ZERO OR NEAR-TO-ZERO LAGRANGE MULTIPLERS IN LINEARLY CONSTRAINED NONLINEAR PROGRAMMING

L. F. ESCUDERO

We discuss in this work the using of Lagrange multiplier estimates in linearly constrained nonlinear programming algorithms and the implication of zero or near-to-zero Lagrange multipliers. Some methods for estimating the tendency of the multipliers are proposed in the context of a given algorithm.

1. INTRODUCTION

The linearly constrained nonlinear programming (LCNP) problem is

\[ \text{minimize } F(X) \quad \text{where} \quad X \in \mathbb{R}^n \]

\[ \Delta(X|B \geq AX, b, U \geq X \geq 1) \quad (1.1) \]

where \( A \) is an \( m \times n \) matrix, \( m \geq n \), and \( F(X) \) is a general nonlinear twice continuously differentiable function, at least, for feasible points such that for all \( X \in \mathbb{R}^n \) the level sets

\[ L(X) \Delta(X|F, X \in \mathbb{R}^n) \quad (1.2) \]

are bounded. Let \( F \) be the set of constraints, \( E \) be the set of equality constraints (such that \( i \in E \) if \( B_1 \subseteq B_i \)), and \( J \) be the set of variables. Let \( \hat{A} \) be the \( \hat{n} \times \hat{n} \) matrix of active constraints at a local optimal point, say \( \hat{X} \) and \( \hat{b} \) the \( \hat{n} \) vector of right-hand-side corresponding to \( \hat{A} \) (i.e., \( \hat{A} \hat{X} = \hat{b} \)), such that \( \hat{i} = |\hat{A}| \) - where \( \hat{A} \) is the set of active constraints and \( \hat{b}_i = B_j \hat{b}_j \) for \( i \in \hat{W} \). Let \( \hat{V} \) be the set of active variables at \( \hat{X} \), such that \( j \in \hat{V} \) if \( \hat{x}_j = U_j \hat{v}_j \) and \( r = |\hat{V}| \). Let \( \hat{I} \) be the \( \hat{n} \times \hat{n} \) identity matrix \( I \) from where the row related to variable \( j \in \hat{V} \) has been deleted.

We shall define vector \( g(X) \) as the vector whose \( j \)-th element is \( \delta F(X)/\delta X \) and the Hessian matrix \( G(X) \) as the symmetric matrix whose \( (i,j) \)-th element is \( \delta^2 F(X)/\delta X_i \delta X_j \).

The algorithm /2/ concerned with this paper - is assumed to generate a sequence of feasible estimates \( \{X(k)\} \) of \( \hat{X} \) (weak local minimum) by obtaining a stepdirection \( d(k) \) and a step-length \( a(k) \) such that \( X(k) = X(k-1) + a(k) d(k) \) - and \( \lim X(k) = \hat{X} \) for \( k \rightarrow \infty \), where \( F(X) \leq F(\hat{X}) \) for \( \hat{X}, \hat{X} \in \mathbb{R}^n \) and \( \| \hat{X} - \hat{X} \| \leq 0 \) being a small enough to define the neighbourhood of \( \hat{X} \).

Note that \( \hat{A} \hat{X} = \hat{b}, \hat{B} \geq \hat{Y} \geq 0 \iff \hat{B} \geq \hat{A} \hat{X} \geq \hat{b} \); then \( i \in \hat{W} \) for \( \hat{v}_i = 0, \hat{b}_i \hat{b}_i \). The X-variables are termed structural; the Y-variables are termed slack.

Because the constraints are a linear system, the properties of linear subspaces make it possible to state a simple characterization - of all feasible moves from a feasible point. Consider the move between two feasible points \( \hat{X} \) and \( \check{X} \) along the manifold defined by the sets \( \hat{V} \) and \( \check{V} \); by linearity \( \hat{A}(\hat{X} - \check{X}) = 0 \) and \( \check{V} = \hat{V} \) so that \( \hat{X} \) and \( \check{X} \) are moveable \( \forall \check{v} \in \check{V} \) and, then,

\[ \hat{A} \hat{d} = 0, \hat{b} \hat{d} = 0 \quad (1.4) \]

where \( d \) is the stepdirection from \( \hat{X} \) to \( \check{X} \) such that \( \check{X} = \hat{X} + \hat{d} \). Any vector \( d \) for which (1.4) ---

- L.F. Escudero, Centro de Investigación UAM-IRN. P. Castellana, 4 - Madrid-1
- Article rebut at Juny de 1982.
holds is a feasible stepposition from \( \hat{x} \) with respect to the above manifold; it is also termed active stepposition; it will be descent if \( F(\hat{x}) < F(x) \). Steplength \( a \) is required to be \( 0 < a \leq a_m \), where \( a_m \) defines the maximum allowed steplength such that \( \hat{x} \) is still feasible and \( F(\hat{x}) \) is descent.

Let us define a non-active stepposition \( d \) as the feasible stepposition such that some constraint or bound is removed from the sets \( \hat{w} \) and \( \hat{v} \), respectively; a feasible stepposition \( d \) is non-active if \( \exists i \in M - \hat{w} \) for which \( A_i d > 0 \) if \( \hat{y}_i = 0 \), \( A_i d < 0 \) if \( \hat{y}_i = L - L_i \), or \( \exists j \in \hat{v} \) for which \( d_j > 0 \) if \( \hat{x}_j = L_j \), \( d_j < 0 \) if \( \hat{x}_j = U_j \).

The paper is organized as follows. Sec 2 describes the optimality conditions for \( \hat{x} \) being a weak local optimum in (1.1)-(1.2), so that the degenerate sets of active constraints -- and bounds are defined. Sec 3 motivates the analysis of these sets, such that it states the reasons for analyzing the zero or near-to-zero Lagrange multipliers estimates. Sec. 4, briefly, describes the formulas that are being used in a given algorithm to obtain Lagrange multipliers estimates. Sec. 5 outlines the deactivating process in that algorithm. Sec. 6, finally, is devoted to some procedures that are proposed to get some insight on the tendency of zero or near-to-zero Lagrange multipliers estimates for optimal or 'quasi-optimal' solutions in the manifold defined by sets \( \hat{w} \) and \( \hat{v} \).

2. OPTIMALITY CONDITIONS

The necessary optimality conditions for \( \hat{x} \) -- being a weak local minimum are as follows. -- See equivalent conditions in /12, /9, /11/ and /8/ among others. As we state them below, they help to analyze the risk of using Lagrange multipliers estimates of nonbasic (structural and slack) variables, mainly when they are zero or near-to-zero, at a given optimal or 'quasi-optimal' point \( \hat{x} \) in the manifold defined by \( \hat{w} \) and \( \hat{v} \).

(1) \( \hat{x} \in F \) (feasible)
(2) The reduced gradient vector, say \( \hat{h} \) of \( F(X) \) vanishes, such that

\[ \hat{h} \leq \hat{a} \leq \hat{g} = 0 \]  

(2.1)

where \( \hat{a} \) is a \((n-t)\times r\) full column rank matrix, whose columns form the null basis of the range of matrix \( \hat{A}^T, \hat{A}^T \) and, then, \( \hat{A}^T \hat{X} = 0, \hat{A}^T \hat{y} = 0 \)  

(2.2)

Based on (1.4) and (2.2), we may note that any linear combination of the columns of \( \hat{a} \) give an active stepposition \( d \),

\[ d = \hat{a} d_s \]  

(2.3)

where \( d_s \) is a \((n-t-r)\)-vector termed reduced stepposition (or superbasic stepposition), such that a vector \( d \) that cannot be expressed by (2.3) is not an active stepposition in the manifold defined by \( \hat{w} \) an \( \hat{v} \).

Any point at which the reduced gradient \( h \) vanishes (2.1) is termed constrained stationary point. To see that, let us examine the Taylor-series expansion of \( F(X) \) about \( \hat{x} \) along an active stepposition \( d \) in the manifold defined by \( \hat{w} \) an \( \hat{v} \):

\[ F(\hat{x} + a d) = F(\hat{x}) + a \hat{a} (\hat{a}^T \hat{a})^{-1} \hat{a} + o(|a|) \]  

(2.4)

where \( \hat{a} \) satisfies \( 0 < \hat{a} < 1 \). Suppose that \( \hat{x} \) is a local minimum, but \( a \hat{a} (\hat{a}^T \hat{a})^{-1} < 0 \); then, there must exist \( \hat{a} > 0 \) such that \( \hat{a} (\hat{a}^T \hat{a})^{-1} + \) \text{--------} \( 1/2 a^2 \hat{a} (\hat{a}^T \hat{a})^{-1} \hat{a} \leq 0 \) for all \( 0 < \hat{a} \) and, then, \( F(\hat{x} + a \hat{a} (\hat{a}^T \hat{a})^{-1} \hat{a}) < F(\hat{x}) \). Similarly, it can be shown that \( \hat{x} \) is non-optimal if \( a \hat{a} (\hat{a}^T \hat{a})^{-1} > 0 \). Therefore, \( a \hat{a} (\hat{a}^T \hat{a})^{-1} \) must be zero in order for \( \hat{x} \) to be a minimum. Thus, a necessary condition -- for \( \hat{x} \) to be a local minimum in the manifold defined by \( \hat{w} \) and \( \hat{v} \) is that \( a \hat{a} (\hat{a}^T \hat{a})^{-1} \) must vanish for every \( d_s \), which implies (2.1) must hold.

The result (2.1) implies that \( \hat{a} \) must be a \( \hat{a} \) near combination of the rows of \( \hat{x} \) and \( \hat{a} \),

\[ \hat{a} = \hat{A}^T B + \hat{f} \hat{g} \]  

(2.5)

for some vectors \( \hat{b} \) and \( \hat{f} \); they are termed -- the Lagrange multipliers of the active constraints and bounds, respectively. Note that (2.5) is equivalent to (2.1) since any n-vector can be expressed as a linear combination.
of the columns of matrices \((\hat{A}^t, \hat{B}^t)\) and \(\hat{\lambda}\) -
and hence
\[
\hat{g} = \hat{A}^t \hat{u} + \hat{B}^t \hat{\lambda} + \hat{b}_g
\]
for some vectors \(\hat{u}\), \(\hat{\lambda}\) and \(\hat{b}_g\). Premultiplying \(\hat{g}\) by \(\hat{B}^t\) and using (2.1) and (2.2), it results
\[
0 = \hat{B}^t \hat{g} = \hat{B}^t \hat{A}^t \hat{u} + \hat{B}^t \hat{B}^t \hat{\lambda} + \hat{B}^t \hat{b}_g = \hat{B}^t \hat{B}^t \hat{\lambda} + \hat{B}^t \hat{b}_g \quad (2.7)
\]
Since \(\hat{B}\) is a full column rank matrix, \(\hat{B}^t \hat{B}\) is nonsingular and, then, (2.7) only holds for \(\hat{b}_g = 0\) such that, by using (2.2), it finally -
results that (2.5) holds. By simple substitution, and using (2.2), we may see that (2.5) also implies (2.1).

(iii) Uniqueness of the Lagrange multipliers.

Let us partition matrix \(\hat{A}\) and gradient \(\hat{g}\) -
such that \(\hat{A} = (\hat{B}^t, \hat{N})\) and \(\hat{g} = (\hat{B}^t \hat{G}^t \hat{N})^t\), where \(\hat{N}\)
is a \(\hat{r} \times \hat{r}\) matrix defined by sets \(\hat{w}\) and \(\hat{v}\) and
\(\hat{G}\) is the gradient of set \(\hat{V}\). Based in (2.5), \(\hat{\lambda}\) can be written,
\[
\hat{\lambda} = \hat{G}^t \hat{u} - \hat{N}^t \hat{v} \quad (2.8)
\]
such that \(\hat{u}\) satisfies the linear system
\[
\hat{g} = \hat{B}^t \hat{u} \quad (2.9)
\]
Point \(\hat{G}\) does not require \(A\) to be a full row rank matrix, but the uniqueness of vectors \(\hat{u}\)
and \(\hat{\lambda}\) require \(\hat{B}\) to have that property.

Assume that \(\hat{\lambda}_1\) and \(\hat{\lambda}_2\) satisfy (2.9). Then,
\[
\hat{g} = \hat{B}^t \hat{u}_1 = \hat{B}^t \hat{u}_2 = \hat{B}^t \hat{u}_3 = (\hat{B}^t)^t (\hat{u}_1 - \hat{u}_2) = 0 \quad (2.10)
\]
If the rows of \(\hat{B}^t\) are linearly independent, \(\hat{u}_1 = \hat{u}_2 = \hat{u}_3 = \hat{0}\).

In any case, computational stability in the algorithms that obtain the sequence \((X^{(k)})\) requires \((A_1)\) to be linearly independent for
\(i \in M\).

There are several ways to characterize matrix
\[
\begin{align*}
 & \begin{bmatrix} \alpha \beta \\
 & \beta \\
 & \delta \\
 & \\
 & \gamma \\
 & \end{bmatrix}
\end{align*}
\]
and \(\beta\) are expected, the most attractive ways to obtain it are based ---
(although it is not explicitly calculated) -
on variable-reduction and QR-factorization -
of matrix \((B^t)\).

(iv) The sign of Lagrange multipliers ---

must be as follows:
\[
\hat{u}_i^s = 0 \quad \text{for } i \in E \quad \text{(equality constraint)},
\]
\[
\hat{u}_i = 0 \quad \text{for } i \notin W \quad \text{(non-active inequality cons-}
\]
\[
\hat{u}_i > 0 \quad \text{for } i \in M-E \cap \hat{W} \quad \text{such that } \hat{y}_i = 0 \quad \text{(active}
\]
\[
\hat{u}_i > 0 \quad \text{for } i \in M-E \cap \hat{W} \quad \text{such that } \hat{y}_i = \hat{x}_i - \hat{b}_i \quad \text{(active}
\]
\[
\hat{u}_j = 0 \quad \text{for } j \in \hat{V} \quad \text{(non-active variable)},
\]
\[
\hat{y}_j \leq 0 \quad \text{for } j \in \hat{V} \quad \text{such that } \hat{y}_j = 1 \quad \text{(active vari-
\]
\[
\hat{y}_j \leq 0 \quad \text{for } j \in \hat{V} \quad \text{such that } \hat{y}_j = \hat{u}_j \quad \text{(active vari-
\]

The set \(D_1 \cup D_2\), where \(D_1 \Delta (i \in M-E \cap \hat{W}\) for -----
\(\hat{u}_i = 0\) and \(D_2 \Delta (j \in \hat{V}\) for \(\hat{x}_j = 0\) is termed --
degenerate set of active constraints and -----

The reason for condition (iv) is as fo-----
llows: Since a non-active stepdirection is ---
also feasible, point \(\hat{z}\) will not be a local mi
nimum if \(P(\hat{X}) < P(\hat{X})\) such that \(\hat{X} = \hat{Z} + \alpha \hat{d}\) and \(d\) is
a non-active stepdirection; to avoid this ---
possibility we must add a condition that ensu
res \(\alpha > 0\) for every non-active stepdirection and
and, then, \(P(\hat{X}) < P(\hat{X})\) where \(P(\hat{X})\) can be ----
written
\[
P(\hat{X} + \alpha \hat{d}) = P(\hat{X}) + \alpha \hat{d}^t \hat{A} + 0(|\hat{d}|) \quad (2.11)
\]
such that for \(\alpha \hat{d} < 0\) there is always a small scalar \(\alpha > 0\) for which
\(P(\hat{X} + \alpha \hat{d}) < P(\hat{X})\).

Note that the optimality of \(\hat{X}\) requires that
\(2.5\) holds; then, it results,
\[
\hat{g}^t \hat{d} = \hat{w}^t \hat{A}d + \hat{\lambda}^t \hat{d} \quad (2.12)
\]
Since $A_i^d=0$ for $i \in E$ (see that $\hat{Y}_i$ is not restricted in sign), $\hat{y}^d$ can be written,

$$\hat{y}^d = \sum_{i \in M \cap E} A_i^d + \sum_{j \in V} A_j^d \geq 0 \quad (2.13)$$

Condition (2.13 holds if $u_i^d \geq 0$ for $\hat{y}_i = 0$, $u_i^d \leq 0$ for $\hat{y}_i = \hat{y}_j$ for $\hat{y}_j = \hat{y}_j$, and $\hat{y}_j = 0$ for $\hat{y}_j = u_j^d$ since, otherwise, it is always possible to find a stepdirection $d$ such that $A_i^d = 0$, $\hat{x}^d = 0$ if $\hat{y}_i = 0$ for $i \in \hat{Y}$, $i \notin E$, or $A_i^d > 0$ if $\hat{y}_i = \hat{y}_j = \hat{y}_j$ for $i \in \hat{Y}$, $i \notin E$. For active stepdirections $d$ (where $A_i^d = 0$, $\hat{x}^d = 0$), but also for non-active stepdirections $d$ where the operators vector $p$ in $\hat{d} \in \hat{d}$, $\hat{d} \in \hat{d}$ includes the feasible inequality in the constraints and bounds whose Lagrange multipliers $\hat{u}_i^d$ and $\hat{y}_i^d$ are zero.

Conditions (i)-(ii) and (iv)-(v) are necessary conditions for local optimality; if $\hat{H}$ is regular (its rows are linearly independent) then the above non-active stepdirection could be very easily found by using its pseudo-inverse matrix; the sign of the Lagrange multipliers may be also proved by using the Farkas'lemma by without requiring the regularity assumption on matrix $\hat{A}$ -- (see /8/).

(v) Positive semi-definiteness of the $\hat{H}$ matrix.

The reduced Hessian matrix $\hat{H}$ must be positive semi-definite, where

$$\hat{X} = \frac{\hat{y}^d}{2} \hat{G} \hat{Z} \quad (2.14)$$

To see that, let us examine the Taylor-series expansion of $F(Y)$ about $X$ along an active stepdirection $d$ such that, by using (2.1) and (2.3), it can be written

$$F(X+\delta d) = F(X) + 1/2 \delta^T \hat{H} \delta + O(\delta^2)$$

(2.15)

If $G$ is indefinite, by continuity $\hat{G}$ will be indefinite for $\hat{y}^d > 0$ being small enough, such that by definition $3d_i$ for which $\hat{h}_i^d < 0$ and, then, $\hat{X}$ is not a local minimum.

Note that the above condition is equivalent to require that matrix $\hat{G}$ be positive semi-definite but only for the active stepdirections in the manifold defined by $\hat{Y}$ and $\hat{V}$.

If the degenerate set $D_1 \cup D_2$ is not empty -- then the positive semi-definiteness property of $\hat{X}$ must be extended to the non-active stepdirection $d$ for which it holds: $A_i^d > 0$ for $\hat{y}_i = \hat{y}_j$, or $A_i^d < 0$ for $\hat{y}_j = \hat{y}_j$, and $A_i^d < 0$ for $\hat{y}_j = \hat{y}_j$ since, otherwise, it is always possible to find a stepdirection $d$ such that $A_i^d = 0$, $\hat{x}^d = 0$ if $\hat{y}_i = 0$ for $i \in \hat{Y}$, $i \notin E$, or $A_i^d > 0$ if $\hat{y}_i = \hat{y}_j$ for $i \in \hat{Y}$, $i \notin E$.

Note that $\hat{y}^d$ given is (2.12) vanishes for active stepdirections $d$ (where $A_i^d = 0$, $\hat{x}^d = 0$), but also for non-active stepdirections $d$ where the operators vector $p$ in $\hat{d} \in \hat{d}$, $\hat{d} \in \hat{d}$ includes the feasible inequality in the constraints and bounds whose Lagrange multipliers $\hat{u}_i^d$ and $\hat{y}_i^d$ are zero.

3. MOTIVATION FOR ZERO OR NEAR-TO-ZERO LAGRANGE MULTIPLIERS ANALYSIS.

Point $X$ is a local minimum in the manifold defined by $\hat{Y}$ and $\hat{V}$ if conditions (i)-(ii) are satisfied and matrix $\hat{H}$ (2.14) is positive definite for all $d$. To test if $X$ is also the solution of problem (1.1)-(1.2), it is required to analyze the sign of the active constraints and bounds Lagrange multipliers, such that if $\hat{u}_i^d$ for $v_i \in E \cap \hat{Y}$ (or $\hat{V}_j$ for $v_j \in \hat{V}$) have not the appropriate signs then constraint $i$ (or bound $j$) must be deactivated (that is, a non-active stepdirection must be obtained). Note that e.g. if $\hat{u}_i^d < 0$ for $i \in E \cap \hat{Y}$ and $\hat{V}_j > 0$, the feasible stepdirection $d$ that deactivates constraint $i$ (and, then, $A_i^d > 0$ such that constraint $i$ is deleted from $\hat{H}$) is descent (see (2.11) and (2.13)); viceversa, a descent stepdirection is also non-active (and, then, feasible). On the other hand, e.g. if $\hat{u}_i^d > 0$ for $i \in E \cap \hat{Y}$ and $\hat{V}_j > 0$, any non-active stepdirec-
tion is non-descent; vice versa, a descent -- stepdirection is non-feasible (and, then, $-A_1d<0$).

Note that the Lagrange multipliers take the first-order rate of change in the objective function (1.1) due to a change in the right-hand-side of the related constraint or bound; see (2.13). A caution has to be made since -- the magnitude of the Lagrange multipliers is not invariant to scaling changes.

We may see that the sign is more important than the magnitude of the Lagrange multipliers. Note that if they are not obtained with exact arithmetic, the 'computed' value of e.g. $\hat{y}_i^+$ for $i \in M-E \cap \hat{W}$ and $\hat{y}_i^+ = 0$ may be 'slightly' negative when the 'exact' value is positive and, then, once the constraint is chosen to be deactivated, the related non-active stepdirection is non-descent; if the 'computed' value is positive when the exact value is negative, then a premature termination of the algorithm may occur without reaching the optimum $\hat{x}$. Any computer algorithm works, by its own nature, with finite precision and the results are subject to unstabilities due to cancellation and rounding errors in intermediate operations -/11/.

An additional difficulty arises in the presence of zero computed value of some Lagrange multiplier since, in that case, there is more uncertainty on the sign of its exact value. -- Recall that if it is zero, the positive definiteness property of matrix $\hat{H}$ is not enough to guarantee (together with the other conditions) that $\hat{x}$ is a local minimum.

Some algorithms deactivate constraints and bounds even before the local optimum is reached such that, once a 'quasi-optimal' solution is obtained, estimates of the Lagrange multipliers are calculated and, based on them, the desactivating process is executed; see /13/ among others. -- These estimations introduce a new uncertainty on the sign of the 'exact' Lagrange multipliers.

4. LAGRANGE MULTIPLIERS ESTIMATES.

See in /10/, /3/, several methods to obtain Lagrange multipliers estimates. We use two types of formulas; the so-called first-order estimates, see in /3/ the motivation for not using second-order estimates.

Following a traditional approach /13/, let -- the active constraints matrix, say $A$ be partitioned as

$$\bar{A}d = (\bar{B}, \bar{S}, \bar{R}) \begin{pmatrix} \bar{d}_B \\ \bar{d}_S \\ \bar{d}_N \end{pmatrix} = 0$$ (4.1)

where the basic stepdirection $\bar{d}_B$ is used to satisfy the constraints set, the superbasic stepdirection $\bar{d}_S$ is allowed to vary to minimize $F(X)$ (1.1) and the nonbasic stepdirection $\bar{d}_N$ is zero, such that set $\hat{V}$ is fixed at any of their bounds. Here $\bar{B} = (\bar{B}, \bar{S})$ and $\bar{B}$ is a $\bar{F}$ nonsingular matrix. At each iteration, the problem then becomes determining vector $d = (d_B, d_S, d_N)^t$ so that it is feasible-descent.

Since $\bar{d}_N = 0$ and $\bar{d}_S$ is allowed to be free, it results

$$d_B = -\bar{B}^{-1}\bar{S} \bar{d}_S$$

such that the variable-reduction characterization of matrix $\bar{S}$ can be written

$$\bar{z} = \begin{pmatrix} -\bar{B}^{-1}\bar{B} \\ 1 \\ 0 \end{pmatrix}$$ (4.3)

so that (2.3) holds.

The quadratic approximation of the unconstrained reduced problem of minimizing $F(X)$ -- in the manifold $\hat{W}$ and $\hat{V}$ as a function of the current superbasic set of variables $\bar{d}_S$ can be written

$$\text{minimize } \bar{h}^T \bar{d}_S + 1/2 \bar{d}_S^T \bar{H} \bar{d}_S$$ (4.4)

where $\bar{h}$ and $\bar{H}$ are given by (2.1) and (2.14), respectively. Note that $\bar{h}$ can also be written

$$\bar{h} = \bar{S} \bar{d}_S - \bar{S} \bar{h}_B$$ (4.5)

where $\bar{h}_B$ solves the linear system

$$\bar{g}_B = \bar{B} \bar{h}_B$$ (4.6)
such that \( \tilde{\mathbf{g}}_{BS} \in (\tilde{\mathbf{g}}_{B}, \tilde{\mathbf{g}}_{S}) \) where \( \tilde{\mathbf{g}}_{B} \) and \( \tilde{\mathbf{g}}_{S} \) are the basic and superbasic gradients, respectively. Theoretically, the algorithm continues till \( ||\tilde{h}|| = 0 \) or the superbasic set is empty and, then, the deactivating process is executed by analyzing the Lagrange multipliers (except if it has been decided to do so in the presence of 'quasi-optimal' solutions so that Lagrange multipliers estimates are obtained).

Let assume that \( \tilde{\mathbf{x}} \) is an optimal point in the manifold \( \tilde{\mathbf{w}} \) and \( \tilde{\mathbf{v}} \). Then, the Lagrange multipliers vector \( \tilde{\mathbf{u}} \) is obtained by solving system (2.9); let \( \tilde{\mathbf{u}}_{BS} \) be the solution. Since \( \|
abla \tilde{h} \| = 0 \) or the superbasic set is empty, it is clear that \( \tilde{\mathbf{u}}_{BS} = \mathbf{u}_{B} \) (4.6). Thus, it is not required to solve (2.9) since \( \mathbf{u}_{B} \) is updated at each iteration to solve problem (4.4) (see /12/, /2/). Note also that \( \mathbf{u}_{B} \) is the negative of the LP simplex multipliers. The Lagrange multipliers vector \( \tilde{\mathbf{u}} \) of the active (nonbasic) structural variables is obtained by using formula (2.8); note also that \( \tilde{\mathbf{h}} \) and \( \tilde{\mathbf{r}} \) are the negative of the LP reduced cost vectors related to the 'nonbasic' and 'superbasic' associated LP subproblems, respectively.

When a 'quasi-optimal' solution, say \( \tilde{\mathbf{x}} \) is obtained in a given manifold defined by \( \tilde{\mathbf{w}} \) and \( \tilde{\mathbf{v}} \), \( \tilde{\mathbf{u}}_{BS}, \mathbf{u}_{B} \) and \( \tilde{\mathbf{r}}_{BS}, \mathbf{r}_{B} \). Let us term \( \tilde{\mathbf{u}}_{B} \) and \( \mathbf{u}_{B} \) as basic-based active constraints and bounds Lagrange multipliers estimates, respectively; and \( \tilde{\mathbf{u}}_{BS} \) and \( \mathbf{u}_{BS} \) as basic-superbasic-based active constraints and bounds Lagrange multipliers estimates, respectively. Of course, estimates based on the basic-superbasic set are generally more accurate than those based in the basic set; see in /11/ a good discussion on the subject.

The motivation for obtaining 'quasi-optimal' solutions and, then, interrupting the mini- mization of the unconstrained reduced nonl2 near problem is based on the assumption that it is likely that current sets \( \tilde{\mathbf{w}} \) and \( \tilde{\mathbf{v}} \) are not the optimal sets \( \mathbf{w} \) and \( \mathbf{v} \) in problem (1.1)-(1.2) and, then, it could be beneficial to analyze if it is worthy to delete some active constraint or bound before reaching the optimum in the reduced problem, but after reaching a solution close to that optimum.

Note that estimation \( \tilde{\mathbf{u}}_{B} \) is already obtained. Estimation \( \tilde{\mathbf{u}}_{BS} \) is based on the QR-factorization of matrix \( \tilde{\mathbf{B}} \). While solving the unconstrained reduced problem, either a basic or a superbasic variable may strike a bound during the search. If a superbasic variable strikes a bound, it is made nonbasic, the dimension of the manifold is reduced by one, and the search continues. If a basic variable strikes a bound, the basic variable is exchanged with an appropriate superbasic variable and the resulting superbasic variable is made nonbasic. The estimation \( \tilde{\mathbf{u}}_{B} \) is updated at each iteration since \( \tilde{\mathbf{r}} \) (4.5) requires it; but \( \tilde{\mathbf{u}}_{BS} \) is only obtained when it is required to execute the deactivating process.

Estimation \( \tilde{\mathbf{u}}_{BS} \) is obtained by minimizing the square of the euclidean length

\[
||\tilde{\mathbf{u}}_{BS} - (\mathbf{B}^T)^{T}\mathbf{u}_{BS}||_2
\](4.7)

For obtaining \( \tilde{\mathbf{u}}_{BS} \) (minimal linear least square solution) we use an implementation of the Gram-Schmidt QR-factorization of a matrix -- where the number of rows is greater than the number of columns; see the motivation and details in /1/, /4/), but the matrices involved are as follows.

Let \( \tilde{\mathbf{e}} \) be a \((n-\tilde{n}) \cdot \tilde{\mathbf{e}} \) orthogonal matrix and \( \tilde{\mathbf{r}} \) a \( \tilde{\mathbf{e}} \cdot \tilde{\mathbf{e}} \) nonsingular upper triangular matrix -- with identity diagonal such that

\[
(\mathbf{B}^T)^T = \tilde{\mathbf{QR}}
\](4.8)
It can be shown that the vector \( \hat{\nu}_{BS} \) that minimize problem (4.7) is also the vector that satisfies the system

\[
\hat{\nu}_{BS} = \hat{Q} \hat{z}_{BS}
\]  
(4.9)

Then, it is required to calculate \( \hat{Q} \) and \( \hat{R} \) -- for obtaining \( \hat{\nu}_{BS} \); they are updated each time a basic-superbasic (structural or slack) variable is made nonbasic or a nonbasic (structural or slack) variable is deactivated; they are calculated anew the first time that \( \hat{\nu}_{BS} \) is used, or after a given number of updating so that some unstability due to too-many intermediate operations is avoided /1, - /4/.

5. DEACTIVATING PROCESS.

Assume that \( \hat{x} \) is an 'optimal' or 'quasi-optimal' solution in the manifold defined by \( \hat{W} \) -- and \( \hat{V} \); assume also that the given algorithm selects, at each time, only one candidate -- nonbasic (structural or slack) variable to be deactivated. Note that not all nonbasic variables are candidates for being deactivated; e.g. if \( \hat{x} \) is only 'quasi-optimal', 'unsafe' nonbasic variables do not belong to the candidate set if an anti-zigzagging strategy is to be used. See /2, /5/. Let \( C_1 \) and \( C_2 \) be candidate sets of structural nonbasic variables and active inequality constraints, respectively. Note that each slack variable is associated with an inequality constraint.

Let \( \gamma \) be an indicator such that \( \gamma = 1 \) means -- the basic superbasic-based estimate \( \hat{\nu}_{BS} \) (4.6) is allowed for \( \hat{x} \); otherwise \( \gamma = 0 \), only the basic-based estimate \( \hat{\nu}_{B} \) (4.9) can be used.

Since \( \hat{\nu}_{BS} = \hat{Q} \hat{z}_{BS} \) for \( \hat{x} \) being optimal in the manifold \( \hat{W} \) and \( \hat{V} \), indicator \( \gamma \) is set to zero -- for the given iteration even if basic-superbasic-based Lagrange multipliers estimates -- are allowed.

Case \( \gamma = 0 \).

The structural nonbasic variable to be deac-

ativated is the variable, say \( k \hat{\nu} \) with the most favorable basic-based Lagrange multi-

plier estimate \( \hat{\nu}_{B_k} \) such that

\[
|\hat{\nu}_{B_k}| = \max (|\hat{\nu}_{B_j}|) \quad j \notin C_1
\]

(5.1)

where \( \epsilon_1 \) is given small positive tolerance (typically, \( \epsilon_1 = 10^{-4} \)) that is intended to give lower priority to variables with zero or near-to-zero Lagrange multipliers estimates.

If \( k = 0 \) then the active inequality constraint, say \( k \hat{W} \) with the most favorable basic-based Lagrange multiplier estimate \( \hat{\nu}_{B_k} \) to be deactivated, such that

\[
|\hat{\nu}_{B_k}| = \max (|\hat{\nu}_{B_k}|) \quad j \notin C_1
\]

(5.2)

for the same tolerance \( \epsilon_1 \).

Assume that \( k = 0 \). If the following condition does not hold

\[
(3j \notin C_1 \land \hat{\nu}_{B_j} \leq \epsilon_1) \lor (3i \notin C_2 \land \hat{\nu}_{B_i} \leq \epsilon_1)
\]

(5.3)

the action to be taken depends on the character of point \( \hat{x} \): if it is a 'quasi-optimal' -- solution in the manifold \( \hat{W} \) and \( \hat{V} \), the next -- iteration obtains the related superbasic stepdirection; otherwise, it is assumed that \( \hat{x} \) is also an optimal solution in problem --- (1.1)---(1.2). If (5.3) holds, let us redefine sets \( D_1 \) and \( D_2 \) (see sec. 2) such that \( D_1 \) takes the subset of \( C_1 \) for which the first part of (5.3) holds and \( D_2 \) takes the subset of \( C_2 \) for which the second part of (5.3) holds. Thus, set \( D_1 \cup D_2 \) defines the zero or near-to-zero Lagrange multipliers estimates.

Case \( \gamma = 1 \).

Note that estimates \( \hat{\nu}_{BS} \) and \( \hat{\nu}_{B} \) (and, then, - \( \hat{\nu}_{BS} \) and \( \hat{\nu}_{B} \)) are to be used. Generally, \( \hat{\nu}_{BS} \) --

\textit{Questa86 – V. 6, n° 2 (juny 1982)}

199
is more accurate than \( \hat{u}_B \) since it uses more information; but, it makes sense to select a candidate nonbasic (structural or slack) variable if among other requirements /5/, both estimates agree in the appropriate sign. In any case, note that the reduced gradient --- (with structural and, although provisionally, slack elements) only uses the basic-based estimate; see (4.5).

The structural nonbasic variable to be deactivated is the variable, say \( k \cdot \hat{v} \hat{W} \) with the most favorable basic-superbasic-based Lagrange multiplier estimate \( \hat{\lambda}_{BS_k} \) whose basic-based estimate \( \hat{\lambda}_{UB_k} \) agrees in sign or, at least, is zero or near-zero, such that

\[
|\hat{\lambda}_{BS_k}| = \max \{ |\hat{\lambda}_{BS_1}|, |\hat{\lambda}_{BS_2}|, \ldots \}
\]

(5.4)

If \( k=0 \) the active inequality constraint to be deactivated is the constraint, say \( k \cdot C_{j} \cdot \hat{w} \hat{W} \) with the most favorable basic-super basic-based Lagrange multiplier estimate --- \( \hat{\lambda}_{BS_k} \) whose estimation \( \hat{\lambda}_{UB_k} \) agrees in sign or, at least, is zero or near-zero, such that

\[
|\hat{\lambda}_{BS_k}| = \max \{ |\hat{\lambda}_{BS_1}|, |\hat{\lambda}_{BS_2}|, \ldots \}
\]

(5.5)

for the same tolerance \( \varepsilon_1 \) used above.

When using (5.4) and (5.5) it is suggested to give the lowest priority to the candidate nonbasic (structural or slack) variables for which the basic-based Lagrange multipliers estimates are zero or near-zero.

When a candidate nonbasic (structural or --- slack) variable is deactivated with zero or near-zero basic-based Lagrange multiplier estimate, its related element in the new reduced gradient \( \hat{h} \) (4.5) will be \( \varepsilon_1 + \varepsilon_2 \) if \( \hat{h}_j = U_j \) or \( \hat{y}_i = B_1 - b_1 \), respectively, and \(-(-\varepsilon_1 \varepsilon_2)\) if \( \hat{h}_j = C_1 \) or \( \hat{y}_i = 0 \), respectively. Note that \( \hat{h} \) will be used for obtaining the superbasic -- stepdirection of the next iteration. \( \varepsilon_2 \) is a given small positive tolerance (typically, \( \varepsilon_2 = 10^{-4} \)) used for the required perturbation in the appropriate sign of the zero or --- near-to-zero basic-based Lagrange multipliers estimates.

Assume that \( k=0 \). If the following condition does not hold

\[
(\exists j \in J) |\hat{\lambda}_{BS_j}| \leq \varepsilon_1 \quad \text{and} \quad (\exists c \in C_1) \{-c \hat{u}_{BS_1} \leq \varepsilon_1 \}
\]

(5.6)

the next iteration obtains the superbasic -- stepdirection; note that \( \hat{h} \) is quasi optimal. Let \( D_1 \cup D_2 \) define the set of zero or near-to-zero basic-superbasic-based Lagrange multipliers estimates as it was defined for \( \gamma=0 \), such that their related basic-based estimates are favorable or, at least, zero or near-to-zero.

6. ESTIMATING THE TENDENCY OF THE LAGRANGE MULTIPLIERS ESTIMATES.

When conditions (5.3) for \( \gamma=0 \) and (5.6) for \( \gamma=1 \) hold, the ambiguity on the Lagrange multipliers estimates does not allow to select a variable from set \( D_1 \cup D_2 \) to be deactivated, except if some perturbation on the active bounds of the X- and Y-variables is produced; in this way, an estimation on the tendency of the zero or near-to-zero Lagrange multipliers estimates could be obtained.

Before describing the procedure and since -- the structural variables may be classified -- according to the linearity of the terms -- of the objective function, let us make the following partition of these variables in pure linear, linear with variable-coefficient and nonlinear; see in /7/ how to use, in a given algorithm, this and other types of variables partition. A variable is pure linear if its coefficient is constant in all terms; a variable is linear with-variable-coefficient if, for a given value of the other variables that are used in the same term, it is a linear function of the given variable; and a variable is nonlinear if, for a given value of the other variables in the same term, it is a nonlinear function of the given variable. An example is as follows: \( F(X) = 4X_1 + X_2 \log X_3 \); variable \( X_1 \) is pure linear, \( X_2 \) is linear with variable-coefficient, and \( X_3 \) is nonlinear. Let \( I \) define the set of li-
near variables and $P$ the set of pure linear variables.

The estimation of the tendency of the zero or near-to-zero Lagrange multipliers estimates is only obtained for the basic-superbasic-based estimates; but, note that $\tilde{\lambda}_{BS}^P_B$ and $\tilde{\lambda}_{BS}^P_B$ (and, then, $\tilde{\lambda}_{BS}^P_B$) for $\gamma=0$.

The tests to select a variable to be deactivated from set $D_1\setminus D_2$ could be as follows.

**Test 1.** Select a (structural) variable from set $D_1$.

It is only performed if $J/\tilde{\Phi}=P$; i.e., the whole basic and superbasic set of variables is pure linear; then, estimation $\tilde{\lambda}_{BS}$ will not be modified by the perturbation to be produced in test 1.

Let us produce an small perturbation on the active bound of variable, say $j$ for $j\in D_1\setminus (J/L)$; let $\bar{g}_j$ be the element in the objective function gradient related to variable $j$ evaluated at the perturbed solution $\tilde{x}$, such that $\bar{x}_j=\tilde{x}_j$ for $j\notin J$, $\bar{x}_j=\tilde{x}_j+\epsilon_2$ if $\tilde{x}_j=1$ and $\tilde{x}_j=U_j$. Note that only element $\bar{g}_j$ is required at each perturbation, being unchanged the other elements of $g$. Note also that variable $j$ is nonlinear, since the gradient elements of linear variables are constant (and then, the solution perturbation has not any effect).

The perturbed basic-superbasic-based estimate $\tilde{\lambda}_{BS}^P_j$ can be written (see (2.8))

$$\tilde{\lambda}_{BS}^P_j = \tilde{\lambda}_{BS}^P_j + \bar{g}_j \epsilon_2$$

(6.1)

If the following condition holds, variable $j$ is to be deactivated (and, then, set $k=j$).

$$(\tilde{\lambda}_{BS}^P_j - \bar{\lambda}_{BS}^P_j)^2 < (\epsilon_3 + 1) \tilde{\lambda}_{BS}^P_j (\tilde{\lambda}_{BS}^P_j - \bar{\lambda}_{BS}^P_j)^2 < (\epsilon_3 \bar{\lambda}_{BS}^P_j)^2$$

(6.2)

Otherwise, it seems that the tendency of $\tilde{\lambda}_{BS}^P_j$ is not favorable and then, variable $j$ is not to be deactivated. $\epsilon_3$ is a given small positive tolerance, such that any number with magnitude less than $\epsilon_3$ will be discarded (i.e., set to zero) as being insincere in any circumstance; typically, $\epsilon_3=10^{-12}$.

The first variable $j\in D_1\setminus (J/L)$ that satisfies (6.2) is the variable, say $k$ to be deactivated.

**Test 2.** Select a (structural) variable from set $D_1$.

If $k=0$ (i.e., test 1 did not select any variable) test 2 will be used; it is not performed if $J/\tilde{\Phi}=P$; i.e., the whole basic and superbasic set of variables is pure linear since, otherwise, the results would be very similar to those obtained in test 1.

Let us perturb simultaneously the active bounds of the whole set $D_1\setminus (J/L)$ with the same criterion used in test 1; gradients $\bar{g}_j$ -- for $\gamma=0$ and $\bar{g}_j$ for $\gamma=1$ are, alternatively, evaluated at the perturbed solution $\tilde{x}$ and, then, the perturbed basic-based estimate $\tilde{\lambda}_{BS}^P$ is obtained for $\gamma=0$ by solving system (4.6) with $\bar{g}_j$ being substituted by $\bar{g}_j$ and the perturbed basic-superbasic-based estimate $\tilde{\lambda}_{BS}^P$ is obtained for $\gamma=1$ by solving problem (4.7) being $\bar{g}_j$ substituted by $\bar{g}_j$.

The scope of this work does not cover the procedures for solving system (4.6), nor problem (4.7); in any case, inverse matrix $B^{-1}$ is not obtained in the first case and system (4.9) is not explicitly solved in the second case. The updated $\bar{L}_0$ and $\bar{B}$ factorizations -- of matrices $\bar{B}$ and $(\bar{B})^t$ respectively are used; see /1,2,4/.

The perturbed estimate $\tilde{\lambda}_{BS}^P_j$ for $j\in D_1\setminus (J/L)$ is obtained by using in (2.8) the perturbed element $\bar{g}_j$ evaluated at the perturbed solution $\tilde{x}$ and the perturbed estimate $\tilde{\lambda}_{BS}^P$ note that $\tilde{\lambda}_{BS}^P \approx \bar{\lambda}_{BS}^P$ for $\gamma=0$.

The first variable $j\in D_1\setminus (J/L)$ that satisfies (6.2) is the variable, say $k$ to be deactivated.

**Test 3.** Select an (active inequality) constraint from set $D_2$.

We will obtain directly, in test $t3$, the basic-superbasic-based estimate $\tilde{\lambda}_{BS}^P$ of the ---
constraints set \( \tilde{d} \) that solves the problem --
\[
\min \| \tilde{a}_{BS} - (\tilde{d} \tilde{x})^T \tilde{u}_{BS} \|_2^2
\]  
(6.3)
for \( \gamma = 1 \), or solves the system
\[
\tilde{a}_{BS} = \tilde{x}^T \tilde{u}_{BS}
\]  
(6.4)
for \( \gamma = 0 \), where \( \tilde{u}_{BS} = \tilde{u} \) for \( \gamma = 0 \),
\[
(\tilde{d} \tilde{x})^T = \tilde{a}_{BS} \equiv (\tilde{x}_{BS}^T \tilde{a}_{BS}^T),
\]
such that \( \tilde{x} \) and \( \tilde{x}_{BS} \) take the optimal or quasi-optimal values of the basic set and basic and superbasic set of variables, respectively, and \( \tilde{a}_{BS} \) are the related deviations from points \( \tilde{x}_{BS} \) and \( \tilde{x} \).

Deviations \( \tilde{d}_{BS} \) satisfies the right-hand-side perturbed original problem, where the nonbasic variables are fixed to their values \( \tilde{x}_N \) and only the active constraints set \( \tilde{a} \) is considered, such that
\[
\tilde{a}_{BS}(\tilde{x}_{BS} + \tilde{d}_{BS}) = \tilde{x}_N^T \tilde{x}_N + \tilde{e}_2^T \tilde{e}_2
\]  
(6.5)
where \( \tilde{x}_j = 1 \) for \( j \in \tilde{a} \), and \( \tilde{x}_j = 0 \) for \( j \notin \tilde{a} \).

Note that \( \tilde{x}_{BS} \) takes the right-hand-side vector to be perturbed in system (6.5), where \( \tilde{d}_{BS} \) takes the constraints matrix such that -- \( \tilde{d}_{BS} \) takes the \( (r,c) \)-th element of matrix \( \tilde{B} \) for \( r = 1, \ldots, t \) and \( c = 1, \ldots, n-t-r \). Let \( i(r) \) define the constraint \( i \in \tilde{a} \) related to index \( r \) in (6.5); similarly, \( j(c) \) defines the basic or superbasic variable \( j \in \tilde{a} \) related to index \( c \) in (6.5). \( \epsilon_2 \) defines a 2-vector such that \( \epsilon_2 = 0 \) for \( r \) such that \( i(r) \notin D_2 \), \( \epsilon_2 = 1 \) for \( r \) such that \( i(r) \in D_2 \), and \( \epsilon_2 = 0 \) for \( r \) such that \( i(r) \notin D_2 \).

In a similar way, deviation \( \tilde{d}_{BS} \) solves the --
\[
\tilde{B}(\tilde{x}_{BS} + \tilde{d}_{BS}) = \tilde{B}_x \tilde{x}_N + \tilde{e}_2^T \tilde{e}_2
\]  
(6.6)
such that \( \tilde{B}_{rc} \) takes the \( (r,c) \)-th element of matrix \( \tilde{B} \) for \( r = 1, \ldots, t \) and \( c = 1, \ldots, t \). Let -- \( i(r) \in \tilde{a} \) define the constraint related to index \( r \) in (6.6); similarly, \( j(c) \) defines the basic variable related to index \( c \) in (6.6). Vector \( \tilde{e}_2 \) in (6.6) is defined as in (6.5).

Note that constraint \( i \in \tilde{a} \rightarrow \tilde{a} \) will be only perturbed if its Lagrange multiplier estimates \( \tilde{u}_{BS} \) is zero or near-to-zero for \( \gamma = 0 \), or its Lagrange multiplier estimates \( \tilde{u}_{BS} \) is zero or near-to-zero and \( \tilde{u}_{BS} \) is not non-favorable (i.e., it is favorable or, at least, zero or near-to-zero) for \( \gamma = 1 \).

For obtaining the perturbed gradient \( \tilde{d}_{BS} \) in (6.3) and \( \tilde{d}_{BS} \) in (6.3) such that the above approach be practical, it is required a fast procedure for obtaining \( \tilde{d}_{BS} \) in (6.6) and \( \tilde{d}_{BS} \) in (6.5).

Case \( \gamma = 0 \).
The deviation \( \tilde{d}_{BS} \) that satisfies (6.6) for \( \gamma = 0 \) perturbation \( \epsilon_2 \) in \( \tilde{B}_{rc} \) for \( i(r) \in \tilde{a} \) is such that
\[
\tilde{d}_{BS} = \epsilon_2 B_{-1} \tilde{e}_2
\]
(6.7)
where \( B_{-1} \) takes the \( r \)-th column in matrix \( \tilde{B} \), such that \( \tilde{d}_{BS} \) takes the deviation from the solution \( \tilde{x}_{j(c)} \). Note that inverse basic matrix \( \tilde{B}^{-1} \) is not explicitly calculated, since its triangular factors \( \tilde{L} \) (lower) and \( \tilde{U} \) (upper) are fresh anew or kept updated in the previous iterations /2/. In any case, \( \tilde{B}^{-1} \) is postmultiplied by a vector for solving (6.7) and premultiplied by a vector for solving (6.6).

Case \( \gamma = 1 \).
The deviation \( \tilde{d}_{BS} \) that satisfies (6.6) for an small enough perturbation \( \epsilon_2 \) in \( \tilde{B}_{BS} \) for \( \gamma = 1 \) can be written
\[
\tilde{d}_{BS} = \epsilon_2 \tilde{B}_{BS} + \tilde{e}_2 \tilde{e}_2
\]  
(6.8)
where
\[
\tilde{B}_{BS} + (\tilde{B}_{BS})^{-1} \tilde{d}_{BS} = \tilde{B}_{BS}
\]  
(6.9)
is the \( (n-t-r) \times (n-t-r) \) pseudo-inverse matrix of the \( (n-t-r) \times t \) matrix \( \tilde{B}_{BS}^T \).
In effect /4,2/, from (6.5) it results
\[
\text{BS}(\hat{\text{BS}}^t) = \hat{\text{BS}} + \varepsilon_2 \hat{I}_2
\]
\[
= \hat{\text{BS}} + \varepsilon_2 \text{BS}(\hat{\text{BS}}^t)(\hat{\text{BS}}(\hat{\text{BS}}^t))^{-1} \hat{I}_2
\]
\[
= \hat{\text{BS}} + \varepsilon_2 \text{BS}(\hat{\text{BS}}^t) \hat{I}_2
\]
\[
= \hat{\text{BS}}(\hat{\text{BS}}^t) \hat{I}_2
\]  
(6.10)

and finally,
\[
\hat{\text{BS}} - \varepsilon_2 \sum_{r \in D_2} (\hat{\text{BS}}^t)^{-1} \hat{I}_2
\]
\[- \sum_{r \in D_2} (\hat{\text{BS}}^t) \hat{I}_2}
(6.11)

where \((\hat{\text{BS}}^t)^{-1}\) takes the r-th row in matrix --
\(\hat{\text{BS}}^t\)\(^{+}\), such that \(\hat{\text{BS}}\) takes the deviation
from the solution \(\hat{x}_j(c)\) for \(j(c) \in J\).\(^{+}\).

Since \((\hat{\text{BS}}^t)^{+} = \hat{\text{BS}}^t(4,8)\) and \((\hat{\text{BS}}^t)^{+}\) can be expressed by (6.9), it results
\[
\hat{R}(\hat{\text{BS}}^t)^{+} = \hat{R}^t
\]  
(6.12)

such that the \(t\)-vector \((\hat{\text{BS}})^{+}\) for
c=1,\ldots,n\^t-r can be written
\[
(\hat{\text{BS}})^{+}_{ct} = \hat{R}^{t}_{ct}
\]  
(6.13a)

\[
(\hat{\text{BS}})^{+}_{rc} = \hat{R}^{t}_{rc} - \sum_{r=t+1}^{t} \hat{R}^{t}_{rc}(\hat{\text{BS}})^{+}_{ct}
\]  
(6.13b)

Note that not all rows \((\hat{\text{BS}})^{+}\) are required, -
but only rows \(r=1,\ldots,t\) such that \(i(r) \in D_2\); -
then, it could be possible that matrix \((\hat{\text{BS}})^{+}\)
is not required to be completely calculated.

Once obtained \(\hat{x}_B + \hat{d}_B\) for \(\gamma=0\) and \(\hat{x}_B + \hat{d}_B\) --
for \(\gamma=1\), perturbed estimates \(\hat{u}_B\) and \(\hat{u}_BS\) are
obtained from (6.4) and (6.3), respectively. Setting \(\hat{u}_BS = \hat{u}_B\) for \(\gamma=0\), the first active ---
inequality constraint \(I_{D_2}\) for which the ---
following condition holds is to be deactiva-
ted (and, then, set \(k=1\)).

\[
(\hat{u}_BS - \hat{u}_BS^t < \varepsilon_3^t \hat{I}_2 = 0) \vee \hat{u}_BS^t > \hat{u}_BS \geq \varepsilon_3 \hat{I}_2 = \hat{d}_1 - \hat{d}_1
\]  
(6.14)

such that a priority is given for \(\gamma=1\) to the
constraint with the maximum absolute value in
its estimate \(\hat{u}_B\) if it is favorable.

If \(k=0\) it seems that the tendency on \(\hat{u}_BS\) is
not favorable; the action to be taken depends
on the character of \(\hat{x}\). If it is a 'quasi-opti-
mal' solution in the manifold \(\hat{\text{w}}\) and \(\hat{\text{v}}\), the
next iteration obtains the related superbasic
stepdirection. If \(\hat{x}\) is an optimal point in --
that manifold, it is also assumed that it is an
optimal point in problem (1.1)-(1.2).

7. CONCLUSION

The influence of the degenerate sets of acti-
ve inequality constraints and bounds in the
optimality conditions has been analyzed for
local optimal points in linearly constrained
nonlinear programming (LCNP) problems. Some
procedures have been described to get some
insight on the tendency of zero or near-to-
zero Lagrange multipliers estimates for non-
basic (structural and slack) variables in --
the frame of a given LCNP algorithm.

8. REFERENCES

/1/ ESCUDERO, L.F., "An implementation of --
the QR factorization for solving overde-
termixed systems of linear equations".

/2/ ESCUDERO, L.F., "An algorithm for large-
scale quadratic programming problems and
its extensions to the linearly constrai-
ned case"; IBM Madrid Scientific Center,
report SCR-01.81, Madrid, 1981.

/3/ ESCUDERO, L.F., "Lagrange multipliers es-
timates for constrained minimization",

/4/ ESCUDERO, L.F., "QR factorization and --
its updatings"; IBM Madrid Scientific Cen-
ter, report SCR-01.82, Madrid, 1982. See
also, "On QR factorization updateings" --
Trabajos de Estadística e Investigación
Operativa (in printing).

/5/ ESCUDERO, L.F., "Dirección de búsqueda --
en programación no lineal en gran escala

QUESTIO - V. 6, n° 2 (juny 1982)
203
con condiciones lineales y escasa densidad. Método Newton-Truncado reducido,
Anales de Ingeniería Mecánica 1 (1982).

See also "On diagonally-preconditioning
the Truncated-Newton method for super-
scale linearly constrained nonlinear
programming, IBM Madrid Scientific Cen-

/6/ ESCUDERO, L.F., "Programación matemática;
División de Matemáticas, Facultad de Cien-
cias, Universidad Autónoma de Madrid, Ma-
drid, 1982.

/7/ ESCUDERO, L.F., "Strategies selection on
linearly constrained nonlinear program-
ning" IBM Madrid Scientific Center, re-
port SCR-02.82, Madrid, 1982.

/8/ Fletcher, R., "Practical methods for op-

/9/ GILL, P.E., and MURRAY, W., "Numerical
methods for constrained optimization" --

/10/ GILL, P.E. and MURRAY, W., "The computa-
tion of Lagrange multipliers estimates
for constrained minimization", Mathema-
tical Programming 17 (1979) 32-60.

/11/ GILL, P.E., MURRAY, W. and WRIGHT,M.H.,
"Practical optimization"(Academic Press,

/12/ MANGASARIAN, O., "Nonlinear programming"

/13/ Murtagh, B.A. and SAUNDERS, M.A.,"Large-
scale linearly constrained optimization"