LAGRANGE MULTIPLIERS ESTIMATES FOR CONSTRAINED MINIMIZATION

L. F. ESCUDERO

We discuss in this work the first-order, second-order and pseudo-second-order estimations of Lagrange multipliers in nonlinear constrained minimization. The paper also justifies estimations and strategies that are used by two nonlinear programming algorithms that are also briefly described.

1. INTRODUCTION

Consider the optimization problem, NLP (Non-Linear Problem)

minimize $F(X)$ $X \in R^n$ (1.1)

subject to $c_i(X) = 0$ $i = 1,2,\dots,m_1$ (1.2)

$c_i(X) \geq 0$ $i = m_1+1,\dots,m$ (1.3)

The functions $F(X)$ and $c_i(X)$ are prescribed nonlinear functions. The function $F(X)$ is usually termed the <u>objective function</u> and the set $\{c_i(X)\}$ is the <u>set of constraint functions</u>. It will be assumed for simplicity that F and $\{c_i\}$ are twice continuously differentiable. Let $c = (c_1, c_2, \dots, c_m)^t$. Under certain mild conditions (see below) on F and c there is defined at solution $\overset{*}{X}$ of NLP a set of scalars known as <u>Lagrange multipliers.</u> These multipliers express the gradient vector $g(X)$ of F as a linear combination of the gradient vectors $\nabla c_i(X)$ (each vector being a column of the Jacobian matrix A) of those element-functions c_i of c which are zero (constraint functions termed <u>active</u>) at $\overset{*}{X}$.

At points which approximate $\overset{*}{X}$ it is possible to define <u>estimates</u> of these Lagrange multipliers. Methods which determine $\overset{*}{X}$ use these estimates in several ways /2/:

a) In conjunction with other information analyze if a given point X is a good approximation to $\overset{*}{X}$.

b) The constrained problem can be transformed into a sequence of unconstrained problems in which these estimates are parameters of the unconstrained objective function.

c) Many methods determine a step-direction by solving a subproblem whose formulation depends on the Lagrange multipliers estimates

This paper is primarily concerned with estimating the Lagrange multipliers of quadratic programming, nonlinear programming with linear constraints and general constrained nonlinear programming; the leading paper by Gill and Murray /8/ is revisited.

We shall define:

The gradient vector

$g(X)$ $(\equiv g)$ as the vector whose j-th element is $\delta F(X)/\delta X_j$

The Hessian matrix

$G(X)$ $(\equiv G)$ as the symmetric matrix whose (i,j)-th element is $\delta^2 F(X)/\delta X_i \delta X_j$.

and the Jacobian matrix

$A(X)$ $(\equiv A)$ as the matrix whose i-th column is the gradient $\nabla c_i(X)$ $(\equiv \nabla c_i)$.

The algorithms concerned with this paper are assumed to generate a <u>sequence of estimates</u> $\{X^k\}$ of $\overset{*}{X}$, by generating a step-direction δ^k, a steplength α^k and a <u>step</u> p^k such that $p^k = \alpha^k \delta^k$ and $X^k = X^{k-1} + p^k$.

L.F. Escudero, Centro de Investigación UAM-IBM. Paseo Castellana, 4. Madrid-1

Article rebut l'abril de 1981.

2. QUADRATIC PROGRAMMING

In general there are no major difficulties in computing Lagrange multipliers estimates for quadratic problems. We consider the quadratic case in detail because methods for estimating Lagrange multipliers for both general-nonlinear constrained problems and linearly constrained problems use the quadratic case. One important feature of this case is that it is possible to define exactly the Lagrange multipliers associated with an equality constrained subproblem at any point.

The methods that estimate the multipliers as we describe in Secs. 2 (quadratic case) and 3 (linearly constrained case) require that the set of active constraints at the current point contains only constraints which are satisfied exactly at the current point. But the general-non-linear constrained case (sec 4) also requires to include in this set those constraints that are violated at the current point. Formally the (QP) quadratic programming /5/ can be expressed.

$$\text{minimize } Q(X) = h^tX + \frac{1}{2} X^tGX \quad (QP) \tag{2.1}$$

$$\text{subject to } A^tX \geqq b \tag{2.2}$$

where G is a constant nxn symmetric matrix, A^t is the mxn constraints matrix, h is a n-vector, b is a m-vector and X is the unknown n-vector.

Let $\overset{*}{A}{}^t$ be the $\overset{*}{t}$xn matrix of constraints active at point $\overset{*}{X}$ and $\overset{*}{b}$ the $\overset{*}{t}$-vector of right-hand-side corresponding to $\overset{*}{A}$. If the following necessary and sufficient conditions hold (see /14, 18, 8 and 11/ among others) then $\overset{*}{X}$ is a global minimum of QP if G is positive definite; otherwise, it is only a local strong minimum.

i) point $\overset{*}{X}$ is feasible.

ii) $\overset{*}{A}$ is a full column rank matrix. It is only necessary for the uniqueness of the Lagrange multipliers. It is important from the computational point of view.

iii) Lagrange multipliers are such that

$\overset{*}{\mu}_i \gtreqless 0$ for $i=1,2,\dots,m_1$ (equality constraints, i.e. set E)

$\overset{*}{\mu}_i = 0$ for $i=m_1+1,\dots,m$ (inequality constraints), if the i-th constraint is non-active (i.e. $A_i^t\overset{*}{X}>b_i$)

$\overset{*}{\mu}_i \geq 0$ for $i=m_1+1,\dots,m$ (inequality constraints), if the i-th constraint is active (i.e. $A_i^t\overset{*}{X}=b_i$).

Let $C_1=\{i|\mu_i>0\}$ and $C_2=\{i|\mu_i=0\}$ be the sets of active inequality constraints for which the corresponding Lagrange multipliers are respectively positive and zero. Let $I = E\cup C_1\cup C_2$ be the set of the $\overset{*}{t}$ active constraints; for this set the vector $\overset{*}{\mu}$ is such that

$$\overset{*}{A}\overset{*}{\mu} = \overset{*}{g} = G\overset{*}{X} + h \tag{2.3}$$

iv) Hessian matrix

$$\overset{*}{H} = \overset{*}{Z}{}^tG\overset{*}{Z} \tag{2.4}$$

is positive definite, that is

$$Y^t\overset{*}{Z}{}^tG\overset{*}{Z}Y > 0 \tag{2.5a}$$

where $\overset{*}{Z}$ is a nx(n-$\overset{*}{t}$) full column rank matrix such that

$$\overset{*}{A}{}^t\overset{*}{Z} = 0 \tag{2.5b}$$

Since $\overset{*}{A}$ belongs to the set of active constraints, $\overset{*}{A}{}^t\overset{*}{X} = \overset{*}{b}$ and for any $X=\overset{*}{X}+p=\overset{*}{X}+\alpha\delta$ feasible solution to this system of active constraints, it results

$$\overset{*}{A}{}^t(\overset{*}{X}+\alpha\delta) = b; \; \overset{*}{A}{}^t\delta = 0 \tag{2.5c}$$

Considering eqs. (2.5b) and (2.5c), then Y will satisfy

$$\delta = \overset{*}{Z}Y \tag{2.5d}$$

That is, for each vector Y there is associated a vector δ that takes the step direction from $\overset{*}{X}$ to feasible point X in the space of active constraints. Then, condition (iv) (2.5a) is equivalent to condition

$$\delta^tG\delta > 0 \tag{2.5e}$$

such that $\overset{*}{A}{}^t\delta = 0$. If $C_2 \neq \emptyset$ then condition (2.5e) must be extended to the feasible region $\overset{*}{A}{}_i^t\delta>0$ for $\forall i\in C_2$. There are several ways to obtain $\overset{*}{Z}$. See /7 and 16/.

Formally, $\overset{*}{X}$ and $\overset{*}{\mu}$ define stationary points of Lagrange function

$$L(X,\mu) = F(X) - \mu^t(\overset{*}{A}{}^tX-b) \tag{2.6a}$$

and, then, satisfy the system

$$\overset{*}{g} - \overset{*}{A}\overset{*}{\mu} = 0 \tag{2.6b}$$

$$\overset{*}{A}{}^t\overset{*}{X} = \overset{*}{b} \tag{2.6c}$$

For QP these eqs. have the form (see /2/)

$$\begin{pmatrix} -G & \overset{*}{A} \\ \overset{*}{A}{}^t & 0 \end{pmatrix} \begin{pmatrix} \overset{*}{X} \\ \overset{*}{\mu} \end{pmatrix} = \begin{pmatrix} h \\ \overset{*}{b} \end{pmatrix} \qquad (2.7)$$

In Sec. 4 we will see that for the non-linear constrained problem, it is not the Hessian matrix of objective function F(X) that is required to be positive definite, but the Hessian matrix of Lagrange function $L(X,\mu)$. In fact, in the linear case both matrices are the same because of the Hessian matrices of linear constraint functions do not exist.

Using the strategy of active constraints at each iteration (in the algorithm described in /5/ they will be the constraints whose slack variable is non-basic and then with zero value), we have an equality-constrained subproblem (EQP) of the form

$$\text{minimize } Q(X) = h^t X + \frac{1}{2} X^t G X \qquad (2.8)$$
$$\text{subject to } \hat{A}^t X = \hat{b} \qquad (2.9) \qquad \text{(EQP)}$$

where the full column rank matrix $\hat{A}$ consists of a selection of $\hat{t}$ columns of matrix A and $\hat{b}$ is the vector composed of the corresponding elements of b. We assume that matrix $\hat{Z}^t G \hat{Z}$ is positive definite where $\hat{Z}$ is a $n \times (n-\hat{t})$ full column rank matrix such that $\hat{A}^t \hat{Z} = 0$. If $\hat{Z}^t G \hat{Z}$ is not positive definite there is no interest in calculating the Lagrange multipliers since the optimum $\hat{X}$ of EQP will not be the optimum $\overset{*}{X}$ of QP (conditions (i)-(ii)). See in /9 and 10/ an algorithm that in this case generates a direction of negative curvature.

Associated with EQP is the t-vector $\hat{\mu}$ that satisfies the overdetermined system of linear eqs.

$$\hat{A}\,\mu = \hat{g} \qquad (2.10)$$

If system (2.10) is compatible, there is only one vector $\hat{\mu}$ that satisfies (2.10); and $\hat{\mu}$ corresponds to $\hat{X}$. If there are elements of $\hat{\mu}$ that are negative and they correspond to the inequality constraints of QP, $\hat{X}$ is not the optimum $\overset{*}{X}$ of QP. If in this case we drop from EQP these constraints, we may obtain a better feasible point $X(Q(X) < Q(\hat{X}))$ in QP; see i.e. /11/.

In summary, the optimum point $\overset{*}{X}$ of QP is the optimum point $\hat{X}$ of EQP (see below) plus the following additional condition: Lagrange multipliers $\hat{\mu}_i$ associated to active inequality constraints (1.3) must be non-negative.

By premultiplying both sides of eq. (2.10) by $\hat{Z}^t$, it results.

$$\hat{Z}^t \hat{A}\,\hat{\mu} = \hat{Z}^t \hat{g} = 0 \qquad (2.11)$$

If $Y^t \hat{Z}^t G \hat{Z} Y > 0$, where Y is any vector such that $\delta = \hat{Z}Y$ is the step direction from $\hat{X}$ to any feasible point, it results that $\hat{X}$ is the optimum point of EQP. In effect, the Taylor series of Q(X) from $Q(\hat{X})$ are

$$Q(\hat{X}+\delta) = Q(\hat{X}) + \hat{g}^t \delta + \frac{1}{2}\,\delta^t G\,\delta \qquad (2.12)$$

Considering eq. (2.11) and substituting δ for $\hat{Z}Y$, it results

$$Q(\hat{X}+\delta) = Q(\hat{X}) + \frac{1}{2}\,Y^t \hat{Z}^t G \hat{Z} Y \qquad (2.13)$$

Notating $\hat{H} \equiv \hat{Z}^t G \hat{Z}$, if $Y^t \hat{H} Y > 0 \rightarrow Q(\hat{X})$ is a strong local minimum.

Points that satisfy eq. (2.11) are termed constrained stationary points and $\hat{Z}^t g$ (where $g \equiv g(X)$) will be the reduced or projected gradient. The solution of eq. (2.10) is

$$\hat{\mu} = \hat{A}^+ \hat{g} \qquad (2.14)$$

where $\hat{A}^+$ is the pseudo-inverse of $\hat{A}$. There are several computational methods to obtain $\hat{A}$. A well-known method computes $\hat{A}^+$ such that (see /7/)

$$\mu = (\hat{A}^t \hat{A})^{-1} \hat{A}^t g \qquad (2.15)$$

In effect, μ has to satisfy (2.10), what is is analogous to minimize the norm

$$\| g - \hat{A}\mu \|^2 \qquad (2.16)$$

whose value is zero for $g=\hat{g}$. The vector that minimizes (2.16) is termed minimal least square solution. The necessary and sufficient condition (see /17/, p. 309) for (2.16) be a minimum is that

$$\hat{A}^t (g - \hat{A}\mu) = 0 \qquad (2.17)$$

Then, theoretically μ is calculated by formula (2.15) if $(\hat{A}^t \hat{A})$ is non-singular; but from a practical point of view even in this case (since the computers have finite precision) the rounding errors produce a solution μ that does not satisfy eq. (2.17). Then formula (2.15) is not recommended. Using it, point X may be the optimum $\hat{X}$ of EQP and this may be the optimum $\overset{*}{X}$ of QP and, since the ill-condi tioning of $\hat{A}^+$, the computed solution μ with formula (2.15) may not satisfy eq. (2.10). On the other hand, the solution of formula (2.15) may give incorrect information, dropping from the active set an inequality constraint of QP if the associated Lagrange multiplier to the corresponding active constraint in EQP is negative (and, because of rounding errors, it is ill-calculated); as a result, this unstable formula may produce a poor converging algorithm.

In order to avoid matrix $\hat{A}^+$, Gill and Murray /8/ (see also /13/) suggest the following method that, based on the QR-factorization of matrix $\hat{A}$ /12, 17, 4, 6/ is more stable than formula (2.15). Let $\hat{A}$ be factorized into the form

$$Q^t\hat{A} = \begin{pmatrix} R \\ 0 \end{pmatrix} \qquad (2.18)$$

where Q is a nxn non-symmetric orthogonal ($Q^tQ=I$) matrix for which it also holds $QQ^t=I$ and R is a $\hat{t}x\hat{t}$ upper-triangular matrix (no necessarily with identity diagonal; see /6/). Matrix Q may be partitioned so that

$$Q = (Q_1\ Q_2) \qquad (2.19)$$

where Q_1 is a $nx\hat{t}$ matrix and Q_2 is a $nx(n-\hat{t})$ matrix, so that

$$\hat{A} = Q_1R \qquad (2.20)$$

Recall that $\hat{A}$ is a full column rank matrix; it results:

a) R is a non-singular matrix

b) $I = QQ^t = (Q_1\ Q_2)\begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix} = Q_1Q_1^t + Q_2Q_2^t$

$I = Q^tQ$ then $Q_1^tQ_1 = I \quad (Q_1^t = Q_1^{-1}),\ Q_1^tQ_2 = 0,$

$$Q_2^tQ_1 = 0,\ Q_2^tQ_2 = I\ (Q_2^t=Q_2^{-1}). \qquad (2.21)$$

c) $$\begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix}\hat{A} = \begin{pmatrix} Q_1^t\hat{A} = R \\ Q_2^t\hat{A} = 0 \end{pmatrix} \qquad (2.22)$$

then matrix Q_2 may be taken as matrix $\hat{Z}$.

d) The vector μ that minimize the norm (2.16), i.e. the minimal least square solution μ_L, is found (see /11 and 6/ among others) by solving

$$R\mu_L = Q_1^tg \qquad (2.23)$$

Formula (2.23) is more stable than formula (2.15), since from (2.23) it results

$$\mu_{L_{\hat{t}}} = Q_{1_{\hat{t}}}^tg/R_{\hat{t}\hat{t}} \qquad (2.24)$$

$$\mu_{L_i} = (Q_{1_i}^tg - \sum_{r=i+1}^{\hat{t}} R_{ir}\mu_{L_r})/R_{ii} \quad \text{for } i=\hat{t}-1,\dots,1$$

where Q_{1_i} is the i-th column of Q_1. Then formula (2.24) is not losing precision, if the computation Q_1^tg is correctly done. It is interesting to point out that in this computation it is implicit the orthogonality of Q_1; if while obtaining Q_1 there are rounding errors then this method, although better than formula (2.15), is also unstable. Escudero /4 and 6/ describes an alternate method.

The vector μ_L (2.24) (or the alternate procedure) minimizes the norm (2.16). If the norm is zero, condition (2.10) is satisfied and point X to which μ_L is associated is the optimum point $\hat{X}$ of EQP (being $g=\hat{g}$) and then, μ_L is the <u>exact</u> vector of Lagrange multipliers of EQP. But if the norm is not zero, point X is not the optimum $\hat{X}$ and then $g(X)$ $(\equiv g) \neq g(\hat{X})$ $(\equiv \hat{g})$. In this case, the next iteration of the algorithm to minimize EQP is to be executed. How close the <u>approximation</u> μ_L is to $\hat{\mu}$?. If we denote the step to the solution $\hat{X}-X$ by d, since EQP is quadratic it results

$$\hat{g} = G\hat{X} + h = GX + h + G(\hat{X}-X) = g + Gd \qquad (2.25)$$

Eq. (2.23) will be for $\hat{X}$

$$R\hat{\mu} = Q_1^t(g+Gd) = Q_1^tGd + R\mu_L \qquad (2.26)$$

from which we obtain

$$\hat{\mu} = R^{-1}Q_1^tGd + \mu_L \qquad (2.27)$$

$$\|\hat{\mu} - \mu_L\| = \|R^{-1}Q_1^tGd\| = M\,\|d\| \qquad (2.28)$$

where M is a constant. The approximation μ_L of Lagrange multipliers whose difference with $\hat{\mu}$ is of this form is termed <u>first-order estimate</u>.

The algorithm described in /5/ obtains $\hat{\mu}$ using the dual of the LP model in which EQP is converted at point $\hat{X}$. The result is the same.

An important simplification occurs when G is the identity matrix, such that (see /7/)

$$\hat{\mu} = \mu_L \qquad (2.29)$$

Note that this is an expression for the exact Lagrange multipliers but the formula is computed at an arbitrary point X. Then -- for G=I eqs. (2.24) give directly μ_L and by using eq. (2.10) we obtain $\hat{X}$.

Returning to the general case EQP, we may -- note that using eq. (2.25) in eq. (2.10), we have

$$\hat{A}\hat{\mu} = g + Gd \qquad (2.30)$$

If G is non-singular, we have

$$G^{-1}\hat{A}\,\hat{\mu} = G^{-1}g + d \qquad (2.31)$$

and by premultiplying eq. (2.31) by $\hat{A}^t$ and - considering eq. (2.5e), it results

$$\hat{\mu} = (\hat{A}^t G^{-1}\hat{A})^{-1}\hat{A}^t G^{-1} g \qquad (2.32)$$

Eq. (2.32), unlike eq. (2.15), obtains exactly vector $\hat{\mu}$, at any arbitrary point X. But - both eqs. are highly unstable /7 pp. 45-47/.

An alternate procedure that has not the above inconvenience, obtains d and, by using the Q_1R factorization of $\hat{A}$ in eq. (2.30), gives the formula

$$\hat{\mu} = R^{-1}Q_1^t(g + Gd) \qquad (2.33)$$

The method for computing $d(\equiv\hat{X}-X)$ is as follows: See eqs. (2.5) and note that we may -- write $d=\hat{Z}\psi$. Then eq. (2.30) gives

$$\hat{Z}^t\hat{A}\hat{\mu} = \hat{Z}^t g + \hat{Z}^t G\hat{Z}\psi = 0 \qquad (2.34)$$

$$\psi = -(\hat{Z}^t G\hat{Z})^{-1}\hat{Z}^t g \qquad (2.35)$$

$$d = -\hat{Z}(\hat{Z}^t G\hat{Z})^{-1}\hat{Z}^t g \qquad (2.36)$$

Matrix $\hat{Z}$ may be matrix Q_2 (eq. 2.22). See in /5/ another method to express $\hat{Z}$. Eq. (2.36) is not less stable then eq. (2.32).

The algorithm described in /5/, that also -- uses matrix $(\hat{Z}^t G\hat{Z})$, explicitly uses the Lagrange multipliers; e.g. to analyze (by -- using the reduced cost of the slack varia-- bles) when a constraint must be dropped from the set of active constraints or when point $\hat{X}$ is the optimum point $\overset{*}{X}$ of QP.

In summary, there are basically four ways to estimate the Lagrange multipliers vector $\hat{\mu}$ for the quadratic problem EQP: formulae -- (2.15), (2.24), (2.32) and (2.33). The first two obtain μ_L (an approximation of $\hat{\mu}$), the - last two obtain exactly the vector $\hat{\mu}$, the -- first and the third are very unstable, the - fourth if recommended.

3. MINIMIZATION OF A GENERAL NON-LINEAR FUNCTION SUBJECT TO LINEAR CONSTRAINTS

Consider the optimization problem, LCP (Linear Constrained Problem)

$$\text{minimize} \quad F(X) \quad X\in R^n \qquad (3.1)$$

(LCP)

$$\text{subject to } A^tX \geqq b \qquad (3.2)$$

Function F(X) is non-linear twice continuously differentiable at least for feasible --- points.

Let $\overset{*}{A}{}^t$ be the $\overset{*}{t}$xn matrix of active cons---- traints at point $\overset{*}{X}$ and $\overset{*}{b}$ the $\overset{*}{t}$-vector of right-hand-side corresponding to $\overset{*}{A}$ (i.e. --- $\overset{*}{A}{}^t\overset{*}{X} = \overset{*}{b}$). If the following conditions hold -- (see /14, 18, 8 and 11/ among others) then $\overset{*}{X}$ is a strong local minimum of LCP (we don't consider, as in the quadratic problem, the case for which $\overset{*}{X}$ is also a maximum):

i) point $\overset{*}{X}$ is feasible
ii) $\overset{*}{A}$ is a full rank matrix. It is only necessary for the uniqueness of the La--- grange multipliers. It is also important from the computational point of view.
iii) Lagrange multipliers are such that
   $\overset{*}{\mu}_i \gtreqless 0$ for $i=1,2,\ldots,m_1$ (equality constraints, i.e. set E).
   $\overset{*}{\mu}_i = 0$ for $i=m_1+1,\ldots,m$ (inequality constraints) if the i-th constraint is non-active.
   $\overset{*}{\mu}_i \geq 0$ for $i=m_1+1,\ldots,m$ (inequality constraints) if the i-th constraint is --- active.

Let C_1, C_2 and I be the same sets notation - used in sec. 2. For the set I, vector $\overset{*}{\mu}$ is - such that

$$\overset{*}{A}\overset{*}{\mu} = \overset{*}{g} \qquad (3.3)$$

iv) Hessian matrix

$$\overset{*}{H} = \overset{*}{Z}{}^t\overset{*}{G}\overset{*}{Z} \qquad (3.4a)$$

is positive definite, that is

$$Y^t\overset{*}{Z}{}^t\overset{*}{G}\overset{*}{Z}Y > 0 \qquad (3.4b)$$

where $\overset{*}{G}\equiv G(\overset{*}{X})$ being $G\equiv G(X)$ the Hessian matrix of F(X) at point X, and being $\overset{*}{Z}$ a nx(n-$\overset{*}{t}$) -- full column rank matrix such that

$$\overset{*}{A}{}^t\overset{*}{Z} = 0 \qquad (3.4c)$$

In a similar way to the quadratic case (Sec. 2) for $\delta = X-\overset{*}{X}$ we have

$$\delta = \overset{*}{Z}Y \qquad (3.4d)$$

Eq. (3.4b) is equivalent to eq. (3.4e)

$$\delta^t\overset{*}{G}\,\delta > 0 \qquad (3.4e)$$

that is, matrix $\overset{*}{G}$ must be positive definite at least for the step direction δ from $\overset{*}{X}$ to any other feasible point X in $\overset{*}{A}{}^tX=\overset{*}{b}$.

Then,

$$\overset{*}{A}{}^t\delta = 0 \qquad (3.4f)$$

If $C_2 \neq \emptyset$ then condition (3.4e) must be extended to the feasible region $\overset{*}{A}{}_i^t\delta > 0$ for $\forall i \in C_2$. We may note the similarity of the conditions to the quadratic case. Conditions (i)-(iii) and $\overset{*}{H}$ being a positive semi-definite matrix are necessary optimality conditions. Conditions (i)-(iv) are only sufficient conditions. Associated to LCP we have the (ELCP) equality-linear constrained problem

$$\text{minimize} \quad F(X) \qquad (3.5)$$
$$\text{subject to} \quad \hat{A}^tX = \hat{b} \qquad (3.6) \qquad \text{(ELCP)}$$

where the nx$\hat{t}$ full column rank matrix $\hat{A}$ is a selection of $\hat{t}$ columns of matrix A and $\hat{b}$ is the $\hat{t}$-vector composed of the corresponding - elements of b.

The optimum point $\hat{X}$ of ELCP must satisfy similar conditions to conditions (i)-(iv) of - LCP, but without any constraint in the sign (positive, zero or negative) of the $\hat{t}$ ele-- ments of $\hat{\mu}$. If the $\hat{t}-m_1$ elements of $\hat{\mu}$ corresponding to the active inequality constraints of LCP are non-negative the optimum $\hat{X}$ is also $\overset{*}{X}$. In other case, these constraints are - to be dropped from ELCP (may be other constraints are to be added to ELCP) and the procedure goes to the next iteration: optimization of the new ELCP.

Then, the vector $\hat{\mu}$ of Lagrange multipliers - of ELCP must satisfy the over-determined system of linear eqs.

$$\hat{A}\hat{\mu} = \hat{g} \qquad (3.7)$$

But, unlike the quadratic case, it is only possible to obtain $\hat{\mu}$ if we know $\hat{X}$ (and its gradient $\hat{g}$). On any arbitrary feasible point X we cannot obtain exactly the value of $\hat{\mu}$;- only we may obtain the <u>first-order estimate</u> μ_L in similar way to eq. (2.23)

$$R\mu_L = Q_1^tg \qquad (3.8)$$

and directly

$$\mu_L = R^{-1}Q_1^tg \qquad (3.9)$$

Estimation μ_L (eqs. 3.8 or 3.9) is calcula-- ted in a similar way to estimation μ_L in EQP (eqs. 2.23 or 2.24). For completeness, the - estimation μ_L that is similar to formula -- (2.15) is

$$\mu_L = (\hat{A}^t\hat{A})^{-1}\hat{A}^tg \qquad (3.10)$$

This formula is also very unstable and it is not recommended. Expanding the gradient $\hat{g}$ in Taylor series around point X gives the analogue of eqs. (2.25) - (2.28).

$$\hat{A}\hat{\mu} = \hat{g}(X+d)$$
$$= g + Gd + 0(\|d\|^2) \qquad (3.11)$$

where $d=\hat{X}-X$. Since R is non-singular and by using eqs. (2.22) and (3.8), we have

$$Q^t\hat{A}\hat{\mu} = Q^tg+Q^tGd+0(\|d\|^2) \qquad (3.12)$$

$$\hat{\mu} = R^{-1}Q_1^tg+R^{-1}Q_1^tGd+0(\|d\|^2) \qquad (3.13)$$

$$\|\hat{\mu}-\mu_L\| = \|R^{-1}Q_1^tGd\| + 0(\|d\|^2) =$$
$$= M\|d\| + 0(\|d\|^2) \qquad (3.14)$$

that is a logical extension of eq. (2.28) -- /8/. Formulae of μ_L (3.9) and (3.10) are termed <u>first-order estimate</u>.

If we know the Hessian matrix $G \equiv G(X)$ (or it is not a time consuming calculation), there are two alternate formulae for calculating the estimate of $\hat{\mu}$. They are similar to the - quadratic case; but in the quadratic case, $\hat{\mu}$ is obtained <u>exactly</u> and for this case only - we may obtain an <u>approximation</u> of $\hat{\mu}$. For -- This we use the Taylor series (3.11); note that $\hat{g}$ can be expressed as follows

$$\hat{g} = g+Gd+D \qquad (3.15)$$

where D takes the truncation of Taylor se--- ries.

Let δ be the estimation of d. Then eq. (3.7) by using eq. (3.15), gives

$$\hat{A}\hat{\mu} = g + G\delta + D' \quad (3.16)$$

where D' takes the deviation due to the estimate δ and the Taylor series truncation.

If $G(\equiv G(X))$ is non-singular, by premultiplying successively eq. (3.16) by G^{-1} and $\hat{A}^t$, we have

$$\hat{\mu} = (\hat{A}^t G^{-1}\hat{A})^{-1}\hat{A}^t G^{-1} g + \bar{D}' \quad (3.17)$$

then, the estimate $\mu_G^{(1)}$ of $\hat{\mu}$ will be

$$\mu_G^{(1)} = (\hat{A}^t G^{-1}\hat{A})^{-1}\hat{A}^t G^{-1} g \quad (3.18)$$

Eq. (3.18) is analogous to eq. (2.32). They are unstable and not recommended.

An alternate way (similar to the quadratic case), that has not the above inconvenience, obtains the estimation δ of step direction $d=\hat{X}-X$ so that, by using in eq. (3.16) the Q_1R factorization of $\hat{A}$ (eq. 2.20), we have

$$\hat{\mu} = R^{-1}Q_1^t(g+G\delta) + D'' \quad (3.19)$$

from where the estimation $\mu_G^{(2)}$ of $\hat{\mu}$ will be

$$\mu_G^{(2)} = R^{-1}Q_1^t(g+G\delta) \quad (3.20)$$

that is similar to formula (2.33) in the quadratic case. Formulae of μ_G (3.18) and (3.20) are termed <u>second-order estimate</u>.

There are several ways to obtain δ. One of them is as follows: It is similar to the method for the quadratic case (eqs. 2.34 to 2.36). Considering eqs. (3.4) and noting that $d=\hat{Z}\phi \therefore \delta=\hat{Z}\nu$, where ν is the estimation of ϕ, eq. (3.16) gives

$$\hat{Z}^t\hat{A}\hat{\mu} = 0 = \hat{Z}^t g + \hat{Z}^t G\hat{Z}\nu + \bar{D}'' \quad (3.21)$$

from where, by deleting the unknown D" from eq. (3.21) since we assume that ν is the best available estimation of ϕ, we have the best computable estimation δ of $d=\hat{X}-X$ such that

$$\nu = -(\hat{Z}^t G\hat{Z})^{-1}\hat{Z}^t g \quad (3.22)$$

$$\delta = -\hat{Z}(\hat{Z}^t G\hat{Z})^{-1}\hat{Z}^t g \quad (3.23)$$

The reasons for the difference between δ (3.23) and d (2.36), quadratic case, are: (1) in ELCP matrix G is the Hessian matrix at arbitrary point X (G is an estimation of $\hat{G}$), and (2) since F(X) is a general non-linear function the vector $g+G\delta$ is not necessarily the vector $\hat{g}$ even for the case in which $\delta=d$, except if $\delta=d=0$ (i,e., for $\hat{X}=X$).

Hessian matrix $\hat{Z}^t G\hat{Z}$ is termed <u>reduced Hessian matrix</u> and gradient $\hat{Z}^t g$ is termed <u>reduced gradient</u>. There are several ways to obtain $\hat{Z}$; e.g. $\hat{Z}$ may be matrix Q_2 (eq. 2.22). Escudero /5/ describes another way to represent $\hat{Z}$; usually it is <u>never</u> explicitly computed.

Normally $G(\equiv G(X))$ is unknown (or it needs much time for its calculation). Usually G is approximated by matrix B (see e.g. in /3/ a survey of the main important methods to calculate B). The estimations $\mu_B^{(1)}$ and $\mu_B^{(2)}$, similar respectively to estimations $\mu_G^{(1)}$, (3.18) and $\mu_G^{(2)}$ (3.20), are.

$$\mu_B^{(1)} = (\hat{A}^t B^{-1}\hat{A})^{-1}\hat{A}^t B^{-1} g \quad (3.24)$$

$$\mu_B^{(2)} = R^{-1}Q_1^t(g+B\delta) \quad (3.25a)$$

where

$$\delta = -\hat{Z}(\hat{Z}^t B\hat{Z})^{-1}\hat{Z}^t g \quad (3.25b)$$

Note that if $\hat{Z}^t B\hat{Z}$ is positive definite and δ is calculated with formula (3.25b), the overdetermined system of linear eqs.

$$\hat{A}\mu = g + B\delta \quad (3.26)$$

from where eqs. (3.24) and (3.25) are derived, is always compatible: $\mu_B^{(2)}$ (and theoretically $\mu_B^{(1)}$) satisfies $\|A\mu-g-B\delta\|^2 = 0$ <u>whatever value of B</u>, since $\mu_B^{(2)}$ (and $\mu_B^{(1)}$) are the <u>exact vector</u> of Lagrange multipliers of the quadratic programming problem.

$$\min \{g^t\delta + \tfrac{1}{2}\delta^t B\delta \mid \hat{A}^t\delta = 0\} \quad (3.27)$$

See eqs. (2.32) and (2.33). Hence, $\mu_B^{(1)}=\mu_B^{(2)}$ at least theoretically; this estimation is termed <u>pseudo-second-order estimate</u>.

In the rest of this work, the notation μ_G and μ_B will be equivalent, respectively, to $\mu_G^{(2)}$ and $\mu_B^{(2)}$ since $\mu_G^{(1)}$ and $\mu_B^{(1)}$ are not recommended and, from a computational point of view, these calculations must be avoided. What of the three vectors μ_L (3.9), μ_G (3.20) and μ_B (3.25) is the best estimation of $\hat{\mu}$?. Let us revise some properties of μ_L, μ_G and μ_B.

If at point X an element of the first-order estimate μ_L (3.9), say μ_{L_i} (that corresponds to the active inequality constraint i of problem ELCP) is negative and, then, constraint i is dropped from matrix $\hat{A}$ without being X - the optimum point $\hat{X}$, /8/ shows that if δ is a Newton (3.23) or Quasi-Newton (3.25b) step direction, where matrix $\hat{A}$ has been substituted by the nx(n-$\hat{t}$+1) matrix $\bar{Z}$ such that $\bar{A}^t\bar{Z}=0$ (where $\bar{A}$ is the new nx(t-1) active constraints matrix) then δ is not necessarily -- feasible (that is, it is possible that $\hat{A}_i^t\delta<0$) even if the corresponding reduced Hessian is positive definite, except if matrix $\bar{Z}^tB\bar{Z}$ is computed as

$$\bar{Z}^tB\bar{Z} = \begin{pmatrix} \hat{Z}^tB\hat{Z} & 0 \\ 0 & 1 \end{pmatrix} \qquad (3.28)$$

in whose case δ is descent feasible (i.e., - $g^t\delta<0$ and $\hat{A}_i^t\delta>0$). Formula (3.28) is used --- when B is not computed, but the Quasi-Newton approximation $\bar{Z}^tB\bar{Z}$ is obtained and vector -- (0 1) is the unique available alternative to compute its last row and column; see /5/.

Gill and Murray (/8/, theorems 4 and 5) also show that if in the above strategy the first order estimate μ_L is substituted by the --- pseudo-second-order estimate μ_B (3.25a) (or by second-order estimate μ_G (3.20)) such that, after evaluating gradient g and reduced Hessian $\hat{Z}^tB\hat{Z}$ (or $\hat{Z}^tG\hat{Z}$) at point X, δ is calculated by formula (3.25b) (or by formula --- 3.23)) and μ_B (or μ_G) is calculated by using δ (and, then, by using $\hat{Z}^tB\hat{Z}$ or $\hat{Z}^tG\hat{Z}$) then -- the new stepdirection, say δ obtained by formula (3.25b) (or by formula 3.23)) is descent feasible if $\hat{Z}$ is substituted by $\bar{Z}$ such that $\bar{A}^t\bar{Z}=0$ as above and the new reduced Hessian $\bar{Z}^tB\bar{Z}$ (or $\bar{Z}^tG\bar{Z}$) is positive definite -- without requiring the special form (3.28).

In any case, note that the Lagrange Multipliers estimates used in the theorems that - have been referenced above are evaluated at point X for μ_L, and at point X+δ for μ_B and μ_G. Note also that μ_G is better than μ_B, but normally it is not easy (or it is time consuming) to calculate G, then generally μ_G is not used. Note that we obtain μ_B by considering that $\hat{X}$ and $\hat{g}$ are estimated respectively by X+δ and g+Bδ (eq. 3.15). Then, μ_B will be better estimation of $\hat{\mu}$ than μ_L if X+δ and -- g+Bδ are, respectively, better estimations - of $\hat{X}$ and $\hat{g}$ than X and g. Note that if we --- solve ELCP by using successively quadratic - approximations as problem (3.27), such that the new point is X+$\alpha\delta$, where the stepdirection δ is the solution (3.25b) of problem -- (3.27) and α is the steplength such that $\alpha>0$, it seems that the first-order estimation μ_L for X+$\alpha\delta$ is at least as good as the pseudo-second-order estimation μ_B. In summary, if B is not a bad approximation of G, μ_B is better than μ_L for X and g(X); μ_L for X+$\alpha\delta$ and ---- g(X+$\alpha\delta$) is better than μ_L for X and g(X); -- and μ_L for X+$\alpha\delta$ and g(X+$\alpha\delta$) is better than - μ_B. Since δ is obtained with formula (3.25b), if B is a bad approximation of G then it results that point X+$\alpha\delta$ will not be a good approximation of $\hat{X}$. Terms "bad" and "good" are intentionally vague. We may conclude that if μ_{L_i} for X+$\alpha\delta$ and g(X+$\alpha\delta$), and μ_{B_i} (being i - each constraint of the set of $\hat{t}$ active constraints) do not have the same sign or

$$\frac{\|\mu_B - \mu_L\|^2}{\|\mu_B\|^2} \not< \varepsilon \qquad (3.29)$$

where ε is a given tolerance, both estimations are likely poor estimations of $\hat{\mu}$; reasons: B is a poor approximation of G, and/or problem (3.27) is a poor approximation of -- ELCP. Usually, $\varepsilon=10^{-4}$. In any case, it does not seem that μ_L (eq. 3.9) for X+$\alpha\delta$ and ---- g(X+$\alpha\delta$) is worse estimation of $\hat{\mu}$ than μ_B --- (eq. 3.25a).

In fact, even if B is close to G or $\hat{Z}^tB\hat{Z}$ is close to $\hat{Z}^tG\hat{Z}$, μ_L is still better than μ_B,- since g+Gδ is only an approximation of ---- g(X+$\alpha\delta$) and point X+$\alpha\delta$ is a better approximation of $\hat{X}$ than point X+δ. Of course, if -- $\alpha\to1$ and point X+$\alpha\delta$ is close to $\hat{X}$, it means - that X+δ is close to X+$\alpha\delta$ and g+Gδ is close to g(X+$\alpha\delta$), since the quadratic approximation of F(X+$\alpha\delta$) based on X (and, then, based on g and G) is close to F(X+$\alpha\delta$); in this latter case, both estimations μ_L and μ_B converge to $\hat{\mu}$. Note that the superiority of μ_L over μ_B in ELCP (even when B is close to G) is not a discrepancy with the conclusion --- from the comparison of eqs. (2.15) and (2.33) in EQP, since the quadratic approximation to F(X) in ELCP is becoming better when the solution is becoming close to $\hat{X}$, and in EQP -- this approximation is always exact by definition.

4. MINIMIZATION OF A NON-LINEAR FUNCTION SUBJECT TO NON-LINEAR CONSTRAINTS

Consider the optimization problem NLCP (Non-linear Constrained Problem)

$$\text{minimize} \quad F(X) \quad X \in R^n \qquad (4.1)$$

(NLCP)

$$\text{subject to } c_i(X) \geqq 0, \; i=1,2,\ldots,m \qquad (4.2)$$

where $F(X)$ and $\{c_i(X)\}$ are nonlinear twice - differentiable functions at least for feasible points. At the strong local minimum $\overset{*}{X}$ of NLCP, a set of $\overset{*}{t}$ constraints (4.2) must be - active; let $\{\overset{*}{c}(X)\}$ be that set, then $\overset{*}{c}(\overset{*}{X})=0$. Let $\overset{*}{A}(X)$ be the nx$\overset{*}{t}$ Jacobian matrix of $\overset{*}{c}(X)$. At point $\overset{*}{X}$, the following conditions hold -- (see Secs. 2 and 3 for the appropriate references and restrictions):

i) point $\overset{*}{X}$ is feasible

ii) $\overset{*}{A}$ is a full column rank matrix. It is - only a condition for the uniqueness of the Lagrange multipliers, and a <u>cons---trained qualification</u> for the other optimality conditions.

iii) Lagrange multipliers are such that
$\overset{*}{\mu}_i \gtrless 0$ for $i=1,2,\ldots,m_1$ (equality cons-- traints)
$\overset{*}{\mu}_i = 0$ for $i=m_1+1,\ldots,m$ (inequality constraints) if the i-th constraint is non-active.
$\overset{*}{\mu}_i \geq 0$ for $i=m_1+1,\ldots,m$ (inequality constraints) if the i-th constraint is ac-- tive.

Let C_1, C_2 and I be the same sets notation - used in previous sections. For the set I, -- vector $\overset{*}{\mu}$ is such that

$$\overset{*}{A}\overset{*}{\mu} = \overset{*}{g} \qquad (4.3)$$

iv) Hessian matrix

$$\overset{*}{H} = \overset{*}{Z}^t\overset{*}{G}\overset{*}{Z} \qquad (4.4a)$$

is positive definite; that is

$$Y^t\overset{*}{Z}^t\overset{*}{G}\overset{*}{Z}Y > 0 \qquad (4.4b)$$

where $\overset{*}{G} \equiv G(\overset{*}{X},\overset{*}{\mu})$ is the Lagrange Hessian matrix at point $(\overset{*}{X},\overset{*}{\mu})$. Note that it is a nxn matrix (it only takes the second derivatives of X - in $L(X,\mu)$ (eq. 2.6a), since those of μ are - zero).

$$G(\overset{*}{X},\overset{*}{\mu}) = G(\overset{*}{X}) - \sum_{i\in I-C_2} \overset{*}{\mu}_i G_i(\overset{*}{X}) \qquad (4.4c)$$

where $G(\overset{*}{X})$ is the Hessian matrix of $F(\overset{*}{X})$, -- $G_i(X)$ is the Hessian matrix of $c_i(\overset{*}{X})$ for --- $i\in I-C_2$, and $\overset{*}{Z}$ is a nx(n-$\overset{*}{t}$) matrix, so that

$$\overset{*}{A}(\overset{*}{X})^t\overset{*}{Z} = 0 \qquad (4.4d)$$

where for $\delta = X-\overset{*}{X}-\phi$ (being ϕ the curvature --- along the feasible arc from $\overset{*}{X}$ to X in continuous system $\overset{*}{c}(X)=0$, δ the limiting vector of ϕ and, then, $\overset{*}{A}^t\delta=0$) it gives

$$\delta = \overset{*}{Z}Y \qquad (4.4e)$$

where Y is any (n-$\overset{*}{t}$) vector. Eq. (4.4b) is - equivalent to

$$\delta^t\overset{*}{G}\delta > 0 \qquad (4.4f)$$

that is, matrix $\overset{*}{G}$ must be positive definite at least for δ, being $\delta+\phi$ the step from the local point $\overset{*}{X}$ to any other near, feasible -- point to the set of active constraints. Formally, δ and ϕ must be sufficiently small -- and must satisfy $\overset{*}{A}^t\delta=0$. If $C_2 \neq \emptyset$ then condition (4.4f) must be extended to the small -- enough stepdirection δ such that $\nabla c_i(\overset{*}{X})^t\delta>0$ for $\forall i\in C_2$. The set C_2 (i.e. $i\in C_2$ means that - $c_i(\overset{*}{X})=0$, $\overset{*}{\mu}_i=0$ and $i=m_1+1,\ldots,m$) is termed -- <u>degenerate inequality</u> constraints set.

Conditions (i)-(iii) and $\overset{*}{H}$ being a positive semi-definite matrix are neccessary optimality conditions. Conditions (i)-(iv) are only sufficient conditions.

There are several ways to calculate $\overset{*}{Z}$ /7/, - /2/ and /5/.

The above conditions are similar to condi-- tions for LCP (Sec. 3). The difference is -- that in the linear constrained case we need in (4.4c) only matrix G(X) since $\{G_i(X)\}$ is zero. Since $\delta=\overset{*}{Z}Y$, $Y^t\overset{*}{Z}^t\overset{*}{G}\overset{*}{Z}Y > 0$ means in LCP -- that $\delta^t\overset{*}{G}\delta>0$ (the Hessian matrix $\overset{*}{G}$ of <u>objective function</u> $F(\overset{*}{X})$ is positive definite) for $\delta=X-\overset{*}{X}$ (being X feasible in $\overset{*}{c}(X)=0$); and in - NLCP, $Y^t\overset{*}{Z}^tG(\overset{*}{X},\overset{*}{\mu})\overset{*}{Z}Y > 0$ means that $\delta^tG(\overset{*}{X},\overset{*}{\mu})\delta>0$ (the Hessian matrix $G(\overset{*}{X},\overset{*}{\mu})$ of <u>Lagrange function</u> $L(\overset{*}{X},\overset{*}{\mu})$ is positive definite, indepen-- dently of $G(\overset{*}{X})$) for $\delta=X-\overset{*}{X}-\phi$ (being X feasible in $\overset{*}{c}(X)=0$) with the specifications given in the previous paragraphs.

Matrix $\overset{*}{H}$ (eq. 4.4a) is termed <u>reduced Lagrange Hessian matrix</u>. It is a (n-$\overset{*}{t}$)x(n-$\overset{*}{t}$) matrix and it is easier than $G(\overset{*}{X},\overset{*}{\mu})$ to handle it.

Like the other two cases, associated to problem NLCP, we have the (ENLCP) equality non-linear constrained problem

$$\text{minimize} \quad F(X) \qquad (4.5)$$

(ENLCP)

$$\text{subject to} \quad \hat{c}(X)=0 \qquad (4.6)$$

where $\hat{c}(X)$ is a selection of $\hat{t}$ elements of -- vector $c(X)$. Let $\hat{A}(X)$ be the Jacobian matrix of $\hat{c}(X)$. It is a $n\times\hat{t}$ full column rank -- matrix.

The optimum point $\hat{X}$ of ENLCP must satisfy -- analogous conditions to conditions (i)-(iv) of NLCP, but without any constraint in the - sign of the $\hat{t}$ elements of $\hat{\mu}$. If the $(\hat{t}-m_1)$ - elements of $\hat{\mu}$ corresponding to the active -- inequality constraints of NLCP are non-negative, the optimum $\hat{X}$ is also $\overset{*}{X}$. In other case these constraints are to be dropped from --- ENLCP (may be other constraints are to be added to ENLCP) and the procedure goes to the next iteration: optimization of the new --- ENLCP.

Like in ELCP also in ENLCP we may use the -- set of active constraints to obtain the firt order, second-order and pseudo-second-order estimations of Lagrange multipliers $\hat{\mu}$. In -- ENLCP many algorithms (see in /2/ the appropriate references) obtain, at intermediate - iterations, points $\{X\}$ that are not feasible, then we may also obtain for these points $\{X\}$ estimations μ_L and μ_B of $\hat{\mu}$ including in the set of active constraints: constraints for - which $\{c_i(X)=0\}$, and also $\{c_i(X)<0\}$.

Let X be an approximation of $\hat{X}$ obtained in - the sequence of iterations solving ENLCP. -- Like in Sec. 3, the estimation μ_L is

$$\mu_L = R^{-1}Q_1^t g \qquad (4.7)$$

and, alternatively (although it is not recommended)

$$\mu_L = (\hat{A}^t\hat{A})^{-1}\hat{A}^t g \qquad (4.8)$$

Gill and Murray (/8/, p. 35) suggest to represent the gradient $\hat{g}$ of optimum point $\hat{X}$ of ENLCP in Taylor series around X, so that --- (see Sec. 3)

$$\hat{A}(X+d)\hat{\mu} = \hat{g}(X+d) = g+Gd+0(\|d\|^2) \qquad (4.9)$$

But $\hat{A}$ is not a constant matrix. Then, representing Jacobian matrix $\hat{A}$ in Taylor series - around point X, we have

$$\hat{A}\hat{\mu}+\sum_{i\in I-C_2}\hat{\mu}_i\hat{G}_i d+0(\|d\|^2)=g+Gd+0(\|d\|^2) \qquad (4.10)$$

where $\hat{A}$, G, $\hat{G}_i$ and g are evaluated at point X, being $\hat{G}_i$ the Hessian matrix of $\hat{c}_i(X)$ the residual values $0(\|d\|^2)$ of both sides are -- not necessarily the same. Fussioning both -- residuals and reordering eq. (4.10), we have

$$\hat{A}\hat{\mu} = g+G(X,\hat{\mu})d+0(\|d\|^2) \qquad (4.11)$$

where

$$G(X,\hat{\mu}) = G(X) - \sum_{i\in I-C_2}\hat{\mu}_i\hat{G}_i(X) \qquad (4.12)$$

Eq. (4.11) is similar to eq. (3.11). If functions $\hat{c}_i(X)$ in ENLCP are linear, then $\hat{G}_i(X)\equiv\equiv 0$ and eq. (4.11)$\equiv$eq. (3.11). $G(X,\hat{\mu})$ is the Hessian matrix of Lagrange function

$$L(X,\hat{\mu}) = F(X) - \hat{\mu}^t\hat{c}(X) \qquad (4.13)$$

See also eq. (4.4c). Comparing eqs. (4.11) - and (3.11) we may recall the remark on condition (iv) in the sense that for the non-li-- near case the matrix to study is the <u>Lagrange Hessian, instead of the Hessian of objective function</u>.

Substituting G by $G(X,\hat{\mu})$ in eqs. (3.11-3.28) of the linear case we may calculate the similar estimations $\mu_G^{(1)}$, $\mu_G^{(2)}$, $\mu_B^{(1)}$, $\mu_B^{(2)}$ -- for the non-linear case. However, we cannot directly use eq. (4.11) to estimate $\hat{\mu}$ since the right-hand-side requires to know $\hat{\mu}$, except if we know some estimation of this vector. It makes sense if this estimation is -- the first-order estimation μ_L (eq. 4.7); then the second and pseudo-second order estima--- tions are (similar to the linear case);

$$\mu_G^{(1)} = (\hat{A}^t G(X,\mu_L)^{-1}\hat{A})^{-1}\hat{A}^t G(X,\mu_L)^{-1}g$$

(not recommended) (4.14)

$$\mu_G^{(2)} = R^{-1}Q_1^t(g+G(X,\mu_L)\delta) \quad \text{(not useful)} \qquad (4.15)$$

$$\mu_B^{(1)} = (\hat{A}^t B(X,\mu_L)^{-1}\hat{A})^{-1}\hat{A}^t B(X,\mu_L)^{-1} g$$

(not recommended) (4.16)

$$\mu_B^{(2)} = R^{-1}Q_1^t(g+B(X,\mu_L)\delta) \qquad (4.17a)$$

where

$$\delta = -\hat{Z}(\hat{Z}^t B(X,\mu_L)\hat{Z})^{-1}\hat{Z}^t g \qquad (4.17b)$$

See e.g. in /2 and 3/ several ways to obtain B. Point here the same remarks made at the - end of Sec. 3 for the comparison among the - estimations μ_L, μ_G and μ_B.

The stepdirection δ is calculated, in a similar way to the linear case, by solving problem

$$\min \{g^t\delta+\tfrac{1}{2}\delta^t B(X,\mu_L)\delta \,|\, \hat{A}^t\delta=-\hat{c}(X)\} \qquad (4.18)$$

since the linearization of the active constraints around point X is produced at the given iteration of the given algorithm, /5/. Note that it is possible that using δ (calculated by eq. (4.18)) some constraint (active or not) may be violated (then if it is non-active we have to add it to the set in the next iteration) or some active constraint may become non-active (then we have to drop it from the set in the next iteration).

Betts /1/ suggests an apparent different way to estimate vector μ in NLCP. Ler μ^+ be this estimation. In order to satisfy conditions (i)-(iii), given the point X and, then A(X) ($\equiv$A) and g(X) ($\equiv$g), vector μ^+ must minimize the euclidean residual

$$e(\mu) = ||P\mu+b||^2 = \mu^t P^t P\mu + 2b^t P\mu + b^t b \qquad (4.19)$$

where μ is the m-unknown vector, P is a (n+m)xm matrix and b is a n+m vector such that

$$P = \begin{pmatrix} -A \\ c \end{pmatrix}; \quad b = \begin{pmatrix} g \\ 0 \end{pmatrix} \qquad (4.20a)$$

where A is the nxm Jacobian matrix of all constraints evaluated at X; c is the diagonal matrix

$$c = \begin{pmatrix} c_1 & & & \\ & c_2 & & \\ & & \ddots & \\ & & & c_m \end{pmatrix} \qquad (4.20b)$$

where $c_i \equiv c_i(X)$ is the i-th constraint evaluated at X, g is the gradient g(X) and 0 is the null m-vector. Then, minimizing (4.19) is equivalent to minimize the deviation from conditions (i) and (iii), so that $\mu_i^+ = 0$ ($i=m_1+1,\ldots,m$) for non-active inequality constraints and with the additional condition $\mu_i^+ \geq 0$ ($i=m_1+1,\ldots,m$) for active inequality constraints.

Since $b^t b$ is fixed, problem (4.19) gives

$$\min_{(\mu)} \bar{e} = b^t P\mu+\tfrac{1}{2}\mu^t(P^t P)\mu \qquad (4.21)$$

with the above conditions on μ. If $P^t P$ is non-singular, it is positive definite. If there are only equality constraints, μ is unrestricted and the vector μ in (4.21) is the vector μ^+ that satisfies the following system of linear eqs.

$$P^t P\mu = -P^t b \equiv (A^t A+c^2)\mu = A^t g \qquad (4.22)$$

where $A^t A$ is a mxm symmetric matrix; then

$$\mu^+ = (A^t A)^{-1} A^t g + \underline{c}^2 A^t g \qquad (4.23)$$

where $\underline{c}$ is a m-diagonal matrix whose i-th element is $1/c_i$. If X is feasible, c=0 and the estimation μ^+(4.23) is the first-order estimation μ_L (4.8); it is unstable. However it may be possible to calculate a QR-factorization of $(A^t A+c^2)$.

5. THE USING OF THE LAGRANGE MULTIPLIERS IN AN ALGORITHM FOR NON-LINEAR CONSTRAINED PROGRAMMING

We describe elsewhere /2/ an algorithm for minimizing function (4.1) subject to constraints (4.2) with bounded variables. It uses the estimation of Lagrange multipliers as follows: At a given iteration we have matrix $B\equiv B(X,\mu_L)$ that (being calculated at the end of previous iteration) approximates matrix $G(X,\mu)$. B is positive definite. We obtain the stepdirection δ by solving the quadratic programming problem (4.18) and as a by-product we obtain μ_B (4.17); in this algorithm we also consider in problem (4.18), the non-active constraints and the bounds on the variables. The strategy that we use (see below) is at an intermediate level between pure active and non-active set strategies; see in /16/ a full discussion of these alternatives. Also we obtain the steplength α and the new point $X \leftarrow X+\alpha\delta$. We obtain the set of active constraints (strictly satisfied and violated) $\{\hat{c}_i(X)\}$, gradient g(X) and Jacobian matrix $\hat{A}(X)$ of $\hat{c}(X)$ all evaluated at the new point X. Finally we use formula (4.7) to obtain the <u>first-order estimation</u> μ_L of Lagrange multipliers. With g(X), $\hat{A}(X)$ and μ_L we correct $B(X,\mu)$ to obtain the new matrix B that will be used by the next iteration.

For solving the quadratic problem (4.18) /5/, we use the strategy that classifies the variables in basic, superbasic and non-basic. The last type of variables are those whose value is temporarily fixed at one of their bounds; the slack variables that are non-basic correspond to the inequality constraints that temporarily are active (then, the value of these variables is zero). The iterations of the quadratic algorithm (termed inner -- iterations) obtain the stepdirection of the superbasic variables by solving an unconstrained reduced quadratic problem. If in a - given inner iteration there are no more su-- perbasic variables or the stepdirection is - zero, the quadratic algorithm tests if some non-basic variable may be converted in superbasic, given priority to the non-slack variables. This test is performed by using the Lagrange multipliers (or also termed reduced - costs) of the non-basic variables, whose formula uses the pseudo-second-order estimation μ_B (4.17) that is obtained as a by-product - while solving problem (4.18). The optimum solution δ on problem (4.18) is obtained when there are no non-basic variables whose status must be changed, and there are no superbasic variables or their step direction is zero. Then in some inner iterations of the quadratic algorithm we use the estimation -- μ_B, but in the major iterations of the non-- linear constrained algorithm we use the ---- first-order estimation μ_L (4.7) for point -- $X+\alpha\delta$ and gradient $g(X+\alpha\delta)$.

Estimation μ_L is no worse then μ_B (see the - final remarks on Sec. 3) and, although the - given algorithm alternatively allows to use μ_B as it is recommended by Powell /19 and 20/ we have better computational results when we use μ_L. Also Murray and Wright /16/ obtain - better results with μ_L. We use the implementation described in /4/ to obtain the QR-factorization of $\hat{A}$ and to calculate the vector μ_L (4.7).

The given algorithm uses also μ_L to analyze if X is close to $\overset{*}{X}$; that is, we use μ_L as a stopping criterium on the sequence of major iterations to solve NLCP. The first convergence criterium that the algorithm uses is as follows:

i) Check the feasibility of X.

ii) Check the sign of the elements of μ_L -- that correspond to the active inequality constraints. In fact if they are -- greater than ε_1 (usually, $\varepsilon_1=10^{-4}$), this condition is considered satisfied.

iii) If the following inequality holds

$$\|g(X)-A(X)\mu_L\|^2 < \varepsilon_2 \qquad (5.1)$$

where ε_2 is a given tolerance (usually, ε_2--- $=10^{-4}$) we consider that X is quasi-optimum (see conditions (i)-(iv)). See in /4/ the -- method to calculate the residual (5.1) with the maximum accuracy that it is possible in a computer.

If criteria (i)-(iii) are satisfied but at - the given major iteration k, $\mu_i^{(k)} \le \varepsilon_1$ for some inequality constraint (4.2), this constraint is temporarily deleted from the set I of active constraints. Other different approach - in dealing with the negative Lagrange multipliers is suggested by Gill and Murray /8/. In this approach, the non-negativity requirement of the Lagrange multipliers of the inequality constraints has top priority over -- the optimality condition(iii) (eq. 4.3). --- Then the first-order estimation given by formula (4.7) to obtain μ may be substituted by the solution in μ of the quadratic problem

$$\min \|R\mu_L - Q_I^t g\|^2 \qquad (5.2a)$$

$$\text{subject to } \mu_{L_i} \ge 0 \qquad (5.2b)$$

where $\{i \in I$ and $i=m_1+1,\ldots,m\}$. With this al-- ternate procedure, it never happens the case for which (X,μ) solves problem ENLCP (4.5) - (4.6) without satisfying restriction (5.2b). We have experimented with this approach, but more computational validation is required.

6. THE USING OF THE LAGRANGE MULTIPLIERS IN AN ALGORITHM FOR LINEARLY CONSTRAINED NON-LINEAR PROGRAMMING

We describe elsewhere (/5/, sec. 16) an algorithm for minimizing function (3.1) subject to lower and upper bounded nonlinear constraints with bounded variables ($\ell \le X \le u$).

In the given algorithm, the status at each major iteration is as follows. Point X is -- feasible, $\bar{A}^t=(\hat{A},\hat{N})$ is the active constraints matrix, $\hat{N}$ is the submatrix of $\bar{A}^t$ related to the non-basic set V of variables j, such --- that $j \in V$ if $X_j=\ell_j \vee u_j$, $\hat{A}$ is the complement -- matrix to $\hat{N}$ in A^t, g is the gradient of the objective function F(X) and H is the reduced Hessian approximated at point X.

At each major iteration, a quadratic problem (QP, see sec. 2) that approximates problem - ELCP (see sec. 3) is solved, although its objective function is not completely minimized in the subspace defined by set J-V, where J is the set of all variables.

At each QP iteration, g and H are approximated and, once a quasi-optimal solution to QP is obtained, the first-order estimates

$$\mu_L : \min \|\hat{A}^t \mu - g\|^2 \tag{6.1}$$

$$\lambda = g_N - \hat{N}^t \mu_L \tag{6.2}$$

of the Lagrange multipliers are obtained, -- such that $\hat{g}$ is the gradient related to set - J-V, and λ is the Lagrange multiplier estimates vector of set V. Eventually, λ is only - obtained for a subset of V, such that the -- zigzagging phenomena is reduced and the multiple-partial pricing is used. See in /4 and 6/ the method to be used for obtaininf the - QR factorization of A^t and its updatings when a constraint or variable is added or deleted. Test (3.29) is used at some major iterations.

A partial (exact) normalization of λ (say λ') and μ are used for selecting the variable -- from set V in the first case and the inequality constraint in the second case, such -- that being dropped (that is, desactivated), it is expected to produce the best descent - feasible stepdirection, so that a new matrix $\bar{A}^t = (\hat{A}, \hat{N})$ is obtained. An special strategy -- is used to avoid the stepdirection being infeasible.

The QP problem is solved by minimizing the - reduced quadratic problem $\{h^t \delta_r + 1/2 \delta_r^t H \delta_r \mid$ -- $\ell - X \le \delta \le u - X\}$, where h is the reduced gradient. Vectors h and λ' are updated by using the Lagrange multipliers of the active constraints in the associated linear program to the QP problem. Note that $\delta = \bar{Z}\delta_r$ where δ_r is the reduced stepdirection and $\bar{Z}$ is such that $\bar{A}^t \bar{Z} = 0$. The given algorithm uses an special strategy for checking the sign of λ and μ_L when they have not any clearly favorable element for - desactivating the non-basic variables or active constraints, but they have zero or near zero elements. The aim of this strategy consists in obtaining more information about -- the trend in the sign of these elements by - means of an small perturbation in the active bound of the variables or in the right-hand side of the constraints. This strategy is very fast since only approximates the gradient g of the new trial point and the rest of the required data are based on the QR factorization to be obtained for calculating μ_L.

7. CONCLUSION

The first-order μ_L, second order μ_G and pseudo-second-order μ_B estimations of the Lagrangian multipliers are discussed. The estima-- tion μ_L may be obtained for points X and -- $X+\alpha\delta$ with gradients $g \equiv g(X)$ and $g(X+\alpha\delta)$; the other two estimations are only obtained for point $X+\delta$ and the gradients are considered - to be $g+B\delta$ for μ_B and $g+G\delta$ for μ_G.

Estimation μ_G is always better than estima-- tions μ_L for point X and μ_B. Estimation μ_B - is better than estimation μ_L for point X, except if B is a poor approximation of the Hessian G. Estimation μ_L for point $X+\alpha\delta$ is, of-course, better than μ_L for point X.

In the quadratic case, estimation μ_G obtains exactly the Lagrange multipliers; μ_L for -- point X is only an approximation.

In the non-linear function case, estimation μ_L for point $X+\alpha\delta$ is better than estimation μ_G; both converge to the true Lagrangian multipliers when the solution X is close to the optimum $\overset{*}{X}$. Estimation μ_L for point $X+\alpha\delta$ is no worse than estimation μ_B; but if B is a - poor approximation of G then both estima--- tions are poor. Since G is not generally --- available and μ_L for point $X+\alpha\delta$ is better -- than μ_L for point X, the selection of the estimation to be used remains between μ_L for - point $X+\alpha\delta$ and μ_B. Both types of estimations have two alternate computational methods; -- formulae (4.7) and (4.17) are, respectively, more stable than formulae (4.8) and (4.16).

In the non-linear constrained algorithm that we describe elsewhere, we have obtained our best results by using estimation μ_B in the - inner iterations of the quadratic program -- solved at each major iteration (note that matrix B is the Hessian matrix of this particular quadratic problem) and by using estimation μ_L for point $X+\alpha\delta$ at each major iteration. In the linearly constrained algorithm, we use μ_L for point $X+\alpha\delta$ at each major iteration and an special strategy to avoid a non feasible stepdirection.

8. ACKNOWLEGEMENT

I wish to thank the helpful comments and --- suggestions of Walter Murray, Phil Gill and

Margaret Wright from the Systems Optimization Laboratory (Stanford University) that strongly improved the formulae and algorithms described in this paper.

9. REFERENCES

/1/ BETTS, J.T.: "An accelerated Multiplier Method for Non-Linear Programming", J. of Optimization Theory and Applications 21 (1977) 137-174.

/2/ ESCUDERO, L.F.: "A Projected Lagrangian Method for Nonlinear Programming" IBM Scientific Center, report G320-3401, Palo Alto, California, 1980a.

/3/ ESCUDERO, L.F.: "On Hessian Matrices in Unconstrained Optimization", IBM Scientific Center, report G320-3407, Palo Alto, California, 1980b.

/4/ ESCUDERO, L.F.: "Implementation of QR-factorization on Solving Overdetermined Systems of Linear Equations", Questiió 4 (1980c) 89-94.

/5/ ESCUDERO, L.F.: "An Algorithm for large-scale quadratic programming problems and its extensions to the linearly constrained case", IBM Scientific Center, report SCR-01.81, Madrid, 1981.

/6/ ESCUDERO, L.F.: "QR factorization and its updatings", Trabajos de Estadística e I-O (1982, in printing).

/7/ GILL, P.E. and MURRAY, W.: "Numerical Methods for Constrained Optimization (Academic Press, London, 1974).

/8/ GILL, P.E. and MURRAY, W.: "The Computation of Lagrange Multipliers Estimates for Constrained Minimization", National Physical Laboratory, NAC report 77, Teddington (UK), 1977a. See also Mathematical Programming 17 (1979) 32-60.

/9/ GILL, P.E. and MURRAY, W.: "Linearly Constrained Problems Including Linear and Quadratic Programming", in: D.A.H. Jacobs (ed.), The State of the Art in Numerical Analysis (Academic Press, London, 1977b) 313-363.

/10/ GILL, P.E. and MURRAY, W.: "Numerically Stable Methods for Quadratic Programming", Mathematical Programming 14 (1978) 349-372.

/11/ GILL, P.E., MURRAY, W. and WRIGHT, M.: "Practical Optimization" (Academic Press, London, 1981).

/12/ GOLUB, G.H.: "Matrix Decomposition and Statistical Calculations", in: Milton and Nedier (eds.), Statistical Computations (Academic Press, London, 1969).

/13/ HEATH, M.T.: "Numerical Algorithms for Nonlinearly Constrained Optimization", Doctoral Thesis, Stanford University, California, 1978.

/14/ MANGASARIAN, O.L.: "Non-Linear Programming" (McGraw-Hill, N.Y., 1969).

/15/ MURRAY, W. and WRIGHT, M.: "Projected Lagrangian Methods Based on the Trajectories of Penalty and Barrier Functions", Systems Optimization Laboratory, report SOL-78-00, Stanford University, California, 1978.

/16/ MURRAY, W. and WRIGHT, M.: "Computation of the Search Direction in Constrained Optimization Algorithms", Systems Optimization Laboratory, report SOL-80-2, Stanford University, California, 1980.

/17/ PETERS, G. and WILKINSON, J.H.: "The Least-Square Problems and Pseudo-Inverses", Computer Journal 13 (1970) 309-316.

/18/ POWELL, M.J.D.: "Introduction to Constrained Optimization", in: P.E. Gill and W. Murray (eds.), Numerical Methods for Constrained Optimization (Academic Press, London, 1974) 1-28.

/19/ POWELL, M.J.D.: "A Fast Algorithm for Nonlinear Constrained Optimization Calculation", Conference on Numerical Analysis, Dundee, 1977.

/20/ POWELL, M.J.D.: "Algorithms for Nonlinear Constraints that use Lagrangian Functions", Mathematical Programming 14 (1978) 224-248.