QÜESTIIÓ, vol. 23, 1, p. 61-83, 1999

ON THE COMPATIBILITY OF CLASSICAL MULTIPLIER ESTIMATES WITH VARIABLE REDUCTION TECHNIQUES WHEN THERE ARE NONLINEAR INEQUALITY CONSTRAINTS

E. MIJANGOS
Zientzi Fakultatea (EHU)*
N. NABONA
Universitat Politècnica de Catalunya**

The minimization of a nonlinear function subject to linear and nonlinear equality constraints and simple bounds can be performed through minimizing a partial augmented Lagrangian function subject only to linear constraints and simple bounds by variable reduction techniques. The first-order procedure for estimating the multiplier of the nonlinear equality constraints through the Kuhn-Tucker conditions is analyzed and compared to that of Hestenes-Powell. There is a method which identifies those major iterations where the procedure based on the Kuhn-Tucker conditions can be safely used and also computes these estimates. This work justifies the extension of the former results to the case of general inequality constraints. To this end two procedures that convert inequalities into equalities are considered.

Keywords: Nonlinear programming, general inequality constraints, augmented lagrangian, variable reduction, Lagrange multiplier estimates.

---

* Department of Applied Mathematic, Statistics and Operations Research, Zientzi Fakultatea (EHU), P.O. Box 644, 48080 Bilbao (Espanya). Corresponding author e-mail: mepmifee@lg.ehu.es.

** Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Pau Gargallo, 5. 08028 Barcelona (Espanya).

– Received February 1998.
– Accepted October 1998.

1. INTRODUCTION AND MOTIVATION

Consider the linearly and nonlinearly constrained problem

$$\begin{array}{llrl} (1) & \text{minimize} & & f(x) \\ (2) & \text{subject to:} & Ax &= b \\ (3) & & c(x) &= 0 \\ (4) & & & l \leqslant x \leqslant u, \end{array} \qquad \textbf{(EP)}$$

where

(1) $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. $f(x)$ is nonlinear and twice continuously differentiable on the feasible set defined by constraints (2–4).

(2) A is an $m \times n$ matrix and b an m-vector.

(3) $c : \mathbb{R}^n \longrightarrow \mathbb{R}^r$, is such that $c = [c_1, \cdots, c_r]^t$, $c_i(x)$ being linear or nonlinear and twice continuously differentiable on the feasible set defined by constraints (2) and (4) $\forall i = 1, \cdots, r$.

(4) $n \gg m + r$.

To solve this problem one could use, among others, partial augmented Lagrangian techniques [1, 2, 3] as in [9, 11, 6], where only the general constraints (3) are included in the Lagrangian. In the application of these techniques there are two fundamental steps. The first solving

$$\begin{array}{llrl} (5) & \underset{x}{\text{minimize}} & & L_\rho(x, \mu) \\ (6) & \text{subject to:} & Ax &= b \\ (7) & & & l \leqslant x \leqslant u, \end{array} \qquad \textbf{(ES)}$$

where $\rho > 0$ and μ are fixed,

$$L_\rho(x, \mu) = f(x) + \mu^t c(x) + \frac{1}{2}\rho c(x)^t c(x).$$

Should the solution $\widetilde{x}$ obtained be infeasible with respect to (3), the second step, which is the updating of the estimate μ of the Lagrange multipliers of constraints (3), is carried out, also updating, if necessary, the penalty coefficient ρ, then going back to the first step. Should $\widetilde{x}$ be feasible (or the violation of constraints (3) be sufficiently small) the procedure ends. It is of paramount importance that the multipliers estimate μ be as accurate as possible, otherwise the convergence of the algorithm can be severely handicapped, as shown in [1, 2, 3].

In practice there are two first-order procedures to estimate μ. On the one hand the method put forward by Hestenes [8] and Powell [15]

$$\widetilde{\mu} = \mu + \rho c(\widetilde{x}),$$

and on the other hand μ_L obtained through the classical solution to the system of Kuhn-Tucker necessary conditions

$$\nabla f(\widetilde{x}) + \nabla c(\widetilde{x})\mu + A^t\pi + \lambda = 0 \tag{8}$$

by least squares, as suggested in [7]. However, whereas the first procedure can be always used for any $\widetilde{x}$, without hampering the convergence [2, 3], this is not the case with the second procedure, as system (8) is only known to be compatible at the optimizer x^*, not being necessarily so at $\widetilde{x}$, thus possibly giving rise to bad estimates μ_L, shown up by large residuals for system (8).

Section 2 of this work presents a study of the viability of using this multiplier estimation technique within the minimization of a Partial Augmented Lagrangian subject to linear constraints and bounds by the Murtagh and Saunders procedure [13] for problem **EP**. (An alternative development of the contents of Section 2 can be found in [10].)

Sections 3 and 4 consider two ways of extending these results to problem

$$\left.\begin{array}{lll} (9) & \text{minimize} & f(x) \\ (10) & \text{subject to:} & Ax = b \\ (11) & & \underline{c} \leq c(x) \leq \overline{c} \\ (12) & & l \leqslant x \leqslant u, \end{array}\right\} \quad \textbf{(IP)}$$

where $\underline{c}_j < \overline{c}_j$, for $j = 1, \ldots, r$. In Section 3, a vector of slacks «y» is used to convert constraints (11) into equalities, and, in Section 4, slacks «y» and artificial variables «w» are used with the same aim. Section 5 contains the conclusions.

2. ANALYSIS OF COMPATIBILITY

The compatibility of the multiplier estimate obtained through the classical solution to the system of Kuhn-Tucker necessary conditions with the variable reduction techniques and its relationship with that of Hestenes and Powell are analyzed along this section.

Let us consider the first-order conditions associated with a local optimizer x^* of the problem **EP**, which are

$$\begin{aligned} \nabla f(x^*) + \nabla c(x^*)\mu^* + A^t\pi^* + \lambda^* &= 0 \\ Ax^* &= b \\ c(x^*) &= 0 \\ \lambda_i^* &\leq 0, \quad \text{if } x_i^* = l_i \\ \lambda_i^* &\geq 0, \quad \text{if } x_i^* = u_i \\ \lambda_i^* &= 0, \quad \text{otherwise} \end{aligned}$$

so that unique vectors μ^*, π^* and λ^* exist, such that the first equation holds. These vectors are denoted *Lagrange multipliers*; being $\nabla c(x) = [\nabla c_1(x), \cdots, \nabla c_r(x)]$. (The gradient is considered to be a column vector).

Throughout this work we assume

AS1. x^* is a *regular point* — i.e., the Jacobian of the active constraints at x^* has full rank.

Solving problem **EP** through a partial augmented Lagrangian techniques consists basically of the following algorithm, where for given $\rho > 0$ and μ the subproblem **ES** (5–7) is successively solved.

Algorithm 2.1.

1. For an initial point x_0 (not necessarily feasible with respect to $c(x) = 0$), a given scalar $\rho > 0$ and vector μ, solve subproblem **ES** and obtain its optimizer $\widetilde{x} = x(\mu, \rho)$.
2. Should this $\widetilde{x}$ make $c(\widetilde{x})$ to be zero or nearly so for a prespecified tolerance, $x^* = \widetilde{x}$ and the problem is solved, otherwise,
3. μ is updated by estimating μ^*, for which two possibilities are considered:
   - U_{KT}. Solving system

$$\nabla f(\widetilde{x}) + \nabla c(\widetilde{x})\mu + A^t\pi + \lambda = 0, \tag{13}$$

(which could be solved through least squares using QR factorization as justified in [7], obtaining $\widetilde{\mu}$, and also $\widetilde{\pi}$).

  - U_{HP}. Setting

$$\widetilde{\mu} = \mu + \rho c(\widetilde{x}), \tag{14}$$

    as established in [2, 3].

4. Should the solution of **ES** not reduce $\|c(x)\|$ sufficiently, ρ would be updated as $\rho = \nu\rho$, where $\nu > 1$.
5. Make $\mu = \widetilde{\mu}$ and $x_0 = \widetilde{x}$, and return to 1.

The issue is now the compatibility of system (13) at $\widetilde{x}$, and in this event the relationship between the procedures U_{KT} and U_{HP} to estimate vector μ^* at that point.

Each time subproblem **ES** is solved exactly by Murtagh and Saunders's active set method [13], we obtain an optimizer $\widetilde{x}$ and an associated partition of matrix $A = [B_A \mid S_A \mid N_A]$, and, thus, the variable reduction matrix Z_A shown below:

$$Z_A = \begin{bmatrix} -B_A^{-1} S_A \\ \mathbb{1} \\ \mathbb{0} \end{bmatrix}, \quad \text{which satisfies} \quad AZ_A = 0. \tag{15}$$

Since $\widetilde{x}$ is an optimizer of problem **ES** the necessary first-order optimality conditions must hold; i.e., there exist unique vectors $\widetilde{\pi}$ and $\widetilde{\lambda}$ such that:

$$\nabla_x L_\rho(\widetilde{x}, \mu) + A^t \widetilde{\pi} + \widetilde{\lambda} = 0 \tag{16}$$

$$A\widetilde{x} = b, \tag{17}$$

$$\widetilde{x}_i = l_i, \quad i = t+1, \cdots, \bar{t} \tag{18}$$

$$\widetilde{x}_i = u_i, \quad i = \bar{t}+1, \cdots, n \tag{19}$$

where t is the number of basic and superbasic variables, and

$$\begin{aligned} \widetilde{\lambda}_i &\le 0, \quad \text{if} \quad \widetilde{x}_i = l_i \\ \widetilde{\lambda}_i &\ge 0, \quad \text{if} \quad \widetilde{x}_i = u_i \\ \widetilde{\lambda}_i &= 0, \qquad \text{otherwise.} \end{aligned}$$

Expression (16) is equivalent to

$$A^t \widetilde{\pi} + \nabla c(\widetilde{x})[\mu + \rho c(\widetilde{x})] + \widetilde{\lambda} = -\nabla f(\widetilde{x}), \tag{20}$$

which in matrix form yields

$$\begin{bmatrix} B_A^t & \nabla_{B_A} c(\widetilde{x}) & \mathbb{0} \\ S_A^t & \nabla_{S_A} c(\widetilde{x}) & \mathbb{0} \\ N_A^t & \nabla_{N_A} c(\widetilde{x}) & \mathbb{1} \end{bmatrix} \begin{bmatrix} \widetilde{\pi} \\ \mu + \rho c(\widetilde{x}) \\ \widetilde{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla_{B_A} f(\widetilde{x}) \\ \nabla_{S_A} f(\widetilde{x}) \\ \nabla_{N_A} f(\widetilde{x}) \end{bmatrix},$$

where $\nabla_{B_A} c(\widetilde{x})$ stands for the rows of $\nabla c(\widetilde{x})$ associated with the rows of B_A^t. Similarly $\nabla_{S_A} c(\widetilde{x})$ and $\nabla_{N_A} c(\widetilde{x})$, and also the partition of $\nabla f(\widetilde{x})$, are defined.

Let us assume that matrix

$$A(\widetilde{x}) = \underbrace{\begin{array}{|c|c|c|} \hline \overbrace{B_A}^{m} & \overbrace{S_A}^{s} & N_A \\ \hline \nabla_{B_A} c(\widetilde{x})^t & \nabla_{S_A} c(\widetilde{x})^t & \nabla_{N_A} c(\widetilde{x})^t \\ \hline \end{array}}_{n} \begin{array}{l} \}m \\ \}r \end{array}$$

has full row rank. It is now possible to get from $A(\widetilde{x})$ a full rank basic matrix B such that it contains matrix B_A, which was obtained at the end of the first step of Algorithm 2.1. Once the basic matrix B has been defined, matrix S can be established such that $[B \mid S]$ contains B_A and S_A as submatrices, thus:
(21)

$$B = \overbrace{\begin{array}{|c|c|c|} \hline B_A & S'_A & \widetilde{N}_A \\ \hline \underbrace{\nabla_{B_A} c(\widetilde{x})^t}_{m} & \underbrace{\nabla_{S'_A} c(\widetilde{x})^t}_{s'} & \underbrace{\nabla_{\widetilde{N}_A} c(\widetilde{x})^t}_{m_1} \\ \hline \end{array}}^{m+r} \begin{array}{l} \}m \\ \}r \end{array} \quad \text{and} \quad S = \begin{array}{|c|} \hline \overbrace{S''_A}^{s-s'} \\ \hline \underbrace{\nabla_{S''_A} c(\widetilde{x})^t}_{s''} \\ \hline \end{array} ,$$

with $S_A = [S'_A \mid S''_A]$. The rest of columns from $A(\widetilde{x})$ makes up submatrix N. Let $\nabla_{\mathcal{B}} v(x)$ denote the gradient with respect to the variables associated with B of any differentiable function $v(x)$. Similarly $\nabla_{\mathcal{S}} v(x)$ and $\nabla_{\mathcal{N}} v(x)$. See [10] for an efficient procedure to build up B from data available at $\widetilde{x}$ when subproblem **ES** has been solved.

From this partition $[B \quad S \quad N]$ of $A(\widetilde{x})$ we have the variable reduction matrix

$$Z(\widetilde{x}) = \begin{bmatrix} -B^{-1}S \\ \mathbb{1} \\ \mathbb{0} \end{bmatrix}, \quad \text{which satisfies} \quad A(\widetilde{x})Z(\widetilde{x}) = 0.$$

Premultiplying by $Z(\widetilde{x})$ both sides of (20) we get the equivalent expression

(22) $$Z(\widetilde{x})^t A^t \widetilde{\pi} + Z(\widetilde{x})^t \nabla c(\widetilde{x})[\mu + \rho c(\widetilde{x})] + Z(\widetilde{x})^t \widetilde{\lambda} = -Z(\widetilde{x})^t \nabla f(\widetilde{x}).$$

According to the definition of $Z(\widetilde{x})$ the following must hold:

$$Z(\widetilde{x})^t \nabla c(\widetilde{x}) = 0 \quad \text{and} \quad Z(\widetilde{x})^t A^t = 0. \tag{23}$$

Furthermore,

$$Z(\widetilde{x})^t \widetilde{\lambda} = \begin{bmatrix} -(B^{-1}S)^t & \mathbb{1} & \mathbb{0} \end{bmatrix} \begin{bmatrix} \widetilde{\lambda}^{\mathcal{B}} \\ \widetilde{\lambda}^{\mathcal{S}} \\ \widetilde{\lambda}^{\mathcal{N}} \end{bmatrix} = -(B^{-1}S)^t \widetilde{\lambda}^{\mathcal{B}} + \widetilde{\lambda}^{\mathcal{S}}. \tag{24}$$

As we can see, $Z(\widetilde{x})^t \nabla f(\widetilde{x})$ will vanish only if the following condition holds

$$\widetilde{\lambda}^{\mathcal{S}} = (B^{-1}S)^t \widetilde{\lambda}^{\mathcal{B}}, \tag{25}$$

which in general is not satisfied, as can be easily proved; see [11].

Let $\sigma_k = [\sigma_{1k}, \cdots, \sigma_{\overline{m}k}]^t$ (with $\overline{m} = m + r$) be the column of $B^{-1}S$ associated with the superbasic variable x_k; then (25) can be recast as

$$\widetilde{\lambda}_k^{\mathcal{S}} = \sum_{j=1}^{\overline{m}} \sigma_{jk} \widetilde{\lambda}_j^{\mathcal{B}}, \quad k \in \mathcal{S},$$

$\mathcal{S}$ being the set of indices associated with the columns of S.

The *basic equivalent path* β_k *of superbasic variable* x_k is defined as the set of basic variables x_l that have a nonzero entry σ_{lk} in the column of $B^{-1}S$ corresponding to variable x_k.

Taking into account the expressions (22)-(24) we get

$$Z(\widetilde{x})^t \nabla f(\widetilde{x}) = \alpha, \tag{26}$$

where α is a vector (whose dimension is the number of columns of S) such that

$$\alpha_k = \sum_{i \in \beta_k \cap \widetilde{N}} \sigma_{ik} \widetilde{\lambda}_i^{\mathcal{B}} - \widetilde{\lambda}_k^{\mathcal{S}}, \tag{27}$$

$\widetilde{N}$ being the index set of the columns of matrix

$$\begin{array}{|c|} \hline \widetilde{N}_A \\ \hline \nabla_{\widetilde{N}} c(\widetilde{x})^t \\ \hline \end{array} \tag{28}$$

selected to make up the basis matrix of $A(\widetilde{x})$ (see expression (21)) and σ_{ik}, as previously defined, entry (i, k) of matrix $B^{-1}S$. Furthermore, since by construction the set of indices associated with the columns of S is a subset of the set of indices associated with the columns of S_A, $\widetilde{\lambda}_k = 0$ holds for all k corresponding to a column of S, hence (27) becomes

$$\alpha_k = \sum_{i \in \beta_k \cap \widetilde{N}} \sigma_{ik} \widetilde{\lambda}_i^{\mathcal{B}}. \tag{29}$$

It must be pointed out that vector α turns out to be the nonzero part of the residual vector corresponding to system (13), when, once the partition of matrix $A(\widetilde{x})$ is fixed, it is solved calculating first (π, μ) through the solution of the compatible system

$$B^t \begin{bmatrix} \pi \\ \mu \end{bmatrix} = -\nabla_{\mathcal{B}} f(\widetilde{x}),$$

and then computing:

$$\lambda^{\mathcal{N}} = -N^t \begin{bmatrix} \pi \\ \mu \end{bmatrix} - \nabla_{\mathcal{N}} f(\widetilde{x}),$$

as, by definition of $Z(\widetilde{x})$ and in view of (13) we are led to

$$\begin{aligned} Z(\widetilde{x})^t \nabla f(\widetilde{x}) &= -(B^{-1}S)^t \nabla_{\mathcal{B}} f(\widetilde{x}) + \nabla_{\mathcal{S}} f(\widetilde{x}) \\ &= S^t[-(B^t)^{-1} \nabla_{\mathcal{B}} f(\widetilde{x})] + \nabla_{\mathcal{S}} f(\widetilde{x}) \\ &= S^t \begin{bmatrix} \pi \\ \mu \end{bmatrix} + \nabla_{\mathcal{S}} f(\widetilde{x}). \end{aligned} \tag{30}$$

Here a series of propositions are presented to be used later.

Let us consider now, for any x, a full-row-rank matrix

$$A(x) = \begin{bmatrix} A \\ \nabla c(x)^t \end{bmatrix}$$

partitioned as $A(x) = [B \quad S \quad N]$, where B is a nonsingular matrix, and S and N have at least one column. Let

$$\widehat{A}(x) = \begin{bmatrix} B & S & N \\ \mathbb{0} & \mathbb{0} & \mathbb{1} \end{bmatrix}, \tag{31}$$

and

$$Z(x) = \begin{bmatrix} -B^{-1}S \\ \mathbb{1} \\ \mathbb{0} \end{bmatrix}, \quad \text{which satisfies} \quad \widehat{A}(x)Z(x) = 0.$$

Let $x = \overline{x}$ be a vector such that system

$$A^t\pi + \nabla c(\overline{x})\mu + \lambda + \nabla f(\overline{x}) = 0 \tag{32}$$

is compatible. Suppose that $\lambda_i = 0$ for all i associated with a column of either B or S. Then, premultiplying this system by $Z(\overline{x})^t$ we have

$$Z(\overline{x})^t A^t\pi + Z(\overline{x})^t\nabla c(\overline{x})\mu + Z(\widetilde{x})^t\lambda + Z(\overline{x})^t\nabla f(\overline{x}) = 0,$$

which, since $\widehat{A}(\overline{x})Z(\overline{x}) = 0$, implies $Z(\overline{x})^t\nabla f(\overline{x}) = 0$.

Proposition 2.1. *System (32) is compatible if and only if*

$$Z(\overline{x})^t\nabla f(\overline{x}) = 0 \tag{33}$$

is verified.

Proof: The previous result to this proposition proves its first part.

To prove the second part we consider the nonsingular $(n \times n)$-matrix

$$W = \begin{bmatrix} \mathbb{1} & \mathbb{0} & \mathbb{0} \\ -(B^{-1}S)^t & \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{1} \end{bmatrix},$$

where $\mathbb{0}$ stands for a zero matrix of suitable dimensions, and such that the unit matrix at the bottom is $(n-t) \times (n-t)$, $(n-t)$ being the number of columns of N.

Let us consider now that system (32) is not compatible. Hence, the matrix $[\widehat{A}(\overline{x})^t \quad \nabla f(\overline{x})]$ of this system has full column rank and we have

$$W[\widehat{A}(\overline{x})^t \quad \nabla f(\overline{x})] = W\begin{bmatrix} B^t & \mathbb{0} & \nabla_{\mathcal{B}} f(\overline{x}) \\ S^t & \mathbb{0} & \nabla_{\mathcal{S}} f(\overline{x}) \\ N^t & \mathbb{1} & \nabla_{\mathcal{N}} f(\overline{x}) \end{bmatrix} = \begin{bmatrix} B^t & \mathbb{0} & \nabla_{\mathcal{B}} f(\overline{x}) \\ \mathbb{0} & \mathbb{0} & Z(\overline{x})^t\nabla f(\overline{x}) \\ N^t & \mathbb{1} & \nabla_{\mathcal{N}} f(\overline{x}) \end{bmatrix},$$

where $\nabla_{\mathcal{M}} f(\overline{x})$ is the gradient of $f(x)$ at $x = \overline{x}$ with respect to the subset of variables associated with M, for all submatrix M formed by columns in $A(\overline{x})$.

Since the product of a nonsingular matrix of order $n \times n$ multiplied by a matrix with full column rank of order $n \times (m + r + (n-t) + 1)$ gives rise to a matrix with these same characteristics, it is shown that the product $Z(\overline{x})^t\nabla f(\overline{x})$ cannot be the null vector.

Therefore, if (33) holds, system (32) is compatible.

■

Let B_A be a nonsingular square submatrix of B made up by columns of A and let S_A be the submatrix formed by the columns common to $[B \quad S]$ and A once removed the columns of B_A. Moreover, N_A is a submatrix constituted by the rest of columns of A, thus their associated variables are the same as those associated with the columns of N.

Let

$$BS = \begin{bmatrix} B_A & S_A \\ \nabla_{B_A} c(\overline{x})^t & \nabla_{S_A} c(\overline{x})^t \end{bmatrix} \tag{34}$$

and Z_A a matrix defined by expression (15) (although using the current B_A and S_A).

Proposition 2.2. *Matrix BS has full row rank if and only if $\nabla c(\overline{x})^t Z_A$ has also full row rank.*

Proof: It is sufficient to take into account the matrix product

$$\begin{bmatrix} B_A & S_A \\ \nabla_{B_A} c(\overline{x})^t & \nabla_{S_A} c(\overline{x})^t \end{bmatrix} \begin{bmatrix} \mathbb{1} & -B_A^{-1} S_A \\ \mathbb{0} & \mathbb{1} \end{bmatrix} = \begin{bmatrix} B_A & \mathbb{0} \\ \nabla_{B_A} c(\overline{x})^t & \nabla c(\overline{x})^t Z_A \end{bmatrix},$$

where the first matrix is BS, see (34), and Z_A is defined by (15).

■

Premultiplying equation (32) by matrix Z_A^t and moving $Z_A^t \nabla f(\overline{x})$ to the right hand side we obtain

$$Z_A^t \nabla c(\overline{x}) \mu = -Z_A^t \nabla f(\overline{x}). \tag{35}$$

Proposition 2.3. *Let BS be a full-row-rank matrix. System (32) is compatible if and only if system (35) is compatible*

Proof: To prove the «only if» part it is enough to premultiply the system (32) by Z_A.

Now, to show that the «if» part holds, let us consider the matrix

$$W_A = \begin{bmatrix} \mathbb{1} & \mathbb{0} & \mathbb{0} \\ -(B_A^{-1} S_A)^t & \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{1} \end{bmatrix}.$$

Suppose that (32) is not compatible, then, since BS has full row rank, matrix $[\widehat{A}(\overline{x})^t \quad \nabla f(\overline{x})]$ has full column rank. Therefore, to prove this part it is sufficient to operate as in proposition 2.1, but now replacing the matrix W with the matrix W_A, thus

$$W_A[\widehat{A}(\overline{x})^t \quad \nabla f(\overline{x})] = W_A \begin{bmatrix} B_A^t & \nabla_{B_A} c(\overline{x}) & \mathbb{0} & \nabla_{B_A} f(\overline{x}) \\ S_A^t & \nabla_{S_A} c(\overline{x}) & \mathbb{0} & \nabla_{S_A} f(\overline{x}) \\ N_A^t & \nabla_{N_A} c(\overline{x}) & \mathbb{1} & \nabla_{N_A} f(\overline{x}) \end{bmatrix}$$

$$= \begin{bmatrix} B_A^t & \nabla_{B_A} c(\overline{x}) & \mathbb{0} & \nabla_{B_A} f(\overline{x}) \\ \mathbb{0} & Z_A^t \nabla c(\overline{x}) & \mathbb{0} & Z_A^t \nabla f(\overline{x}) \\ N_A^t & \nabla_{N_A} c(\overline{x}) & \mathbb{1} & \nabla_{N_A} f(\overline{x}) \end{bmatrix}.$$

Note that if matrix $[\widehat{A}(\overline{x})^t \quad \nabla f(\overline{x})]$ has full column rank, $[Z_A^t \nabla c(\overline{x}) \quad Z_A^t \nabla f(\overline{x})]$ has also full column rank.

■

Corollary 2.1. *Under the same conditions of the previous propositions, system (35) is compatible if and only if (33) holds.*

Proof: It is a direct result of the propositions 2.1 and 2.3.

■

Now we consider $\overline{x} = \widetilde{x} = x(\mu, \rho)$ (i.e. the optimizer of **ES** considered at the beginning of this section).

As a result of these propositions and corollary, we have the following consequences:

- If $\alpha \neq 0$ (see expression (26)), system (13) is not compatible, and neither is system

  (36) $$Z_A^t \nabla c(\widetilde{x}) \mu = -Z_A^t \nabla f(\widetilde{x}).$$

  Therefore, in this case the estimate of μ^* by the U_{KT} procedure is not reliable, because system (13) is not compatible.

- If when building up the basic matrix B of $A(\widetilde{x})$ it is not necessary to take columns of $A(\widetilde{x})$ that contain columns of N_A, we have

  (37) $$Z(\widetilde{x})^t \nabla f(\widetilde{x}) = 0,$$

and the associated systems (13) and (36) are compatible. Moreover, the μ obtained by solving both (13) and (36) will be the same, see proposition 2.2 and corollary 2.1. Therefore, it is enough to solve (36) to calculate μ in procedure U_{KT}.

- If problem **EP** has not simple bounds, i.e., if it is only defined by (1-3), then it automatically holds $\alpha = 0$ for all $\widetilde{x} = x(\mu, \rho)$ optimizing

$$\begin{array}{ll} \underset{x}{\text{minimize}} & L_\rho(x,\mu) \\ \text{subject to:} & Ax = b, \end{array}$$

  where $L_\rho(x,\mu) = f(x) + \mu^t c(x) + \frac{\rho}{2}\|c(x)\|_2^2$ is the augmented Lagrangian function. Therefore, for all vector $\widetilde{x}$ obtained in this way, it is enough to solve (36) — with $\overline{x} = \widetilde{x}$ — to calculate μ in procedure U_{KT}.

- It is clear from the above results that in order to build up matrix B it is preferable to use only columns of BS. Should this submatrix of $A(\widetilde{x})$ not have full row rank, one must initially search for suitable columns among the l columns associated with N_A such that $\widetilde{\lambda}_l = 0$, for $l \in \widetilde{N}$, if any, see (29).

Next, by means of a proposition we analyze the general case considered at the beginning of this section.

Proposition 2.4. *If $\alpha = 0$ and* **ES** *(5–7) is solved* ***with exact minimization****, the U_{KT} procedure is valid for estimating μ^*. Furthermore, when B is built up following the rule given at the last item, the results obtained coincide with those found by means of the U_{HP} procedure. Otherwise, even if $\alpha = 0$, the estimates of μ^* obtained through U_{KT} and U_{HP} are not the same.*

Proof: As shown above, $\alpha = 0$ implies the validity of procedure U_{KT} for estimating μ^*, because of (26) and Proposition 2.1.

The second part of this proposition is proved next.

The solution of **ES** by means of exact minimization implies that $Z_A^t \nabla_x L_\rho(\widetilde{x}, \mu) = 0$. This is equivalent to

$$Z_A^t \nabla c(\widetilde{x})(\mu + \rho c(\widetilde{x})) = -Z_A^t \nabla f(\widetilde{x}). \tag{38}$$

If the set $\overline{BS}$ of indices associated with $[B \quad S]$ coincides with the set $\overline{BS}_A$ of the indices associated with $[B_A \quad S_A]$, $\alpha = 0$ and the conditions of Propositions 2.1, 2.2 and 2.3 are fulfilled directly, and therefore the system

$$Z_A^t \nabla c(\widetilde{x})\mu = -Z_A^t \nabla f(\widetilde{x}) \tag{39}$$

is compatible and has a single solution $\overline{\mu}$, which compared with (38) implies that

$$\overline{\mu} = \mu + \rho c(\widetilde{x}),$$

whose second member corresponds to the estimate of μ^* by means of the U_{HP} procedure.

Due to the length of the part of the proof corresponding to the case in which $\overline{BS}$ contains strictly $\overline{BS}_A$, it is divided into several sections.

(i) Definition of matrix $\widetilde{N}_A$ to fill up the row rank of $[B \quad S]$.

Let us consider again the submatrix given in (28)

$$\begin{array}{|c|}\hline \widetilde{N}_A \\ \hline \nabla_{\widetilde{N}} c(\widetilde{x})^t \\ \hline \end{array},$$

whose columns are selected among the columns of $A(\widetilde{x})$ associated with nonbasic variables (with regard to subproblem **ES**) so that the matrix B of expression (21) is nonsingular. Let $\widetilde{N}$ be, as above, the set of indices associated with the columns of $\widetilde{N}_A$. Hence $\overline{BS}_A \cup \widetilde{N} = \overline{BS}$. From now on, we consider that the columns of this submatrix appear arranged in A immediately after those of S_A, without loss of generality.

(ii) Definition of matrix $\overline{Z}_A$.

A new reduction matrix $\overline{Z}_A$ is also defined for matrix A such that

$$\overline{Z}_A = [Z_A \quad Z_{\widetilde{N}}] = \begin{bmatrix} -B_A^{-1}S_A & -B_A^{-1}\widetilde{N}_A \\ \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{1} \\ \mathbb{0} & \mathbb{0} \end{bmatrix}. \tag{40}$$

The matrices $\mathbb{1}$, from left to right, have their dimensions fixed, respectively, by the number of columns in S_A and the number of columns in $\widetilde{N}_A$. The matrix $\overline{Z}_A$ has full column rank.

(iii) Compatibility of system $\overline{Z}_A^t \nabla c(\widetilde{x})\mu = -\overline{Z}_A^t \nabla f(\widetilde{x})$.

Since $\alpha = 0$, system

$$A^t\pi + \nabla c(\widetilde{x})\mu + \lambda = -\nabla f(\widetilde{x}),$$

has one only solution $(\overline{\pi}, \overline{\mu}, \overline{\lambda})$ (see Proposition 2.1). Moreover, by Proposition 2.3, the following system is compatible:

$$\overline{Z}_A^t \nabla c(\widetilde{x})\mu = -\overline{Z}_A^t \nabla f(\widetilde{x}). \tag{41}$$

The solution of this is $\overline{\mu}$, which is unique due to matrix $\overline{Z}_A^t \nabla c(\widetilde{x})$ having full column rank. To prove this last, it is sufficient to take into account that

$$[B \quad S] = \begin{bmatrix} B_A & S_A & \widetilde{N}_A \\ \nabla_B c(\widetilde{x})^t & \nabla_S c(\widetilde{x})^t & \nabla_{\widetilde{N}} c(\widetilde{x})^t \end{bmatrix}, \tag{42}$$

is of full rank $(m + r)$ and to construct a suitable matrix W. Let W be

$$W = \begin{bmatrix} \mathbb{1} & -B_A^{-1} S_A & -B_A^{-1} \widetilde{N}_A \\ \mathbb{0} & \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{1} \end{bmatrix}.$$

This matrix is nonsingular and of order $(m + s + m_1) \times (m + s + m_1)$, being m_1 the number of columns in $\widetilde{N}_A$ and such that $s + m_1 \geq r$, by construction of $[B \quad S]$ (42).

Multiplying and taking into account (40)

$$[B \quad S]W = \begin{bmatrix} B_A & \mathbb{0} & \mathbb{0} \\ \nabla_B c(\widetilde{x})^t & \nabla c(\widetilde{x})^t Z_A & \nabla c(\widetilde{x})^t Z_{\widetilde{N}} \end{bmatrix} = \begin{bmatrix} B_A & \mathbb{0} \\ \nabla_B c(\widetilde{x})^t & \nabla c(\widetilde{x})^t \overline{Z}_A \end{bmatrix},$$

then because of the features of the factor matrices with respect to the rank, the product is a $(m+r) \times (m+s+m_1)$-matrix that has full row rank, and since B_A is a nonsingular matrix with rank m, the submatrix $\nabla c(\widetilde{x})^t \overline{Z}_A$ has rank r, or in other words, it has full row rank.

(iv) Calculation of $\overline{Z}_A^t \nabla_x L_\rho(\widetilde{x}, \mu)$.

Let us return to the subproblem **ES** and perform the product $\overline{Z}_A^t \nabla_x L_\rho(\widetilde{x}, \mu)$, the result is

$$\overline{Z}_A^t \nabla_x L_\rho(\widetilde{x}, \mu) = \begin{bmatrix} Z_A^t \\ Z_{\widetilde{N}}^t \end{bmatrix} \nabla_x L_\rho(\widetilde{x}, \mu) = \begin{bmatrix} Z_A^t \nabla_x L_\rho(\widetilde{x}, \mu) \\ Z_{\widetilde{N}}^t \nabla_x L_\rho(\widetilde{x}, \mu) \end{bmatrix}.$$

$Z_A^t \nabla_x L_\rho(\widetilde{x}, \mu)$ is null because **ES** is solved by means of exact minimization. Besides,

$$Z_{\widetilde{N}}^t \nabla_x L_\rho(\widetilde{x}, \mu) + Z_{\widetilde{N}}^t A^t \widetilde{\pi} + Z_{\widetilde{N}}^t \widetilde{\lambda} = 0.$$

Let us also observe that the second term on the left is zero, as $Z_{\widetilde{N}}$ expands a null subspace of A and

$$Z^t_{\widetilde{N}}\widetilde{\lambda} = (-B_A^{-1}\widetilde{N}_A)^t\widetilde{\lambda}^B + \widetilde{\lambda}^{\widetilde{N}} = \widetilde{\lambda}^{\widetilde{N}},$$

since $\widetilde{\lambda}^B = 0$ — which is associated with B_A. Thus,

$$\overline{Z}^t_A \nabla_x L_\rho(\widetilde{x}, \mu) = \begin{bmatrix} \mathbb{0} \\ -\widetilde{\lambda}^{\widetilde{N}} \end{bmatrix},$$

which, by developing $\nabla_x L_\rho(\widetilde{x}, \mu)$, is equivalent to

$$\overline{Z}^t_A \nabla c(\widetilde{x})(\mu + \rho c(\widetilde{x})) = -\overline{Z}^t_A \nabla f(\widetilde{x}) + \begin{bmatrix} \mathbb{0} \\ -\widetilde{\lambda}^{\widetilde{N}} \end{bmatrix}. \tag{43}$$

If $\widetilde{\lambda}^{\widetilde{N}}$ is null — i.e., if following the rule put forward in the item previous to this proposition we are able to make up a basis matrix B —, then we have

$$\overline{Z}^t_A \nabla c(\widetilde{x})(\mu + \rho c(\widetilde{x})) = -\overline{Z}^t_A \nabla f(\widetilde{x}). \tag{44}$$

(v) Comparison of expressions (41) and (44).

Finally, if we compare expressions (41) and (44), taking into account the compatibility of system (41) and the uniqueness of its solution, the conclusion reached is that $\overline{\mu} = \mu + \rho c(\widetilde{x})$ is fulfilled if $\widetilde{\lambda}^{\widetilde{N}}$ is null. In this case the procedures U_{KT} and U_{HP} produce the same estimate of μ^*.

Note that if $\widetilde{\lambda}^{\widetilde{N}}$ is not null and $\alpha = 0$ with exact minimization, the U_{KT} procedure is reliable (in the sense of that the residuals of both systems (13) and (41) are null), although the estimate of $\overline{\mu}$ that it provides is different from that given by U_{HP}, see (43). ■

In reference [10] there is also an efficient procedure for computing α, and practicalities related to the implementation of Algorithm 2.1 with the problem considered here (code PFNRN [12]) and computational results.

3. EXTENSION BY USING VECTOR y

In this section the above results are extended to the case of problems with general inequality contraints (problem **IP** (9–12)) by using a vector y, taking into account the technique put forward by Conn *et al* in [5], which is also employed by Murtagh and Saunders in [14].

Through this section and the following we add to assumption **AS1**, given in §2, the new one:

AS2. μ^* satisfies the strict complementarity condition

$$\begin{aligned} \text{if } c_j(x^*) = \underline{c}_j \quad &\Rightarrow \quad \mu_j^* < 0, \\ \text{if } c_j(x^*) = \overline{c}_j \quad &\Rightarrow \quad \mu_j^* > 0, \\ &\text{otherwise } \mu_j^* = 0. \end{aligned}$$

Solving problem **IP** is equivalent to solving:

$$\begin{aligned}
&(45) \qquad \text{minimize} \;\; f(x) \\
&(46) \qquad \text{subject to:} \;\; Ax = b \\
&(47) \qquad c(x) - y = 0 \qquad \textbf{(EPy)} \\
&(48) \qquad l \le x \le u \\
&(49) \qquad \underline{c} \le y \le \overline{c},
\end{aligned}$$

where y represents the slacks vector that turns the inequalities (11) into equalities.

As regards problem **EP** (1–4), the difference between this and problem **EPy** is that in the latter there are constraints (47) and (49) instead of constraints (3). Therefore we must analyze the effect of replacing (3) with (47) and (49) in the results of the former Section to see whether they can still be applied to the current problem.

First, it can be observed that the Kuhn-Tucker conditions of an optimal solution (x^*, y^*) to problem **EPy** are

$$\nabla f(x^*) + \nabla c(x^*)\mu + A^t\pi + \lambda = 0 \tag{50}$$

$$-\mu + \gamma = 0 \tag{51}$$

$$Ax^* = b \tag{52}$$

$$c(x^*) - y^* = 0, \tag{53}$$

such that unique vectors μ^*, π^*, λ^* and γ^* exist that satisfy equations (50) and (51), and with γ^* also satisfying

$$\begin{aligned} &\gamma_i^* \le 0 \;\text{ if }\; y_i^* = \underline{c}_i \\ &\gamma_i^* \ge 0 \;\text{ if }\; y_i^* = \overline{c}_i \\ &\gamma_i^* = 0 \quad\; \text{otherwise,} \end{aligned} \tag{54}$$

and λ^*

$$\begin{array}{ll} \lambda_i^* \leq 0 \text{ if } x_i^* = l_i \\ \lambda_i^* \geq 0 \text{ if } x_i^* = u_i \\ \lambda_i^* = 0 \quad \text{otherwise.} \end{array} \tag{55}$$

Note that the only consequence on the multipliers of the introduction of the slacks y are expressions (51) and (54), the first being used to determine γ^* once the rest of the variables have been found through (50). Therefore, in order to obtain a first-order estimate of all multipliers at point (x, y) it suffices, as pointed out in [7], to solve

$$\nabla f(x) + \nabla c(x)\mu + A^t\pi + \lambda = 0, \tag{56}$$

just as in the case of problem **EP**, obtaining γ afterwards through (51). Furthermore, expression (13) does not change — let us compare it with (56).

Here the associated subproblem is defined as

$$\begin{array}{lll} (57) & \underset{x,y}{\text{minimize}} & L_\rho(x, y, \mu) \\ (58) & \text{subject to:} & Ax = b \\ (59) & & l \leq x \leq u \\ (60) & & \underline{c} \leq y \leq \overline{c} \end{array} \qquad \textbf{(ESy)}$$

where

$$L_\rho(x, y, \mu) = f(x) + \mu^t[c(x) - y] + \frac{1}{2}\rho\|c(x) - y\|_2^2 \tag{61}$$

is the augmented Lagrangian function. Furthermore, in Algorithm 2.1, subproblem **ES** (5–7) is replaced by subproblem **ESy** and $c(x)$ with $c(x) - y$, hence the U_{HP} type estimate at the optimizer $(\widetilde{x}, \widetilde{y})$, obtained from the solution of **ESy**, can be written as:

$$\widetilde{\mu} = \mu + \rho[c(\widetilde{x}) - \widetilde{y}], \tag{62}$$

and the U_{KT} type estimate is obtained solving

$$\nabla f(\widetilde{x}) + \nabla c(\widetilde{x})\mu + A^t\pi + \lambda = 0,$$

as in §2 with (13). Thus, as before in §2, we can now define matrices $A(\widetilde{x})$, $Z(\widetilde{x})$ and Z_A, arriving at the results analogous to those obtained in §2, by propositions 2.1-2.3 and corollary 2.1.

Other modifications of the algorithm are associated with the substitution of x with x, y. From the Kuhn-Tucker conditions for this subproblem we obtain

$$\nabla f(\widetilde{x}) + \nabla c(\widetilde{x})\{\mu + \rho[c(\widetilde{x}) - \widetilde{y}]\} + A^t\pi + \lambda = 0 \tag{63}$$

$$-\{\mu + \rho[c(\widetilde{x}) - \widetilde{y}]\} + \gamma = 0 \tag{64}$$

such that vectors $\pi = \widetilde{\pi}$ and $\lambda = \widetilde{\lambda}$ exist satisfying the first equation and a vector $\widetilde{\gamma}$ can be obtained through the second one. As in §2 for expresion (20), here (63) is the key expression for studying the compatibility of estimate U_{KT} when using the active set techniques of Murtagh and Saunders [13] to solve **ESy**, obtaining the results equivalent to those of the proposition 2.4.

In conclusion, all the analysis after expression (20) in §2 is also applicable to this case, with the only exception that expression $\mu + \rho c(\widetilde{x})$ must be replaced by $\mu + \rho[c(\widetilde{x}) - \widetilde{y}]$.

4. EXTENSION BY USING VECTORS y AND w

Here the results of §2 are extended to the case of problems with general inequality contraints by using vectors y and w.

This section describes an alternative way of dealing with problem **IP** (9–12). Solving problem **IP**, as put forward in [9] (using slack variables raised to the square, as done by Rockafellar in [16]), is equivalent to solving problem:

$$\begin{array}{llr}
(65) & \text{minimize } f(x) & \\
(66) & \text{subject to: } Ax = b & \\
(67) & \quad (c(x) - \underline{c}) - y^2 = 0 & \textbf{(EPyw)} \\
(68) & \quad l \le x \le u & \\
(69) & \quad y^2 + w^2 = \overline{c} - \underline{c}, &
\end{array}$$

where $y, w \in \mathbb{R}^r$ are auxiliary vectors of free variables — whithout bounds — that through vectors

$$y^2 = \begin{bmatrix} y_1^2 \\ \vdots \\ y_r^2 \end{bmatrix} \quad \text{and} \quad w^2 = \begin{bmatrix} w_1^2 \\ \vdots \\ w_r^2 \end{bmatrix}$$

allow the transformation of inequality (11) into an equality.

As regards the reference problem **EP** (1–4), the difference with problem **EPyw** (65–69) is that in the latter instead of constraint (3) there are constraints (67) and (69), but contrary to what happens in Section 2, the number of bounds does not increase. Therefore, we must examine the effect of replacing (3) with (67) and (69) in the main calculation steps involved in the results of Section 2 and extend them to the type of problem now considered.

It must be noticed first that the Kuhn-Tucker conditions to be satisfied by an optimizer (x^*, y^*, w^*) of problem **EPyw** are

$$\nabla f(x^*) + \nabla c(x^*)\mu + A^t\pi + \lambda = 0 \tag{70}$$

$$(-\mu + \gamma)^t y^* = 0 \tag{71}$$

$$\gamma^t w^* = 0 \tag{72}$$

$$Ax^* = b \tag{73}$$

$$(c(x^*) - \underline{c}) - (y^*)^2 = 0, \tag{74}$$

$$(y^*)^2 + (w^*)^2 = \overline{c} - \underline{c}, \tag{75}$$

such that unique vectors μ^*, π^*, λ and γ^* exist that satisfy equations (70), (71) and (72), and λ^* such that

$$\begin{array}{l} \lambda_i^* \leq 0 \text{ if } x_i^* = l_i \\ \lambda_i^* \geq 0 \text{ if } x_i^* = u_i \\ \lambda_i^* = 0 \quad \text{ otherwise.} \end{array}$$

Note that the only consequence of the introduction of slacks y (apart from the constraints where they appear) are expressions (71) and (72), and γ^* can be obtained from these once the remainig multipliers have been computed from (70), given that in case $y^* \neq 0$, $\gamma^* = \mu^*$ (due to (71)), otherwise, $w^* \neq 0$ yields $\gamma^* = 0$ (due to (72)). Therefore, to obtain a first-order estimate of the U_{KT} type at point (x, y, w) of all multipliers it suffices to solve

$$\nabla f(x) + \nabla c(x)\mu + A^t\pi + \lambda = 0, \tag{76}$$

as in problem **EP** (1–4) and then compute γ through (71) and (72). Furthermore, expression (13), which coincides with (76), does not change.

Here the associated subproblem would be

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & \overline{L}_\rho(x, y, \mu) \\ \text{subject to:} & Ax = b \\ & l \leq x \leq u \\ & y^2 + w^2 = \overline{c} - \underline{c}, \end{array} \tag*{\textbf{(ESyw)}}$$

where

$$\overline{L}_\rho(x,y,\mu) = f(x) + \mu^t[(c(x)-\underline{c}) - y^2] + \frac{\rho}{2}\|(c(x)-\underline{c}) - y^2\|_2^2 \tag{77}$$

is the augmented Lagrangian function. Furthermore, see [9], the constraints where variables (y,w) appear can be eliminated by replacing the above augmented Lagrangian by

$$L_\rho(x,\mu) = f(x) + \sum_{j=1}^{r}\left\{\mu_j\varphi_j[c_j(x),\mu_j,\rho] + \frac{1}{2}\rho|\varphi_j[c_j(x),\mu,\rho]|^2\right\}, \tag{78}$$

where

$$\varphi_j[c_j(x),\mu_j,\rho] = \begin{cases} c_j(x)-\overline{c}_j & \text{if} \quad \mu_j+\rho[c_j(x)-\overline{c}_j] > 0 \\ c_j(x)-\underline{c}_j & \text{if} \quad \mu_j+\rho[c_j(x)-\underline{c}_j] < 0 \\ -\mu_j/\rho & \text{otherwise,} \end{cases}$$

the expression in braces that appears in (78) being continuously differentiable with respect to x, for f and c defined as in §1, and (bearing in mind the strict complementarity assumption **AS2** in §2) twice continuously differentiable with respect to x if $x \in \mathcal{X}$, where

$$\mathcal{X} = \left\{x \mid \mu_j+\rho[c_j(x)-\overline{c}_j] \neq 0, \mu_j+\rho[c_j(x)-\underline{c}_j] \neq 0, \forall j=1,\cdots,r\right\}.$$

Therefore, according to [9], solving problem **ESyw** is equivalent to solving problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & L_\rho(x,\mu) \\ \text{subject to:} \quad & Ax = b \\ & l \leq x \leq u. \end{aligned} \tag{ESx}$$

Moreover, in Algorithm 2.1, replacing subproblem **ES** (5–7) with **ESx** and $c_j(x)$ with $\varphi_j[c_j(x),\mu_j,\rho]$, for $j=1,\cdots,r$, we have that the U_{HP} type estimate at the optimizer $\widetilde{x}$, obtained through the solution of **ESx**, can be expressed by (see [2, 9])

$$\widetilde{\mu} = \mu + \rho\varphi[c(\widetilde{x}),\mu,\rho], \tag{79}$$

such that $\varphi \equiv [\varphi_1,\cdots,\varphi_r]^t$, and the U_{KT} type estimate is obtained solving

$$\nabla f(\widetilde{x}) + \nabla c(\widetilde{x})\mu + A^t\pi + \lambda = 0,$$

as in §2 with (13). Thus, as before in §2, we can now define matrices $A(\widetilde{x})$, $Z(\widetilde{x})$ and Z_A, arriving at results analogous to those obtained in §2, by propositions 2.1-2.3 and corollary 2.1.

In addition, from the Kuhn-Tucker conditions associated with subproblem **ESx** we obtain:

$$\nabla f(\widetilde{x}) + \nabla c(\widetilde{x}) \{\mu + \rho\varphi[c(\widetilde{x}), \mu, \rho]\} + A^t\pi + \lambda = 0, \tag{80}$$

where vectors $\pi = \widetilde{\pi}$ and $\lambda = \widetilde{\lambda}$ exist satisfying this equation.

Hence, from (80), by an analogous process to that followed from expression (20), we are led to the same results as are obtained in §2 when using the active set techniques of Murtagh and Saunders [13] to solve **ESx**, obtaining the equivalent results to those of the proposition 2.4.

In conclusion, all the analysis after expression (20) in §2 is also applicable to this case, with the only exception that expression $\mu+\rho c(\widetilde{x})$ must be replaced by $\mu+\rho\varphi[c(\widetilde{x}), \mu, \rho]$.

5. CONCLUSIONS

This work shows when one may compute under certain guarantees the firts-order multiplier estimate based on the Khun-Tucker conditions. In addition, it proves the equivalence between this type of first-order estimate and that obtained through the original multiplier method (Hestenes and Powell's method) when the exact minimization is used to solve the subproblem and it is not necessary to use columns of the constraint matrix that correspond to strongly active variables (with respect to the subproblem) to obtain a submatrix of the Jacobian (not including the simple bounds) having full row rank. Nevertheless, in practice not even in the latter case both procedures give the same estimate, as usually an inexact minimization of the subproblem is carried out. This work puts forward also a procedure to compute the multiplier estimates by solving a reduced system (41), instead of having to solve the large system (13), if the conditions so permit (see first lines of this paragraph).

In the previous two sections specific procedures for transforming problems of type **IP** (9–12) into problems of type **EP** (1–4) have been considered.

The procedure described in Section 4 as compared with that described in section 3 has the advantage that it does not increase subproblem size with respect to the original problem **IP**. Results of numerical tests comparing both procedures have not been included yet as, in our view, the construction of appropiate software to exploit the technique put forward by Conn *et al* in [5] $-$once fitted to the structure of problem **EPy** (45-49)$-$ is by no means trivial and its coding is still underway.

Using these procedures, inequalities in general constraints are eliminated. The validity for these problems of results and procedures put forward in [10] for type **EP** problems only has also been proved.

6. REFERENCES

[1] **Bertsekas D.P.** (1977). «Approximation procedures based on the method of multipliers». *J. Opt. Th. and Appl.*, **23**, 487–510.

[2] **Bertsekas D.P.** (1982). *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York.

[3] **Bertsekas D.P.** (1995). *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts.

[4] **Bertsekas D.P.** (1998). *Network Optimization: Continuous and Discrete Models.* Athena Scientific, Belmont, Massachusetts.

[5] **Conn, A.R., Gould, N.** and **Toint, Ph.L.** (1994). «A note on exploiting structure when using slack variables». *Mathematical Programming*, **67**, 89–97.

[6] **Conn, A.R., Gould, N., Saternaer, A.** and **Toint, Ph.L.** (1996). «Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints». *SIAM J. on Optimization*, **6, 3**, 674–703.

[7] **Gill, P.E., Murray, W.** and **Wright, M.H.** (1981). *Practical Optimization.* Academic Press, New York.

[8] **Hestenes, M.R.** (1969). «Multiplier and gradient methods». *J. Opt. Th. and Appl.*, **4**, 303-320.

[9] **Mijangos, E.** and **Nabona, N.** (1996). «The application of the multipliers method in nonlinear network flows with side constraints». *Technical Report 96/10*, Dept. of Statistics and Operations Research. Universitat Politècnica de Catalunya, 08028 Barcelona, Spain.

[10] **Mijangos, E.** and **Nabona, N.** (1997). «On the compatibility of classical first order multiplier estimates with variable reduction techniques.»*Technical Report 97/06*, Dept. of Statistics and Operations Research. Universitat Politècnica de Catalunya, 08028 Barcelona, Spain.

[11] **Mijangos, E.** (1996). «Optimización de flujos no lineales en redes con restricciones laterales mediante técnicas de multiplicadores». *Ph.D. Thesis, Statistics and Operations Research Dept.*. Universitat Politècnica de Catalunya, 08028 Barcelona, Spain.

[12] **Mijangos, E.** (1997). «PFNRN03 user's guide». *Technical Report 97/05*, Dept. of Statistics and Operations Research. Universitat Politècnica de Catalunya, 08028 Barcelona, Spain.

[13] **Murtagh, B.A.** and **Saunders, M.A.** (1978). «Large–scale linearly constrained optimization». *Mathematical Programming*, **14**, 41–72.

[14] **Murtagh, B.A.** and **Saunders, M.A.** (1983). *MINOS 5.0. User's guide*. Dept. of Operations Research, Standford University, CA 9430, USA.

[15] **Powell, M.J.D.** (1969). «A method for nonlinear constraints in minimization problems». *Optimization* (R. Fletcher, ed.), 283–298. Academic Press New, York.

[16] **Rockafellar, R.T.** (1971). «New applications of duality in convex programming». *Proc. Confer. Probab.*, 4th, Brasov, Romania, 73–81.