

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:

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

one has to solve an infinite number of unconstrained minimizations of the augmented Lagrangian functional: $\min_x L(x, u^k, C)$ with $u^k \rightarrow u^*$ 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 $\nabla_{xx}^2 L(x, u(x), C)$ 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 $C \geq 0$, 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 $C \geq C_0$ (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, \dots, g_m are twice continuously differentiable functionals on R^n , with $m \leq n$.

We recall that the notation $g(x) = (g_1(x), \dots, g_m(x))^T$, $u = (u_1, \dots, u_m)^T$ $u(x) = (u_1(x), \dots, u_m(x))^T$ 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 ; $\nabla g(x) = (\nabla g_1(x), \dots, \nabla g_m(x))$ and $\nabla^2 f(x), \nabla^2 g_i(x)$ denote their Hessians.

Given $x \in R^n$ we put

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

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

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

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

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

$$(4) \quad g_i(x) = 0 \quad i = 1, \dots, p,$$

$$(5) \quad g_i(x) \leq 0 \quad i = p+1, \dots, m,$$

$$(6) \quad u_i \geq 0 \quad i = p+1, \dots, m,$$

$$(7) \quad u \cdot g(x) = 0.$$

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 $i = p+1, \dots, m$.

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:

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

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

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

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 $u: R^n \rightarrow R^m$ 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):

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

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*

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

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 $\hat{C} \geq 0$ such that for all $C \geq \hat{C}$ the Hessian $\nabla_{xx}^2 L(x^*, u^*, C)$ 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 $\hat{C} \geq 0$ such that for all $C \geq \hat{C}$*

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

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

$$Z(x, C) = \{i = p+1, \dots, m \mid Cg_i(x) = -u_i(x)\}$$

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

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

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

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

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

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:

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

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

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

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

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

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 $C \geq \hat{C}$.

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 $\hat{C} \geq 0$ such that for all $C \geq \hat{C}$*

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

is positive semidefinite.

Proof. By the second order necessary condition for a local minimum (8) the Hessian $\nabla_{xx}^2 F(x^*, u^*)$ 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 $y^T \cdot \nabla_{xx}^2 F(x^*, u^*) \cdot y = 0$ we get

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

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

$$(17) \quad \nabla_x L(x, u(x), C) = 0.$$

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 -quadratic 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: $x^k \neq x^*$, $\nabla_x L(x^k, u(x^k), C_k) = 0$, $k = 1, 2, \dots$, and $C_k \rightarrow \infty$, $x^k \rightarrow \bar{x} \in S$ (by the compactness of S) as $k \rightarrow \infty$. We remember that

$$(18) \quad \begin{aligned} \nabla_x L(x^k, u(x^k), C_k) &= \nabla f(x^k) + \sum_{i=1}^p u_i(x^k) \cdot \nabla g_i(x^k) + C_k \sum_{i=1}^p g_i(x^k) \cdot \nabla g_i(x^k) \\ &\quad + \sum_{i=p+1}^m (C_k g_i(x^k) + u_i(x^k))_+ \cdot \nabla g_i(x^k) \\ 0 &= \lim_{k \rightarrow \infty} \left\{ \nabla f(x^k) + \sum_{i=1}^p u_i(x^k) \cdot \nabla g_i(x^k) + \right. \\ &\quad \left. + C_k \left[\sum_{i=1}^p g_i(x^k) \cdot \nabla g_i(x^k) + \sum_{i=p+1}^m \left(g_i(x^k) + \frac{u_i(x^k)}{C_k} \right)_+ \cdot \nabla g_i(x^k) \right] \right\}. \end{aligned}$$

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

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

as $k \rightarrow \infty$, 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

$$\begin{aligned} g_i(\bar{x}) &= 0 & i = 1, 2, \dots, p, \\ g_i(\bar{x}) &\leq 0 & i = p+1, \dots, m \end{aligned}$$

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

Setting

$$(19) \quad \begin{aligned} \bar{u}_i &= \lim_{k \rightarrow \infty} (C_k g_i(x^k) + u_i(x^k)) & i = 1, 2, \dots, p \\ \bar{u}_i &= \lim_{k \rightarrow \infty} (C_k g_i(x^k) + u_i(x^k))_+ & i = p+1, \dots, m \end{aligned}$$

we have $\bar{u}_i \geq 0$ for $i = p+1, \dots, m$. 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 $k \rightarrow \infty$, 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 $d_k = \frac{x^k - x^*}{\|x^k - x^*\|}$ we can assume, without loss of generality that $d_k \rightarrow d$.

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 $C \geq C_0$, then $Cg_i(x) + u_i(x) \leq 0$ for all $i \notin I(x^*)$. This proves that $J(x^k, C_k) = \{1, \dots, p\} \cup \{i = p+1, \dots, m \mid C_k g_i(x^k) + u_i(x^k) > 0\} \subset I(x^*)$ for all $k \geq k_0$, where k_0 is large enough.

Let us denote

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

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 $k \geq k_0$.

