T$ $u(x) = (u_1(x), \ldots, u_m(x))^T$ is used. Let $\nabla f(x)$, $\nabla g_i(x)$, $\nabla u_i(x)$ denote, respectively, the gradient column vectors of the functionals f, g_i, u_i; $\nabla g(x) = (\nabla g_1(x), \ldots, \nabla g_m(x))$ and $\nabla^2 f(x), \nabla^2 g_i(x)$ denote their Hessians.

Given $x \in R^n$ we put

$$I(x) = \{1, \ldots, p\} \cup \{i = p+1, \ldots, m | g_i(x) \geqslant 0\}$$

Assume that the constraints of problem (1) satisfy the following *regularity condition*: for all $x \in R^n$ the gradients

(2) $$\{\nabla g_i(x) | i \in I(x) \quad \text{are linearly independent.}$$

For $x \in R^n$ and $u \in R^m$ the functional

$$F(x, u) = f(x) + u^T \cdot g(x)$$

is said to be *the Lagrangian functional* for problem (1). A vector $x \in R^n$ is said to be *a critical (or Kuhn-Tucker) point* of problem (1) if at x the following Kuhn-Tucker necessary condition for local solution holds: there exists a Lagrange multiplier $u \in R^m$ such that

$$\nabla_x F(x, u) = \nabla f(x) + \nabla g(x) \cdot u = 0, \tag{3}$$

$$g_i(x) = 0 \quad i = 1, \ldots, p, \tag{4}$$

$$g_i(x) \leqslant 0 \quad i = p+1, \ldots, m, \tag{5}$$

$$u_i \geqslant 0 \quad i = p+1, \ldots, m, \tag{6}$$

$$u \cdot g(x) = 0. \tag{7}$$

The relation (7) is said to be *the complementary condition.* It is said that at a critical point x the strict complementarity holds if $g_i(x) = 0$ implies $u_i > 0$, for all $i = p+1, \ldots, m$.

It has been shown [3] that when the regularity condition (2) holds, every local solution x of problem (1) is also a critical point. Moreover, at x the following *second order necessary condition* for local minimum is satisfied:

$$\nabla^2_{xx} F(x, u) = \nabla^2 f(x) + \sum_{i=1}^{m} u_i \cdot \nabla^2 g_i(x) \tag{8}$$

is positive semidefinite on the manifold $M = \{y \in R^n | y^T \cdot \nabla g_i(x) = 0 \text{ for all } i \in I(x)\}$. Conversely, if x is a critical point with the strict complementarity and if

$$\nabla^2_{xx} F(x, u) \text{ is positive definite on the manifold } M, \tag{9}$$

then x is a strict local minimum of problem (1). The condition (9) is said to be *the second order sufficient condition* for the local minimum.

In our approach the Lagrange multiplier function is defined in a natural way.

DEFINITION 2.1. A function u: $R^n \to R^m$ is said to be *the Lagrange multiplier function for problem* (1) if $u(x)$ is continuously differentiable in a neighborhood of x^* and $u(x^*) = u^*$ whenever x^* is a critical point of problem (1), where u^* is the associated Lagrange multiplier.

Given a nonnegative number C, we consider the *augmented Lagrangian functional* of Rockafellar [12] for the general constrained optimization problem (1):

$$L(x, u, C) = f(x) + \sum_{i=1}^{p} u_i \cdot g_i(x) + \frac{C}{2} \sum_{i=1}^{p} g_i^2(x) + \frac{1}{2C} \sum_{i=p+1}^{m} [(C \cdot g_i(x) + u_i)_+^2 - u_i^2], \tag{10}$$

where $a_+ = \max\{a; 0\}$, $a_- = \min\{a; 0\}$.

By differentiating (10) with respect to x we obtain

$$(11)\quad \nabla_x L(x,u,C) = \nabla f(x)+\sum_{i=1}^{p} u_i\cdot\nabla g_i(x)+C\sum_{i=1}^{p} g_i(x)\cdot\nabla g_i(x)+\sum_{i=p+1}^{m}(C\cdot g_i(x)+u_i)_+\cdot\nabla g_i(x)\,.$$

Substituting u by $u(x)$ in (11) we have

$$(12)\quad \nabla_x L(x,u(x),C) = \nabla f(x)+\sum_{i=1}^{p} u_i(x)\cdot\nabla g_i(x)+C\sum_{i=1}^{p} g_i(x)\cdot\nabla g_i(x) + \sum_{i=p+1}^{m}(C\cdot g_i(x)+u_i(x))_+\cdot\nabla g_i(x)\,.$$

The following result is obvious from the definition of the Lagrange multiplier function and from the relation (3)–(7).

PROPOSITION 2.1. *Assume that $u(x)$ is a Lagrange multiplier function and x^* is a critical point of problem* (1). *Then*

$$\nabla_x L(x^*,u(x^*),C) = 0 \quad \textit{for all } C\geqslant 0\,.$$

It is shown in [6] that if x^* is a nonsingular local minimum of problem (1) (i.e. at x^* the second order sufficient condition for local minimum holds) and, moreover, at x^* the strict complementary condition holds, then there exists $\hat{C}\geqslant 0$ such that for all $C\geqslant\hat{C}$ the Hessian $\nabla^2_{xx}L(x^*,u^*,C)$ is positive definite. In the same way we can prove the analogous result when the Lagrange multiplier function $u(x)$ is taken instead of u.

PROPOSITION 2.2 *Assume that $u(x)$ is a Lagrange multiplier function for problem* (1). *Let x^* be a nonsingular local minimum of problem* (1) *and at x^* the strict complementarity holds. Then there exists $\hat{C}\geqslant 0$ such that for all $C\geqslant\hat{C}$*

$$\partial\nabla_x L(x^*,u(x^*),C) \quad \textit{is positive definite.}$$

Proof. Given $x\in R^n$ we denote

$$Z(x,C) = \{i=p+1,\ldots,m \mid Cg_i(x) = -u_i(x)\}$$
$$J(x,C) = \{1,2,\ldots,p\}\cup\{i=p+1,\ldots,m \mid Cg_i(x)>-u_i(x)\}\,.$$

Differentiating (12) with respect to x we get

$$(13)\quad \partial\nabla_x L(x,u(x),C) = \nabla^2 f(x)+\sum_{i=1}^{p} u_i(x)\cdot\nabla^2 g_i(x)+C\sum_{i=1}^{p} g_i(x)\cdot\nabla^2 g_i(x) + \sum_{i=p+1}^{m}(Cg_i(x)+u_i(x))_+\cdot\nabla^2 g_i(x)+C\sum_{i\in J(x,C)}\nabla g_i(x)\cdot\nabla g_i(x)^T + \sum_{i\in J(x,C)}\nabla g_i(x)\cdot\nabla u_i(x)^T\,.$$

We observe that whenever $Z(x,C)\neq\emptyset$ the two last terms of the right-hand side of (13) have a jump at x. It causes the discontinuity of $\partial\nabla_x L(x,u(x),C)$ at x. Whereas for x satisfying $Z(x,C)=\emptyset$, $\partial\nabla_x L(x,u(x),C)$ is well-defined and is continuous because

f, g_i are twice continuously differentiable by assumption. Assuming that the strict complementary condition is satisfied at x^*, we obtain

$$Z(x^*, C) = \{i = p+1, \ldots, m \mid Cg_i(x^*) = u_i(x^*) = 0\} = \emptyset$$

for all $C>0$ by (5) and (6). So $\partial\nabla_x L(x^*, u(x^*), C)$ is well-defined and $\partial\nabla_x L(x, u(x), C)$ is continuous in a neighborhood of x^*.

To prove that $\partial\nabla_x L(x^*, u(x^*), C)$ is positive definite we show that

$$y^T\cdot\partial\nabla_x L(x^*, u(x^*), C)\cdot y>0$$

for all $y\in R^n$, $y\neq 0$.

Given $y\in R^n$, $y\neq 0$ we decompose y into a sum of two vectors $y = y^1+y^2$, where y^1 satisfies the condition $(y^1)^T\cdot\nabla g_i(x^*) = 0$ for all $i\in I(x^*)$ and y^2 is a linear combination of the gradients $\nabla g_i(x^*)$, $i\in I(x^*)$. By the regularity of the constraints we may write

$$\begin{aligned}(14)\qquad y &= y^1+\nabla g_I(x^*)[\nabla g_I(x^*)^T\cdot\nabla g_I(x^*)]^{-1}\cdot z\\ &= y^1[\nabla g_I(x^*)^+]^T\cdot z,\end{aligned}$$

where $\nabla g_I(x^*)$ is the matrix whose columns are the gradients of the active constraints at x^*, and $A^+ = (A^T\cdot A)^{-1}\cdot A^T$.

From (13), (14) we have:

$$\begin{aligned}(15)\quad y^T\cdot\partial\nabla_x L(x^*, u(x^*), C)\cdot y &= (y^1)^T\cdot\nabla^2_{xx}F(x^*, u^*)\cdot y^1+\\ &+2(y^1)^T\cdot\nabla^2_{xx}F(x^*, u^*)\cdot[\nabla g_I(x^*)^+]^T\cdot z+z^T\cdot[\nabla g_I(x^*)^+]\cdot\nabla^2_{xx}F(x^*, u^*)\cdot[\nabla g_I(x^*)^+]^T\cdot z\\ &+C\cdot z^T\cdot z+z^T\cdot\nabla u_I(x^*)^T\cdot y^1+z^T\cdot\nabla u_I(x^*)^T\cdot[\nabla g_I(x^*)^+]^T\cdot z.\end{aligned}$$

By the second order sufficient condition (9) there exists a positive number a such that $y^T\cdot\nabla^2_{xx}F(x^*, u^*)\cdot y\geqslant a\|y\|^2$ for all y belonging to the manifold

$$M = \{y\in R^n|y^T\cdot\nabla g_i(x^*) = 0,\ i\in I(x^*)\}.$$

Setting $b = \|\nabla^2_{xx}F(x^*, u^*)\cdot[\nabla g_I(x^*)^+]^T\|$, $c = \|\nabla u_I(x^*)\|$, $d = \|[\nabla g_I(x^*)^+]\cdot\nabla^2_{xx}F(x^*, u^*)\cdot[\nabla g_I(x^*)^+]^T\|$ and $e = \|[\nabla g_I(x^*)^+]\cdot\nabla u_I(x^*)\|$ we obtain

$$(16)\qquad y^T\cdot\partial\nabla_x L(x^*, u(x^*), C)\cdot y\geqslant a\|y^1\|^2-(2b+c)\|y^1\|\|z\|+(C-d-e)\|z\|^2.$$

Hence, due to the fact that $y^1\neq 0$ or $z\neq 0$, we get the existence of $\hat{C}$ guaranting the positive definiteness of $\partial\nabla_x L(x^*, u(x^*), C)$ for all $C\geqslant\hat{C}$.

The proposition is false when at the local minimum x^* the second order sufficient condition does not hold. However, in this case we obtain the following

PROPOSITION 2.3. *Assume that $u(x)$ is a Lagrange multiplier function for problem* (1). *Let x^* be a local minimum of problem* (1). *Then there exists $\hat{C}\geqslant 0$ such that for all $C\geqslant\hat{C}$*

$$\partial\nabla_x L(x^*, u(x^*), C)$$

is positive semidefinite.

Proof. By the second order necessary condition for a local minimum (8) the Hessian $\nabla^2_{xx}F(x^*, u^*)$ is positive semidefinite on the manifold M. The existence of $\hat{C}$ follows from (16) with $a = 0$.

We can observe that in this case $\partial\nabla_x L(x^*, u(x^*), C)$ is never positive definite, because for $y \in R^n$ such that $y^T \cdot \nabla g_i(x^*) = 0$, $i \in I(x^*)$ and $y^T, \nabla^2_{xx} F(x^*, u^*) \cdot y = 0$ we get

$$y^T \cdot \partial\nabla_x L(x^*, u(x^*), C) \cdot y = y^T \cdot \nabla^2_{xx} F(x^*, u^*) \cdot y = 0 .$$

III. The equation approach to the constrained optimization. In the sequel we shall consider the general constrained optimization problem (1) with the twice continuously differentiable functionals f, g_i which fulfil the regularity condition (2).

Let $u(x)$ be a Lagrange multiplier function for problem (1). Let x^* be a nonsingular local minimum of problem (1) and at x^* the strict complementarity condition is satisfied. So, by Proposition 2.1, x^* is a solution of the nonlinear system of equations

$$\nabla_x L(x, u(x), C) = 0 . \tag{17}$$

Moreover, by Proposition 2.2 there exists $\hat{C} \geq 0$ such that if $C \geq \hat{C}$ then $\partial\nabla_x L(x^*, u(x^*), C)$ is positive definite and $\partial\nabla_x L(x, u(x), C)$ is well-defined in a neighborhood of x^*.

Hence, to locate x^* we can apply the efficient methods for the nonlinear system (17), for instance, Newton's method or the secant quasi-Newton methods. These methods are locally well-defined, and if the generated sequence $\{x^k\}$ converges to x^*, then the convergence rate is Q-qadratical and Q-superlinear, respectively.

The following theorem guarantees the local convergence of the sequence $\{x^k\}$ to x^* when arbitrary convergent methods for nonlinear system (17) are used (assuming that C is sufficiently large).

THEOREM 3.1. *Assume that $u(x)$ is a Lagrange multiplier function for problem* (1). *Let x^* be a nonsingular local solution of problem* (1) *with the strict complementarity. Let S be compact set in which x^* is the unique critical point of problem* (1). *Then there exists a positive number C_S such that for all $C \geq C_S$, x^* is the unique solution in S of the nonlinear system* (17).

Proof. Suppose the theorem is false. Then we can assume, without loss of generality, that there exist sequences $\{C_k\}$, $\{x^k\} \subset S$ such that: $x^k \neq x^*$, $\nabla_x L(x^k, u(x^k), C_k) = 0$, $k = 1, 2, \ldots$, and $C_k \to \infty$, $x^k \to \bar{x} \in S$ (by the compactness of S) as $k \to \infty$. We remember that

$$\nabla_x L(x^k, u(x^k), C_k) = \nabla f(x^k) + \sum_{i=1}^{p} u_i(x^k) \cdot \nabla g_i(x^k) + C_k \sum_{i=1}^{p} g_i(x^k) \cdot \nabla g_i(x^k) + \sum_{i=p+1}^{m} (C_k g_i(x^k) + u_i(x^k))_+ \cdot \nabla g_i(x^k)$$

$$0 = \lim_{k \to \infty} \left\{ \nabla f(x^k) + \sum_{i=1}^{p} u_i(x^k) \cdot \nabla g_i(x^k) + C_k \left[\sum_{i=1}^{p} g_i(x^k) \cdot \nabla g_i(x^k) + \sum_{i=p+1}^{m} \left(g_i(x^k) + \frac{u_i(x^k)}{C_k} \right)_+ \cdot \nabla g_i(x^k) \right] \right\} . \tag{18}$$

Thus from the assumption $C_k \to \infty$ it follows that

$$\left[\sum_{i=1}^{p} g_i(x^k)\cdot\nabla g_i(x^k)+\sum_{i=p+1}^{m}\left(g_i(x^k)+\frac{u_i(x^k)}{C_k}\right)_+\cdot\nabla g_i(x^k)\right]\to 0$$

as $k \to \infty$, and by the continuity we obtain

$$\sum_{i=1}^{p} g_i(\bar{x})\cdot\nabla g_i(\bar{x})+\sum_{i=p+1}^{m} g_i(\bar{x})_+\cdot\nabla g_i(\bar{x}) = 0\,.$$

Owing to the regularity of the constraints we get

$$g_i(\bar{x}) = 0 \qquad i = 1, 2, \ldots, p\,,$$
$$g_i(\bar{x}) \leqslant 0 \qquad i = p+1, \ldots, m$$

i.e. $\bar{x}$ is a feasible point for problem (1).

Setting

(19) $$\bar{u}_i = \lim_{k\to\infty} \left(C_k g_i(x^k)+u_i(x^k)\right) \qquad i = 1, 2, \ldots, p$$
$$\bar{u}_i = \lim_{k\to\infty} \left(C_k g_i(x^k)+u_i(x^k)\right)_+ \qquad i = p+1, \ldots, m$$

we have $\bar{u}_i \geqslant 0$ for $i = p+1, \ldots, m$. The existence of the limits in (19) follows from the regularity of the constraints and from (18). It is obvious from (19) that $\bar{u}_i = 0$ for all i satisfying $g_i(x)<0$.

From (18), letting $k \to \infty$, we obtain

$$\nabla f(\bar{x})+\sum_{i=1}^{m} \bar{u}_i\cdot\nabla g_i(\bar{x}) = 0$$

This shows that $\bar{x}$ is a critical point of problem (1). By the assumption of the theorem, $\bar{x} = x^*$ by force.

Setting $d_k = \dfrac{x^k-x^*}{\|x^k-x^*\|}$ we can assume, without loss of generality that $d_k \to d$.

Because of the continuity of $g_i(x)$, $u_i(x)$ there exists a neighborhood of x^* in which $g_i(x)<\delta<0$ for all $i \notin I(x^*)$ (i.e. for i such that $g_i(x^*)<0$). So there exists C_0 such that in this neighborhood if $C\geqslant C_0$, then $Cg_i(x)+u_i(x)\leqslant 0$ for all $i \notin I(x^*)$. This proves that $J(x^k, C_k) = \{1, \ldots, p\} \cup \{i = p+1, \ldots, m \mid C_k g_i(x^k)+u_i(x^k)>0\} \subset I(x^*)$ for all $k\geqslant k_0$, where k_0 is large enough.

Let us denote

(20) $$N(C_k) = \{x \mid J(x, C_k) = I(x^*)\}\,.$$

By the definition of $J(x, C)$ and the strict complementarity condition, $N(C_k)$ is a neighborhood of x^* and in this neighborhood the derivative $\partial\nabla_x L(x, u(x), C_k)$ is continuous because $Z(x, C_k) = \emptyset$.

Now we show that if k_0 is sufficiently large, then $x^k \in N(C_k)$ for all $k\geqslant k_0$.

Suppose that the converse holds. Then there exists an infinite subsequence of $\{x^k\}$, say $\{\tilde{x}^k\}$, which is indexed by $K \subset \{1, 2, \ldots\}$ such that $J\{\tilde{x}^k, C_k) \neq I(x^*)$ for all $k \in K$. Because of that the number of subsets of $I(x^*)$ is finite, we can assume, without loss of generality, that $J(\tilde{x}^k, C_k) = J$, $J \subset I(x^*)$, $J \neq I(x^*)$ for all $k \in K$. Now from (19) we get

$$u_i = \lim_{k\to\infty} (C_k g_i(x^k) + u_i(x^k))_+ = \lim_{\substack{k\to\infty \\ k\in K}} (C_k g_i(\tilde{x}^k) + u_i(\tilde{x}^k))_+ = 0$$

for all $i \in I(x^*)$, $i \notin J$. So it follows from the regularity condition (2) that $u_i^* = u_i = 0$, $g_i(x^*) = 0$ for all $i \in I(x^*)$, $i \notin J$. This contradicts the strict complementarity.

Now we have

$$\begin{aligned} 0 &= \nabla_x L(x^k, u(x^k), C_k) - \nabla_x L(x^*, u(x^*), C_k) \\ &= \nabla f(x^k) + \sum_{i\in I(x^*)} u_i(x^k)\cdot\nabla g_i(x^k) - \nabla f(x^*) - \sum_{i\in I(x^*)} u_i^*\cdot\nabla g_i(x^*) + \\ &\qquad + C_k \sum_{i\in I(x^*)} g_i(x^k)\cdot\nabla g_i(x^k)\,. \end{aligned}$$

Setting $F_I(x, u) = f(x) + \sum_{i\in I(x^*)} u_i\cdot g_i(x)$ we can rewrite the above equality in the form:

$$\begin{aligned} 0 &= \nabla_x F_I(x^k, u(x^k)) - \nabla_x F_I(x^*, u^*) + C_k \sum_{i\in I(x^*)} g_i(x^k)\cdot\nabla g_i(x^k) \\ &= \nabla_x F_I(x^k, u^*) - \nabla_x F_I(x^*, u^*) + \sum_{i\in I(x^*)} (u_i(x^k) - u_i^*)\cdot\nabla g_i(x^k) + \\ &\qquad + C_k \sum_{i\in I(x^*)} g_i(x^k)\cdot\nabla g_i(x^k)\,. \end{aligned}$$

Dividing both sides of the above equality by $\|x^k - x^*\|$ and letting $k \to \infty$ we obtain:

$$\begin{aligned} 0 &= \nabla^2_{xx} F(x^*, u^*)\cdot d + \sum_{i\in I(x^*)} \nabla u_i(x^*)^T\cdot d\cdot\nabla g_i(x^*) + \\ &\qquad + \lim_{k\to\infty} C_k \sum_{i\in I(x^*)} \frac{g_i(x^k) - g_i(x^*)}{\|x^k - x^*\|}\cdot\nabla g_i(x^k)\,. \end{aligned} \tag{21}$$

Hence, in view of that $\nabla g_i(x^k) \to \nabla g_i(x^*)$ and $\{\nabla g_i(x^*) | i \in I(x^*\}$ are linearly independent, we can conclude that the limits

$$\lim_{k\to\infty} C_k \frac{g_i(x^k) - g_i(x^*)}{\|x^k - x^*\|} \tag{22}$$

exist and are finite for all $i \in I(x^*)$.

It follows from (21), owing to the fact that $C_k \to \infty$, that

$$\lim_{k\to\infty} \sum_{i\in I(x^*)} \frac{g_i(x^k) - g_i(x^*)}{\|x^k - x^*\|}\cdot\nabla g_i(x^k) = \sum_{i\in I(x^*)} \nabla g_i(x^*)^T\cdot d\cdot\nabla g_i(x^*) = 0\,.$$

Therefore, we get

$$\nabla g_i(x^*)^T\cdot d = 0 \quad \text{for all } i \in I(x^*)\,. \tag{23}$$

Multiplying by d^T both sides of (21) we obtain:

$$(24)\qquad 0 = d^T\cdot\nabla^2_{xx}F(x^*, u^*)\cdot d + \lim C_k \sum_{i\in I(x^*)} \frac{g_i(x^k)-g_i(x^*)}{\|k^k - x^*\|}\cdot d^T\cdot\nabla g_i(x^k)\,.$$

By using (22) and (23) we can conclude that the last term of (24) is equal to zero. Therefore, we have $d^T\cdot\nabla^2_{xx}F(x^*, u^*)\cdot d = 0$. This, in combination with (23), contradicts the assumption that x^* is a nonsingular solution of problem (1). Thus the proof of theorem is completed.

It is necessary to observe that, by Proposition 2.2, the derivative $\partial\nabla_x L(x^*, u(x^*), C)$ is positive definite at nonsingular local solution x^* for all $C \geqslant \hat{C}$ and because of this there exists a neighborhood of x^* in which the equation $\nabla^* L(x, u(x), C) = 0$ has the only solution x^*. Theorem 3.1 gives the same conclusion but for any compact set S containing the unique critical point x^*. Moreover, the theorem permits also to conclude that if problem (1) is nonsingular (i.e. every critical point of the problem is also a nonsingular solution) then, for any compact set S there exists C_S such that for all $C \geqslant C_S$ the solution x^* of (17) is also a local solution of problem (1).

It is obvious, from the proof of the theorem, that if problem (1) has a singular local solution or the penalty constant C is not large enough, then, generally, a solution x of the nonlinear system (17) may not be a critical point. Moreover, x may not be a feasible point of problem (1). However, we can omit this difficulty in two following ways: by constructing an algorithm model with increasing of C when the increasing is necessary or by choosing a special Lagrange multiplier function $w(x)$ satisfying the condition (e) mentioned in the section 1.

IV. The test function and algorithm model. To improve the unconstrained minimization approach of Fletcher to constrained minimization problem, Glad and Polak [8] have constructed a test function $t(x, C)$ and an algorithm model with increasing C such that the generated sequence $\{x^k\}$ converges to a critical point of problem (1). Observing that, in fact, the algorithm of Glad and Polak intends to locate the stationary point of the augmented Lagrangian functional $L(x, u(x), C)$ i.e. the solution of $\partial L(x, u(x), C) = 0$ we can adapt their test function to our approach, where we solve the nonlinear system (17). All results are established in the same way as in [8].

Assume that $u(x)$ is a multiplier function. Consider the following *test function*:

$$(25)\qquad t(x, C) = -\|\nabla_x L(x, u(x), C)\|^2 + \frac{1}{C}\left\{\sum_{i=1}^{p} g_i(x)^2 + \sum_{i=p+1}^{m}\left[\left(g_i(x)+\frac{u_i(x)}{C}\right)_+ - \frac{u_i(x)}{C}\right]^2\right\}.$$

PROPOSITION 4.1. *If $\bar{x}$ is a critical point of problem* (1), *then*

$$\nabla_x L(\bar{x}, u(\bar{x}), C) = 0\,, \quad t(\bar{x}, C) \leqslant 0 \quad \textit{for all } C > 0\,.$$

Conversely, if for a positive number C $\bar{x}$ satisfies

$$\nabla_x L(\bar{x}, u(\bar{x}), C) = 0, \quad t(\bar{x}, C) \leqslant 0 \tag{26}$$

then $\bar{x}$ is a critical point of problem (1).

Proof. The first conclusion of the proposition is obvious by (4)–(7) and by Proposition 2.1. To prove the second conclusion of the proposition we observe that from (25) and (26) it follows

$$\sum_{i=1}^{p} g_i(\bar{x})^2 + \sum_{i=p+1}^{m} \left[\left(g_i(\bar{x}) + \frac{u_i(\bar{x})}{C}\right)_+ - \frac{u_i(\bar{x})}{C}\right]^2 = 0$$

so, $g_i(\bar{x}) = 0$, $i = 1, 2, \ldots, p$ and $\left(g_i(\bar{x}) + \frac{u_i(\bar{x})}{C}\right)_+ = \frac{u_i(\bar{x})}{C}$, $i = p+1, \ldots, m$. But the last equalities show that

$$\begin{aligned} u_i(\bar{x}) &\geqslant 0 \quad i = p+1, \ldots, m, \\ g_i(\bar{x}) &\leqslant 0 \quad i = p+1, \ldots, m, \\ \text{and} \quad u_i(\bar{x}) \cdot g_i(\bar{x}) &= 0 \quad i = p+1, \ldots, m. \end{aligned}$$

Therefore, $\bar{x}$ is a feasible point and setting $\bar{u} = u(\bar{x})$, we see that $\bar{x}$ is a critical point of problem (1) with the associated Lagrange multiplier $\bar{x}$.

Hence, given an iterative procedure $A(x, C)$ for the nonlinear system (17) we can construct the algorithm model with the test function (25) in the same way as in [8].

ALGORITHM MODEL

Step 1: Determine $x^0 \in R^n$; $C_0 > 0$; set $k := 0$; $l := 0$

Step 2: if $t(x^k, C_k) > 0$ is true, go to step 3, if not go to step 4

Step 3: set $\tilde{x}^l := x^k$; $C_{l+1} := \phi(C_1)$; $l := l+1$ go to step 2

Step 4: if $\nabla_x L(x^k, u(x^k), C_k) = 0$ stop, if not compute $x^{k+1} = A(x^k, C_l)$ set $k := k+1$; go to step 1,

where the functional ϕ: $R^1_+ \to R^1_+$ satisfies $\phi(C) > C$ for all $C \in R^1_+$ and $\phi(C) \to \infty$ as $C \to \infty$.

THEOREM OF CONVERGENCE. ([8], Theorem 1). *Let the sequences $\{x^k\}$, $\{\tilde{x}^l\}$, $\{C_l\}$ be generated by the algorithm model. Assume that*

(i) *For all l, $A(x, C_l)$ is a convergent iterative operator for the nonlinear system $\nabla_x L(x, u(x), C) = 0$;*

(ii) *For all $C > 0$ the test function $x \to t(x, C)$ is continuous;*

(iii) *If $\nabla_x L(x, u(x), C) = 0$ and $t(x, C) \leqslant 0$, then x is a critical point of problem* (1);

(iv) *Given $\hat{x} \in R^n$, there exists $\hat{C} \geqslant 0$, $\hat{\varepsilon} > 0$ such that $t(x, C) \leqslant 0$ for all $C \geqslant \hat{C}$ and $x \in \bar{B}(\hat{x}, \hat{\varepsilon}) = \{x \mid \|x - \hat{x}\| \leqslant \hat{\varepsilon}\}$.*

Then,

(1) *If the sequence* $\{x^k\}$ *is finite (consequently, the sequence* $\{\tilde{x}^l\}$ *is also finite) then, the last element is a critical point of problem* (1).

(2) *If the sequence* $\{\tilde{x}^l\}$ *is finite and the sequence* $\{x^k\}$ *is infinite, then every accumulation point of* $\{x^k\}$ *is critical.*

(3) *If the sequence* $\{\tilde{x}^l\}$ *is infinite then this sequence has no accumulation point.*

From (25) and Proposition 4.1 it follows that the test function $t(x, C)$ defined by (25) satisfies the conditions (ii), (iii), of the theorem.

Now we shall prove that $t(x, C)$ satisfies also condition (iv), provided that at every critical point x^* of problem (1) the second order sufficient condition for local minimum and the strict complementarity hold.

LEMMA 4.1. *Let* $t(x, C)$ *be the test function defined by* (25). *Assume that at every critical point* x^* *of problem* (1) *the second order sufficient condition* (9) *and the strict complementarity hold. Then* $t(x, C)$ *satisfies the condition* (iv) *of the convergence theorem.*

Proof. First we assume that $\hat{x}$ is not a critical point. Then there exists $\hat{\varepsilon}>0$ such that the closed ball $\bar{B}(\hat{x}, \hat{\varepsilon})$ does not contain any critical point of problem (1). Now we prove that there exist $C_0 \geqslant 0$, $\delta>0$ such that

(27) $$\|\nabla_x L(x, u(x), C)\| \geqslant \delta \quad \text{for all } C \geqslant C_0,\ x \in \bar{B}(\hat{x}, \hat{\varepsilon}).$$

So, by (25) and the continuity of $g_i(x)$, $u_i(x)$ we shall have $t(x, C) \leqslant 0$ for all $x \in \bar{B}(\hat{x}, \hat{\varepsilon})$ and $C \geqslant \hat{C}$, where $\hat{C}$ is large enough.

To prove (27) we assume that the contrary holds. Then there exist sequences $\{C_k\}$, $\{x^k\} \subset \bar{B}(\hat{x}, \hat{\varepsilon})$ such that $\lim \nabla_x L(x^k, u(u^k), C_k) = 0$ and $C_k \to \infty$, $x^k \to \bar{x} \in \bar{B}(\hat{x}, \hat{\varepsilon})$ (by the compactness of $\bar{B}$). Using the same reasoning as in the first part of the proof of Theorem 3.1, we can conclude that $\bar{x}$ is a critical point of problem (1). This contradicts the construction $\bar{B}(\hat{x}, \hat{\varepsilon})$ and (27) is established.

Let $\hat{x}$ be a critical point of problem (1). We show that there exist $C_0 \geqslant 0$, $\delta>0$, $\hat{\varepsilon}>0$ such that

(28) $$\|\nabla_x L(x, u(x), C)\| \geqslant \delta \|x-\hat{x}\| \quad \text{for all } x \in \bar{B}(\hat{x}, \hat{\varepsilon}) \text{ and } C \geqslant C_0.$$

Assume that $\hat{\varepsilon}$ is sufficiently small such that in the closed ball $\bar{B}(\hat{x}, \hat{\varepsilon})$ $\hat{x}$ is the unique critical point. The existence of such an $\hat{\varepsilon}$ is guaranteed by the second order sufficient condition at $\hat{x}$. If (28) is false, then there exist the sequences $\{C_k\}$, $\{x^k\} \subset \bar{B}(\hat{x}, \hat{\varepsilon})$ such that

(29) $$\lim_{k\to\infty} \frac{\nabla_x L(x^k, u(x^k), C_k)}{\|x^k-\hat{x}\|} = 0$$

and $C_k \to \infty$, $x^k \to \bar{x} \in \bar{B}(\hat{x}, \hat{\varepsilon})$ as $k \to \infty$. Therefore, in view of $\lim\limits_{k\to\infty} \nabla_x L(x^k, u(x^k), C_k) = 0$, we can show that $\bar{x}$ is a critical point. By construction of $\bar{B}(\hat{x}, \hat{\varepsilon})$ $\bar{x}$ must coincide with $\hat{x}$. Setting $d_k = \dfrac{x^k-\hat{x}}{\|x^k-\hat{x}\|}$ we can assume, without loss of generality, that $d_k \to d$. From (29) it follows that

$$\lim_{k\to\infty} \frac{\nabla_x L(x^k, u(x^k), C_k) - \nabla_x L(\hat{x}, u(\hat{x}), C_k)}{||x^k - \hat{x}||} = 0 .$$

Using the same reasoning as in the proof of the theorem 3.1 we obtain the contradiction with the second order sufficient condition at $\hat{x}$.

Let us notice that in the close ball $\bar{B}(\hat{x}, \hat{\varepsilon})$ the functionals $g_i(x)$, $u_i(x)$ are Lipschitz continuous, so $\left[\left(g_i(x) + \frac{u_i(x)}{C}\right)_+ - \frac{u_i(x)}{C}\right]$ is Lipschitz continuous uniformly with respect to $C \in [C_0, \infty)$. At the critical point $\hat{x}$ we have $g_i(\hat{x}) = 0$ for $i = 1, \ldots, p$ and $\left[\left(g_i(\hat{x}) + \frac{u_i(\hat{x})}{C}\right)_+ - \frac{u_i(\hat{x})}{C}\right] = 0$ for $i = p+1, \ldots, m$. So, we get

$$\text{(30)} \qquad \begin{cases} |g_i(x)| \leqslant K||x - \hat{x}|| & i = 1, \ldots, p \\ \left|\left(g_i(x) + \frac{u_i(x)}{C}\right)_+ - \frac{u_i(x)}{C}\right| \leqslant K||x - \hat{x}||, & i = p+1, \ldots, m \end{cases}$$

for all $C \geqslant C_0$ and $x \in \bar{B}(\hat{x}, \hat{\varepsilon})$.

By (25), (28) and (30) we have $t(x, C) \leqslant \left(-\delta^2 + \frac{m \cdot K^2}{C}\right) ||x - \hat{x}||^2$.

Setting $\hat{C} = \max\left\{C_0, \frac{m \cdot K^2}{\delta^2}\right\}$ we obtain $t(x, C) \leqslant 0$ for all $C \geqslant \hat{C}$ and $x \in \bar{B}(\hat{x}, \hat{\varepsilon})$.

The condition (iv) guarantees that if the sequence $\{x^k\}$ is bounded the increasing of C must stop after a finite number of increasings, because the test $t(x, C) \leqslant 0$ will be automatically satisfied.

V. The Lagrange multiplier function for constrained optimization with equality constraints. For the sake of simplicity, we deal with the equality constrained optimization problem in this section, while the inequality constrained problem will be considered in the next section. The Lagrange multiplier function can be generalized in a natural way, when the constraint set contains a mixture of equality and inequality constraints.

Consider problem (1) with $p = m$

$$\text{(31)} \qquad \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) = 0 \quad i = 1, \ldots, m \end{cases}$$

Assume that the following regularity condition holds:

$$\text{(32)} \qquad \{\nabla g_i(x) | i \in I(x)\} \quad \text{are linearly independent,}$$

where $I(x) = \{i | g_i(x) = 0\}$.

The Kuhn–Tucker necessary condition for a local minimum of problem (31) is the following: there is $u^* \in R^m$ such that

(33) $$\nabla_x F(x^*, u^*) = \nabla f(x^*) + \nabla g(x^*) \cdot u^* = 0\,,$$

(34) $$g(x^*) = 0\,.$$

Let us put $G(x) = \operatorname{diag}(g_1(x), \ldots, g_m(x))$.

PROPOSITION 5.1. *For all $\alpha \geqslant 0$, $b \geqslant 0$ the function*

(35) $$u(x) = [\nabla g(x)^T \cdot \nabla g(x) + a \cdot G(x)^2]^{-1}(b \cdot g(x) - \nabla g(x)^T \cdot \nabla f(x))$$

is the Lagrange multiplier function for problem (31).

Proof. The matrix $[\nabla g(x)^T \cdot \nabla g(x) + a \cdot G(x)^2]$ is positive semidefinite for all $x \in R^n$ and $a \geqslant 0$. Using the regularity condition (32) we can prove that this matrix, with $a > 0$, is positive definite for all $x \in R^n$. In the case where $a = 0$, the matrix $\nabla g(x)^T \cdot \nabla g(x)$ is positive definite at every feasible point of problem (31). Thus $u(x)$ is well-defined in a neighborhood of x^* whenever x^* is a critical point. It is continuously differentiable because f, g_i are twice continuously differentiable.

To prove that $u(x^*) = u^*$ whenever x^* is a critical point, we rewrite (35) in the form:

$$[\nabla g(x)^T \cdot \nabla g(x) + a \cdot G(x)^2] u(x) = b \cdot g(x) - \nabla g(x)^T \cdot \nabla f(x)\,.$$

Setting $x = x^*$, we get

$$[\nabla g(x^*)^T \cdot \nabla g(x^*)] u(x^*) = -\nabla g(x^*)^T \cdot \nabla f(x^*)\,.$$

Multiplying both sides of (33) by $\nabla g(x^*)^T$ we get

$$[\nabla g(x^*)^T \cdot \nabla g(x^*)] u^* = -\nabla g(x^*) \cdot^T \nabla f(x^*)\,.$$

So the condition $u(x^*) = u^*$ easily follows from the positive definiteness of the matrix $[\nabla g(x^*) \cdot^T \nabla g(x^*)]$.

Notice that, setting $a = 0$, $b = 0$ in (35) we get the projection multiplier update ([9], [13]). The Lagrange multiplier function (35) with $a = 1$, $b = 0$ was suggested by Martensson [9].

PROPOSITION 5.2. *Let $w(x)$ be the following Lagrange multiplier function*

(36) $$w(x) = [\nabla g(x)^T \cdot \nabla g(x)]^{-1}(b \cdot g(x) - \nabla g(x)^T \cdot \nabla f(x)) \quad \textit{with } b > 0\,.$$

Then for every fixed $C \geqslant 0$, the vector x^* solves the nonlinear system of equations

(37) $$\nabla_x L(x, w(x), C) = \nabla f(x) + \nabla g(x) \cdot w(x) + C \cdot \nabla g(x) \cdot g(x) = 0$$

if and only if x^* is a critical point of problem (31).

Proof. By Proposition 2.1 we need only prove that a solution x^* of (37) is also a critical point of problem (31). Let us set $x = x^*$ in (37) and multiply both sides by

$\nabla g(x^*)^T$. We get

$$\nabla g(x^*)^T \cdot \nabla f(x^*) + \nabla g(x^*)^T \cdot \nabla g(x^*) \cdot w(x^*) + C \cdot \nabla g(x^*)^T \cdot \nabla g(x^*) \cdot g(x^*) = 0 .$$

From (36) we have

$$\nabla g(x^*)^T \cdot \nabla f(x^*) + \nabla g(x^*)^T \cdot \nabla g(x^*) \cdot w(x^*) - b \cdot g(x^*) = 0 .$$

Substracting the two last equalities side by side, we obtain

$$[C \cdot \nabla g(x^*)^T \cdot \nabla g(x^*) + b \cdot I] g(x^*) = 0 . \tag{38}$$

Hence, owing to the fact that the matrix $[C \cdot \nabla g(x^*)^T \cdot \nabla g(x^*) + b \cdot I]$ is positive definite for all $C \geqslant 0$, $b > 0$, we obtain $g(x^*) = 0$. This means that x^* is a feasible point of problem (31).

Setting $u^* = w(x^*)$, from (37) we have

$$\nabla f(x^*) + \nabla g(x^*) \cdot u^* = 0 .$$

This proves that x^* is a critical point of problem (31) with the associated Lagrange multiplier u^*.

Notice that setting $b = 1$ in (36) we get the Lagrange multiplier function investigated by Tapia in [13].

The following lemma will be used in the proof of the convergence of the quasi-Newton methods presented in section 7.

LEMMA 5.3. *Let $u(x)$ be the Lagrange multiplier function defined by* (35) *with $a \geqslant 0$, $b \geqslant 0$. Then the derivative $\partial \nabla_x L(x, u(x), C)$ is symmetric at every critical point of problem* (31).

Proof. From (35) and (36), by straightforward calculation, we obtain $\nabla u(x^*)^T = (\nabla g(x^*)^T \cdot \nabla g(x^*))^{-1} [\nabla g(x^*)^T (b \cdot I - \nabla^2_{xx} L(x^*, u(x^*), C))]$, where x^* is a critical point of problem (31). So

$$\begin{aligned} \partial \nabla_x L(x^*, u(x^*), C) &= \nabla^2_{xx} L(x^*, u(x^*), C) + \nabla g(x^*) \cdot \nabla(x^*)^T \\ &= \nabla^2_{xx} L(x^*, u(x^*), C) + \nabla g(x^*) (\nabla g(x^*)^T \cdot \nabla g(x^*))^{-1} \nabla g(x^*)^T [b \cdot I - \nabla^2_{xx} L(x^*, u(x^*), C)] \end{aligned} \tag{39}$$

and the conclusion of the lemma is obvious.

VI. The Lagrange multiplier function for problem with inequality constraints. Consider the inequality constrained optimization problem

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leqslant 0 \quad i = 1, \dots, m \end{cases} \tag{40}$$

The regularity condition (2) of the constraints becomes:

$$\{\nabla g_i(x) | i \in I(x)\} \text{ are linearly independent, where } I(x) = \{i | g_i(x) \geqslant 0\} . \tag{41}$$

The Kuhn–Tucker necessary condition for a local minimum is the following: there is $u^* \in R^m$ such that

(42) $$\nabla f(x^*) + \nabla g(x^*) \cdot u^* = 0 ,$$

(43) $$u^{*T} \cdot g(x^*) = 0 ,$$

(44) $$g(x^*) \leqslant 0 ,$$

(45) $$u^* \geqslant 0 .$$

Let us set $G(x) = \operatorname{diag}(g_1(x), \dots, g_m(x))$, $G(x)_- = \operatorname{diag}(g_1(x)_-, \dots, g_m(x)_-)$

PROPOSITION 6.1. *For all $\alpha > 0$, $b \geqslant 0$ the functions*

(46) $$u(x) = (\nabla g(x)^T \cdot \nabla g(x) + a \cdot G(x)^2)^{-1} (b \cdot g(x)_+^2 - \nabla g(x)^T \cdot \nabla f(x)) ,$$

(47) $$w(x) = (Vg(x)^T \cdot \nabla g(x) + a \cdot G(x)_-^2)^{-1} (b \cdot g(x)_+^2 - \nabla g(x) \cdot \nabla f(x))$$

are the Lagrange multiplier functions for problem (40).

Proof. The matrices $(\nabla g(x)^T \cdot \nabla g(x) + a \cdot G(x)^2)$, $(\nabla g(x)^T \cdot \nabla g(x) + a \cdot G(x)_-^2)$ with $a > 0$ are positive semidefinite. Using the regularity condition (41) we can easily prove that they are positive definite for all $x \in R^n$. Hence, $u(x)$ and $w(x)$ are well-defined for all $x \in R^n$. They are continuously differentiable because f, g_i are twice differentiable. The condition $u(x^*) = u^*$, $w(x^*) = u^*$ at every critical point x^* can be verified in the same way as in the proof of Proposition 5.1.

Notice that the Lagrange multiplier function $u(x)$ defined by (46) with $a = 1$, $b = 0$ was investigated by Glad and Polak [8].

The following proposition demonstrates, like Proposition 5.2, that the Lagrange multiplier function $w(x)$ satisfies the condition (e) mentioned in Section 1.

PROPOSITION 6.2. *Assume that the regularity condition* (41) *holds. Let $w(x)$ be the Lagrange multiplier function defined by* (47) *with $a > 0$, $b \geqslant 0$. Then for any fixed $C > 0$, the vector x^* solves the nonlinear system of equations*

(48) $$\nabla_x L(x, w(x), C) = \nabla f(x) + \nabla g(x) (Cg(x) + w(x))_+ = 0$$

if and only if x^* is a critical point of problem (40).

Proof. By Proposition 2.1 we need only show the following implication: if x^* solves (48) then x^* is a critical point of (40).

Setting $x = x^*$ in (48) and multiplying both sides of (48) by $\nabla g(x^*)^T$ we get

$$\nabla g(x^*)^T \cdot \nabla f(x^*) + \nabla g(x^*)^T \cdot \nabla g(x^*) (Cg(x^*) + w(x^*))_+ = 0 .$$

From (47) we have

$$\nabla g(x^*)^T \cdot \nabla f(x^*) + (\nabla g(x^*)^T \cdot \nabla g(x^*) + a \cdot G(x^*)_-^2) w(x^*) - b \cdot g(x^*)_+^2 = 0 .$$

Substracting the two last equalities side by side we obtain:

(49) $$(Vg(x^*)^T\cdot\nabla g(x^*)+a\cdot G(x^*)_-^2)w(x^*)$$
$$-\nabla g(x^*)^T\cdot\nabla g(x^*)\big(Cg(x^*)+w(x^*)\big)_+-bg(x^*)_+^2 = 0\,.$$

From (49), by straightforward calculation, we have

(50) $$[w(x^*)-\big(Cg(x^*)+w(x^*)\big)_+]^T(\nabla g(x^*)^T\cdot\nabla g(x^*)+a\cdot G(x^*)_-^2)[w(x^*)$$
$$-\big(Cg(x^*)+w(x^*)\big)_+]+a\cdot[w(x^*)-\big(Cg(x^*)+w(x^*)\big)_+]^T\cdot G(x^*)_-^2\big(Cg(x^*)+w(x^*)\big)_+$$
$$-b\cdot[w(x^*)-\big(Cg(x^*)+w(x^*)\big)_+]^T\cdot g(x^*)_+^2 = 0\,.$$

Since the matrix $(\nabla g(x^*)^T\cdot\nabla g(x^*)+a\cdot G(x^*)_-^2)$ is positive definite, the first term of (50) is nonnegative. We shall prove that the second and the third terms of (50) are also nonnegative. In consequence, all three terms will be equal to zero.

By definition, $g_i(x^*)_-^2 = 0$ for i such that $g_i(x^*)\geqslant 0$. If $g_i(x^*)<0$ and $w_i(x^*)\leqslant 0$ then $\big(Cg_i(x^*)+w_i(x^*)\big)_+ = 0$. If $g_i(x^*)<0$ and $w_i(x^*)>0$ then $g_i(x^*)_-^2>0$, $\big(Cg_i(x^*)+w_i(x^*)\big)_+\geqslant 0$, $w_i(x^*)-\big(Cg_i(x^*)+w_i(x^*)\big)_+\geqslant 0$. Therefore, the second term is nonnegative. In the same way we prove that the third term is nonnegative. We have $g_i(x^*)_+^2 = 0$ for i such that $g_i(x^*)\leqslant 0$. If $g_i(x^*)>0$ then $w_i(x^*)-\big(Cg_i(x^*)+w_i(x^*)\big)_+\leqslant 0$. Hence

$$-[w(x^*)-\big(Cg(x^*)+w(x^*)\big)_+]^Tbg(x^*)_+^2\geqslant 0\,.$$

Now we get from (50):

$$[w(x^*)-\big(Cg(x^*)+w(x^*)\big)_+]^T(\nabla g(x^*)^T\cdot\nabla g(x^*)+a\cdot G(x^*)_-^2)$$
$$\times[w(x^*)-\big(Cg(x^*)+w(x^*)\big)_+] = 0\,.$$

By the positive definiteness of the matrix $(\nabla g(x^*)^T\cdot\nabla g(x^*)+a\cdot G(x^*)_-^2)$ we obtain:

(51) $$w(x^*)-\big(Cg(x^*)+w(x^*)\big)_+ = 0\,.$$

So

$$w(x^*) = \big(Cg(x^*)+w(x^*)\big)_+\geqslant 0\,.$$

From (51) we get also $g(x^*)\leqslant 0$, because if there were $g_i(x^*)>0$ then there would be $\big(Cg_i(x^*)+w_i(x^*)\big)_+ = Cg_i(x^*)+w_i(x^*)>w_i(x^*)$, what contradicts (51).

In the end, it must be $w(x^*)^T\cdot g(x^*) = 0$, because if it were $g_i(x^*)<0$ and $w_i(x^*)>0$ it would be $w_i(x^*)>\big(Cg_i(x^*)+w_i(x^*)\big)_+$, what would contradict (51).

Therefore, x^* is a critical point of problem (40) with the associated Lagrange multiplier $u^* = w(x^*)$.

The following Lemma, like the lemma 5.3, can be proved by straightforward calculation.

LEMMA 6.3. *Let $u(x)$, $w(x)$ be the Lagrange multiplier functions defined by* (46), (47) *with $a>0$, $b\geqslant 0$. Then the derivatives $\partial\nabla_x L(x^*, u(x^*), C)$ and $\partial\nabla_x L(x^*, w(x^*), C)$ are symmetric at every critical point of problem* (40).

VII. The algorithm and conclusion. Assume the Lagrange multiplier function $w(x)$ satisfies the condition (e) of Section 1. Then one can locate any critical point of problem (1) by solving the nonlinear system (17) with fixed C.

$$\nabla_x L(x, w(x), C) = 0 .$$

Let $A(x, C)$ be a convergent iterative operator for the system (17). Then the iteration

$$x^{k+1} = A(x^k, C)$$

generates the sequence $\{x^k\}$ convergent to a critical point of (1).

In order to ensure the high rate of convergence in a neighborhood of x^* which is a nonsingular local solution of problem (1) we may use Newton's method or the secant methods with C large enough.

THEOREM 7.1. *Assume that x^* is a nonsingular local solution of problem* (1) *and at x^* the strict complementary condition holds. Assume that there exists a neighborhood of x^* and a positive number K such that for all x belonging to this neighborhood we have*

$$\|\partial\nabla_x L(x, w(x), C) - \partial\nabla_x L(x^*, w(x^*), C)\| \leqslant K\|x - x^*\| . \tag{52}$$

Then for all $C \geqslant \hat{C}$ ($\hat{C}$ is large enough) Newton's iteration

$$x^{k+1} = x^k - (\partial\nabla_x L(x^k, w(x^k), C))^{-1} \cdot \nabla_x L(x^k, w(x^k), C)$$

is locally well-defined and locally Q-quadratically convergent to x^*.

Proof. It was shown in the proof of Proposition 2.2 that there is a neighborhood of x^* in which $\partial\nabla_x L(x, w(x), C)$ is continuous and $\partial\nabla_x L(x^*, w(x^*), C)$ is positive definite for all $C \geqslant \hat{C}$ ($\hat{C}$ large enough). Hence, Newton's iteration is locally well-defined. The conclusion on convergence rate is obvious (see Thm 10.2.2 [11]).

Now we consider the secant methods for system (17)

ALGORITHM

Step 0: Determine x^0; $w^0 = w(x^0)$; H_0; set $k := 0$.

Step 1: $x^{k+1} = x^k - H_k \cdot \nabla_x L(x^k, w^k, C)$, $w^{k+1} = w(x^{k+1})$, $s^k = x^{k+1} - x^k$

$y^k = \nabla_x L(x^{k+1}, w^{k+1}, C) - \nabla_x L(x^k, w^k, C)$.

Step 2: $H_{k+1} = \mathscr{H}(s^k, y^k, H_k)$.

Step 3: go to Step 1.

($\mathscr{H}$ is an inverse update.)

Assume that the following inverse updates are used (see [2])

(53) $$\mathscr{H}_{\mathrm{B}}(s, y, H) = H + \frac{(s - Hy)s^T \cdot H}{s^T \cdot Hy},$$

(54) $$\mathscr{H}_{\mathrm{DFB}}(s, y, H) = H + \frac{s \cdot s^T}{s^T \cdot y} - \frac{Hyy^T H}{y^T Hy},$$

(55) $$\mathscr{H}_{\mathrm{BFGS}}(s, y, H) = \left(\frac{I - sy^T}{y^T s}\right) H \left(\frac{I - ys^T}{y^T s}\right) + \frac{ss^T}{y^T s}.$$

THEOREM 7.2. *Suppose that the assumptions of Theorem* 7.1 *hold. Then for* $C \geqslant \hat{C}$ ($\hat{C}$ *large enough*) *the Broyden* (53), *Davidon–Fletcher–Powell* (54), *Broyden–Fletcher–Goldfarb–Shanno* (55) *methods are locally well-defined and locally* Q*-superlinearly convergent to* x^*.

Proof. The methods are locally well-defined, because there is a neighborhood of x^* in which $\partial\nabla_x L(x, w(x), C)$ is continuous and, for $C \geqslant \hat{C}$ ($\hat{C}$ large enough) $\partial\nabla_x L(x^*, w(x^*), C)$ is positive definite. (Proposition 2.2). This also proves that the Broyden method is Q-superlinearly convergent, owing to the assumption (52), (see [1]). Moreover, the derivative $\partial\nabla_x L(x^*, w(x^*), C)$ is symmetric at every critical point x^* (Lemma 5.3, 6.3). Since it is also positive definite, the DFP and BFGS methods are Q-superlinearly convergent (see [1]).

It is worth noticing that the same amount of calculation is required at each stage of this Algorithm as at each stage of the diagonalized quasi-Newton multiplier methods using the projection multiplier update ([7], [13]). However, the last methods are locally Q-linearly convergent ([7]), while here the local Q-superlinear convergence is proved.

Acknowledgment. The author is indebted to Prof. Z. Denkowski who, after reading the manuscript of the present work, provided many helpful suggestions.

References

[1] C. G. Broyden, J. E. Denis, J. J. More, *On the Local and Superlinear Convergence of Quasi-Newton Methods*, J. of the Institute of Math. and its Applications, Vol. 12 (1973).

[2] J. E. Denis, J. J. More, *Quasi-Newton Methods — Motivation and Theory*, SIAM Riview, Vol. 19 (1977).

[3] A. Fiaco, P. McCormic, *Nonlinear Programing: Sequential Unconstrained Minimization Techniques*, (in Russian) Moscow 1972.

[4] R. Fletcher, *A Classes of Methods for Nonlinear Programing with Termination and Convergence Properties*, Integer and Nonlinear Programing, North Holand, Amsterdam 1970.

[5] R. Fletcher, *An Exact Penalty Function for Nonlinear Programing with Inequalities*, Math. Programing, Vol. 5 (1973).

[6] R. Fletcher, *An Ideal Penalty Function for Constrained Optimization*, J. of Inst. of Math. and its Applications, Vol. 15 (1975).

[7] S. T. Gland, *Properties of Updating Methods for the Multipliers in Augmented Lagrangians*, J.O.T.A. Vol. 18 (1979).

[8] S. T. Glad, E. Polak, *A Multiplier Method with Automatic Limitation of Penalty Growth*, Math. Programing, Vol. 17 (1979).

[9] K. Martensson, *A New Approach to Constrained Function Optimization*, J.O.T.A. Vol. 12 (1973).

[10] H. Mukai, E. Polak, *A Primal-Dual Algorithm with Global Convergence Properties for Solving Optimization Problems with Equality Constraints*, Math. Programing Vol. 9 (1975).

[11] J. M. Ortega, W. C. Rheinboldt, *Iterative Solution of Nonlinear Equation in Several Variables*, (in Russian) Moscow 1975.

[12] R. T. Rockafellar, *Augmented Lagrange Multiplier Functions and duality in Nonconvex Programing*, SIAM Journal on Control, Vol. 12 (1974).

[13] A. Tapia, *Diagonalized Multiplier Methods and Quasi-Newton Methods for Constrained Optimization*, J.O.T.A. Vol. 22 (1977).

Received March 1, 1980.