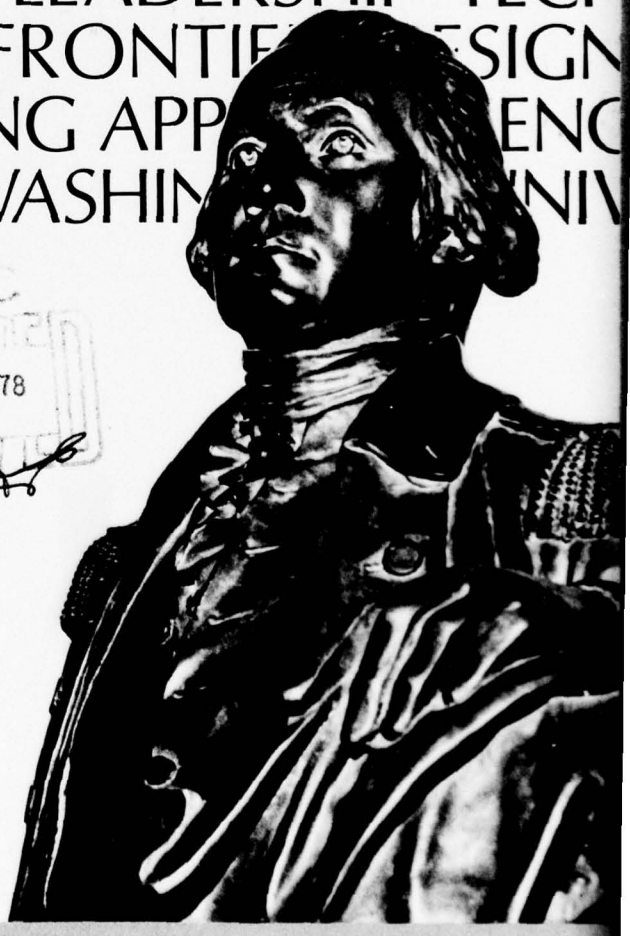
AD-A050 792 GEORGE WASHINGTON UNIV WASHINGTON D C INST FOR MANAG--ETC F/G 12/1
EXACT SENSITIVITY ANALYSIS USING AUGMENTED LAGRANGIANS.(U)
MAR 77 R L ARMACOST, A V FIACCO DAAG-29-76-G-0150
UNCLASSIFIED SERIAL-T-349 ARO-13622.3-M NL

| OF |
AD
A050 792

END
DATE
FILMED
4-78
DDC

THE
GEORGE
WASHINGTON
UNIVERSITY

STUDENTS FACULTY STUDY R
ESEARCH DEVELOPMENT FUT
URE CAREER CREATIVITY CO
MMUNITY LEADERSHIP TECH
NOLOGY FRONTIE ESIGN
ENGINEERING APP ENC
GEORGE WASHIN NIV



INSTITUTE FOR MANAGEMENT
SCIENCE AND ENGINEERING
SCHOOL OF ENGINEERING
AND APPLIED SCIENCE

(10)

EXACT SENSITIVITY ANALYSIS USING AUGMENTED LAGRANGIANS

by

Robert L. Armacost
Anthony V. Fiacco

Serial T-349
22 March 1977

The George Washington University
School of Engineering and Applied Science
Institute for Management Science and Engineering

U.S. Army Research Office - Durham
Contract Number DAAG29-76-G-0150

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

None

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: T-349 (14) SERIAL-T-349
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): (6) EXACT SENSITIVITY ANALYSIS USING AUGMENTED LAGRANGIANS,
5. TYPE OF REPORT & PERIOD COVERED: (9) Interim Technical rept.,
6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): (10) ROBERT L. ARMACOST, ANTHONY V. FIACCO
8. CONTRACT OR GRANT NUMBER(s): (15) DAAG-29-76-G-0150

9. PERFORMING ORGANIZATION NAME AND ADDRESS
THE GEORGE WASHINGTON UNIVERSITY
INSTITUTE FOR MANAGEMENT SCIENCE AND ENGINEERING
WASHINGTON, D.C. 20037

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. ARMY RESEARCH OFFICE
BOX 12211
RESEARCH TRIANGLE PARK, N.C. 27709

12. REPORT DATE: (11) 22 March 1977
13. NUMBER OF PAGES: 34 (12) 39 P.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
(18) ARO
(19) 13622.3-M

15. SECURITY CLASS. (of this report): None
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION OF THIS REPORT IS UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

NONLINEAR PROGRAMMING	AUGMENTED LAGRANGIAN
SENSITIVITY ANALYSIS	KUHN-TUCKER TRIPLE
PARAMETERS	OPTIMAL VALUE FUNCTION

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper reviews several recently developed sensitivity analysis results for a general class of parametric nonlinear programming problems and develops formulas for calculating the sensitivity measures, i.e., the partial derivatives of a Kuhn-Tucker triple and the optimal value function, using an augmented Lagrangian. The theory, developed in the general case in terms of the ordinary Lagrangian and extended to allow the utilization
(Cont'd)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601 |

None

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406 743. JOB

No. 20 Continued.

of traditional penalty functions, is applied. Recently obtained results (using a different manner of proof) are shown to follow immediately using the present approach. Furthermore, the relationship of the calculations to those involving the usual Lagrangian are made explicit, hence all the results obtained previously in terms of the usual Lagrangian are immediately applicable.

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Abstract
of
Serial T-349
22 March 1977

EXACT SENSITIVITY ANALYSIS USING AUGMENTED LAGRANGIANS

by

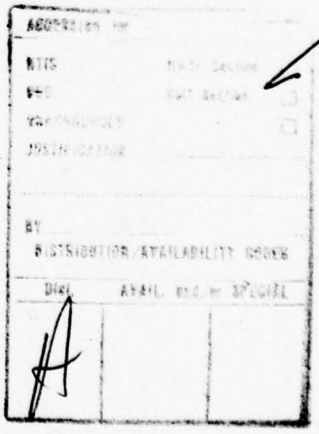
Robert L. Armacost*
Anthony V. Fiacco

This paper reviews several recently developed sensitivity analysis results for a general class of parametric nonlinear programming problems and develops formulas for calculating the sensitivity measures, i.e., the partial derivatives of a Kuhn-Tucker triple and the optimal value function, using an augmented Lagrangian. The theory, developed in the general case in terms of the ordinary Lagrangian and extended to allow the utilization of traditional penalty functions, is applied. Recently obtained results (using a different manner of proof) are shown to follow immediately using the present approach. Furthermore, the relationship of the calculations to those involving the usual Lagrangian are made explicit, hence all the results obtained previously in terms of the usual Lagrangian are immediately applicable.

*U.S. Coast Guard Headquarters, Washington, D.C. Portions of this paper were published in a Doctoral Dissertation submitted to the School of Engineering and Applied Science.

Research Sponsored by
U.S. Army Research Office - Durham

TABLE OF CONTENTS



THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

EXACT SENSITIVITY ANALYSIS USING AUGMENTED LAGRANGIANS

by

Robert L. Armacost
Anthony V. Fiacco

1. Introduction

For a general class of parametric, twice continuously differentiable nonlinear programming problems, Fiacco (1976) obtained a theoretical basis for characterizing the differentiability properties of a local solution and the associated optimal Lagrange multipliers, with respect to parametric variation, and established the use of the usual Lagrangian to calculate exact sensitivity information and a penalty function to estimate sensitivity information. Armacost and Fiacco (1975) obtained first- and second-order changes in the optimal value function and pursued the computational aspects of computing the first-order changes of a Kuhn-Tucker triple. Formulas were obtained for calculating these changes when a Kuhn-Tucker triple is known, and for estimating these changes when a penalty function minimum is known. Armacost and Fiacco (1976a) further refined these results for the problem where the parameters are restricted to be the right-hand side components of the constraints.

The procedure developed by Fiacco (1976) was implemented computationally by Armacost and Mylander (1973) with computational experience reported by Armacost and Fiacco (1974). The computational procedures were extended and additional computational experience reported by Armacost (1976a) and by Armacost and Fiacco (1976b). Recently, Buys and Gonin (1975), paralleling

the cited results developed by way of usual penalty functions, obtained the first-order sensitivity results for the Kuhn-Tucker triple and the optimal value function in terms of an augmented Lagrangian. Armacost (1976b) subsequently used the approach developed by Armacost and Fiacco (1975) to develop these sensitivity results and extend the penalty function approximation procedure to a more general type of sequential algorithm.

This paper presents a much simpler development of the sensitivity results using augmented Lagrangians (obtained independently by Buys and Gonin (1975)) following the direct approach of Fiacco (1976). We also show that these results are exact and equivalent to the sensitivity computations developed by Armacost and Fiacco (1975) when a Kuhn-Tucker triple is known. As indicated, this is based in part on material in the dissertation by Armacost (1976b).

In Section 2, we review the supporting theory from Fiacco (1976) and Armacost and Fiacco (1975). In Section 3, we present the necessary background on augmented Lagrangians. In Section 4, we develop the sensitivity results using augmented Lagrangians, and in Section 5 show their exactness and equivalence to the computed expressions for the case when a Kuhn-Tucker triple is known. Several conclusions and extensions are noted in Section 5.

2. Supporting Theory

The parametric mathematical programming problems considered here are of the form

$$\begin{aligned} &\underset{x \in E^n}{\text{minimize}} \quad f(x,\varepsilon) \\ &\text{subject to} \quad g_i(x,\varepsilon) \geq 0 , \quad i = 1,\ldots,m , \qquad P(\varepsilon) \\ &\qquad\qquad\quad h_j(x,\varepsilon) = 0 , \quad j = 1,\ldots,p , \end{aligned}$$

where x is the usual vector of variables and ε is a k-component vector of numbers called "parameters." It is desired ultimately to develop a complete

characterization of a solution $x(\varepsilon)$ of Problem $P(\varepsilon)$ as a function of ε. In our current work, we have concentrated on certain recently computationally tractable measures of change in a solution as ε is perturbed from a specified value. (Without loss of generality, we assume that the specified value is $\varepsilon = 0$.)

The Lagrangian for Problem $P(\varepsilon)$ is defined as

$$L(x,u,w,\varepsilon) \equiv f(x,\varepsilon) - \sum_{i=1}^{m} u_i g_i(x,\varepsilon) + \sum_{j=1}^{p} w_j h_j(x,\varepsilon) , \qquad (1)$$

where u_i, $i = 1,\ldots,m$ and w_j, $j = 1,\ldots,p$ are "Lagrange multipliers" associated with the inequality and equality constraints, respectively. Any vector $(\bar{x},\bar{u},\bar{w})$ satisfying the usual (first order) Kuhn-Tucker conditions (Fiacco and McCormick, 1968) of Problem $P(\bar{\varepsilon})$ is called a Kuhn-Tucker triple.

The following four assumptions are sufficient to establish the results and are assumed to hold throughout the paper:

A1 -- The functions defining Problem $P(\varepsilon)$ are twice continuously differentiable in (x,ε) in a neighborhood of $(x^*,0)$.

A2 -- The second order sufficient conditions for a local minimum of Problem $P(0)$ hold at x^* with associated Lagrange multipliers u^* and w^*.

A3 -- The gradients $\nabla_x g_i(x^*,0)$ for all i such that $g_i(x^*,0) = 0$, and $\nabla_x h_j(x^*,0)$, $j = 1,\ldots,p$ are linearly independent.

A4 -- Strict complementary slackness holds at $(x^*,0)$ (i.e., $u_i^* > 0$ for all i such that $g_i(x^*,0) = 0$).

Lemma 1: (Local characterization of a Kuhn-Tucker triple, Fiacco, (1976) of Problem $P(\varepsilon)$). If Assumptions A1, A2, A3 and A4 hold for Problem $P(\varepsilon)$ at $(x^*,0)$, then

(a) x^* is a local isolated minimizing point of Problem $P(0)$ and the associated Lagrange multipliers u^* and w^* are unique;

(b) for ε in a neighborhood of 0, there exists a unique, once continuously differentiable vector function $y(\varepsilon) = (x(\varepsilon),u(\varepsilon),w(\varepsilon))^T$ satisfying the second order sufficient conditions for a local minimum of Problem $P(\varepsilon)$ such that $y(0) = (x^*,u^*,w^*)^T = y^*$ and hence, $x(\varepsilon)$ is a locally unique, local minimum of Problem $P(\varepsilon)$ with associated unique Lagrange multipliers $u(\varepsilon)$ and $w(\varepsilon)$; and

(c) for ε near 0, the set of binding inequalities is unchanged, strict complementary slackness holds for $u_i(\varepsilon)$ for i such that $g_i(x(\varepsilon),\varepsilon) = 0$, and the binding constraint gradients are linearly independent at $x(\varepsilon)$.

This result provides a characterization of a local solution of Problem $P(\varepsilon)$ and its associated optimal Lagrange multipliers near $\varepsilon = 0$. It generalizes a theorem first presented by Fiacco and McCormick (1968, Theorem 6) and is closely related to a generalization of the same theorem provided independently by Robinson (1974). It shows that the Kuhn-Tucker triple $y(\varepsilon)$ is unique and well behaved, under the given conditions. Since $y(\varepsilon)$ is once differentiable, the partial derivatives of the components of $y(\varepsilon)$ are well defined. This fact and Assumption A1 also mean that the functions defining Problem $P(\varepsilon)$ are once continuously differentiable functions of ε along the "solution trajectory" $x(\varepsilon)$ near $\varepsilon = 0$, and the Lagrangian is a once continuously differentiable function of ε along the "Kuhn-Tucker point trajectory."

We are thus motivated to determine a means to calculate the various partial derivatives, since this yields a first order estimate of the locally optimal Kuhn-Tucker triple and the problem functions near $\varepsilon = 0$.

Denote by $\nabla_\varepsilon x(\varepsilon) \equiv (\partial x_i(\varepsilon)/\partial \varepsilon_j)$, $i = 1,\ldots,n$, $j = 1,\ldots,k$, the $n \times k$ matrix of partial derivatives of $x(\varepsilon)$ with respect to ε , and define $\nabla_\varepsilon u(\varepsilon)$ and $\nabla_\varepsilon w(\varepsilon)$ in a similar fashion. We then define

$\nabla_\varepsilon y(\varepsilon) \equiv (\nabla_\varepsilon^T x(\varepsilon), \nabla_\varepsilon^T u(\varepsilon), \nabla_\varepsilon^T w(\varepsilon))^T$, an $(n+m+p) \times k$ matrix.

When $y(\varepsilon)$ is available, $\nabla_\varepsilon y(\varepsilon)$ can be calculated by noting that Conclusion (b) of the theorem implies the satisfaction of the Kuhn-Tucker conditions for $P(\varepsilon)$ at $y(\varepsilon)$ near $\varepsilon = 0$, i.e.,

$$\nabla_x L(x(\varepsilon), u(\varepsilon), w(\varepsilon), \varepsilon) = 0 , \tag{2}$$

$$u_i(\varepsilon) g_i(x(\varepsilon), \varepsilon) = 0 , \quad i = 1,\ldots,m , \tag{3}$$

$$h_j(x(\varepsilon), \varepsilon) = 0 , \quad j = 1,\ldots,p . \tag{4}$$

Since the Jacobian, $M(\varepsilon)$, of this system with respect to (x,u,w) (i.e., the matrix obtained by differentiating the left side of (2) - (4) with respect to the components of (x,u,w)) is nonsingular under the given assumptions, the total derivative of the system with respect to ε is well defined and must equal zero. This yields

$$M(\varepsilon) \nabla_\varepsilon y(\varepsilon) = N(\varepsilon) , \tag{5}$$

where $N(\varepsilon)$ is the negative of the Jacobian of the Kuhn-Tucker system with respect to ε , and hence

$$\nabla_\varepsilon y(\varepsilon) = M(\varepsilon)^{-1} N(\varepsilon) . \tag{6}$$

The next result applies this theory to an analysis of the optimal value function of Problem $P(\varepsilon)$ along the Kuhn-Tucker point trajectory $[x(\varepsilon), u(\varepsilon), w(\varepsilon)]$.

The optimal value function is defined as:

$$f^*(\varepsilon) \equiv f[x(\varepsilon), \varepsilon] , \tag{7}$$

and the "optimal value Lagrangian" is defined as:

$$L^*(\varepsilon) = L[x(\varepsilon),u(\varepsilon),w(\varepsilon),\varepsilon] \ . \tag{8}$$

Lemma 2: (First and second order changes in the optimal value function, Armacost and Fiacco (1975)). If Assumptions A1 - A4 hold for Problem $P(\varepsilon)$, then for ε near 0, $f^*(\varepsilon)$ is a twice continuously differentiable function of ε, and

(a) $f^*(\varepsilon) = L^*(\varepsilon)$;

(b) $$\nabla_\varepsilon f^*(\varepsilon) = \nabla_\varepsilon L(x,u,w,\varepsilon) \Big|_{(x,u,w) = (x(\varepsilon),u(\varepsilon),w(\varepsilon))}$$
$$= \nabla_\varepsilon f(x,\varepsilon) - \sum_{i=1}^{m} u_i \nabla_\varepsilon g_i(x,\varepsilon)$$
$$+ \sum_{j=1}^{p} w_j \nabla_\varepsilon h_j(x,\varepsilon) \Big|_{(x,u,w) - (x(\varepsilon),u(\varepsilon),w(\varepsilon))} ;$$

(c) $$\nabla_\varepsilon^2 f^*(\varepsilon) = \nabla_\varepsilon (\nabla_\varepsilon L(x(\varepsilon),u(\varepsilon),w(\varepsilon),\varepsilon)^T) \ .$$

As noted following Theorem 1, when $y(\varepsilon) = [x(\varepsilon),u(\varepsilon),w(\varepsilon)]^T$ is available, $\nabla_\varepsilon y(\varepsilon) = [\nabla_\varepsilon^T x(\varepsilon),\nabla_\varepsilon^T u(\varepsilon),\nabla_\varepsilon^T w(\varepsilon)]^T$ can be calculated. We briefly recapitulate and then analyze various cases in some detail. Recall that the Kuhn-Tucker first-order necessary conditions satisfied by $y(\varepsilon)$ for Problem $P(\varepsilon)$ are

$$\nabla_x L(y(\varepsilon),\varepsilon) \equiv \nabla_x f(x(\varepsilon),\varepsilon) - \sum_{i=1}^{m} u_i(\varepsilon) \nabla_x g_i(x(\varepsilon),\varepsilon)$$
$$+ \sum_{j=1}^{p} w_j(\varepsilon) \nabla_x h_j(x(\varepsilon),\varepsilon) = 0 \tag{2}$$

$$u_i(\varepsilon) g_i(x(\varepsilon),\varepsilon) = 0 \ , \quad i = 1,\ldots,m \ , \tag{3}$$

$$h_j(x(\varepsilon),\varepsilon) = 0 \ , \quad j = 1,\ldots,p \ . \tag{4}$$

Let $N(\varepsilon) = (-\nabla_{\varepsilon x}^{T} L, \ldots, -u_i \nabla_{\varepsilon}^{T} g_i, \ldots, -\nabla_{\varepsilon}^{T} h_j, \ldots)^T$ and let

$$M(\varepsilon) = \left[\begin{array}{c|c|c} \nabla_x^2 L & \cdots -\nabla_x g_i^T \cdots & \cdots \nabla_x h_j^T \cdots \\ \hline \vdots & \ddots & \\ u_i \nabla_x g_i & g_i & 0 \\ \vdots & & \ddots \\ \hline \vdots & & \\ \nabla_x h_j & 0 & 0 \end{array}\right]$$

Since the system (2) - (4) is identically satisfied for ε near 0, it can be differentiated with respect to ε to obtain

$$M(\varepsilon)\nabla_{\varepsilon} y(\varepsilon) = N(\varepsilon) . \tag{5}$$

Under the conditions of Theorem 1, $M(\varepsilon)$ has an inverse, thus

$$\nabla_{\varepsilon} y(\varepsilon) = M(\varepsilon)^{-1} N(\varepsilon) ; \tag{6}$$

$M(\varepsilon)$ is an $n+m+p$ square matrix and $\nabla_{\varepsilon} y(\varepsilon)$ and $N(\varepsilon)$ are $n+m+p \times k$ matrices.

Clearly, any method of solving (5), a system of linear equations, is satisfactory and $M(\varepsilon)$ need not be inverted as in (6). However, under the given assumptions the work involved in calculating $M(\varepsilon)^{-1}$ can be significantly reduced, as will become evident. Assume throughout this section that the conditions of Theorem 1 hold, and suppose, without loss of generality, that the first r inequality constraints are binding. Let $\bar{M}(\varepsilon)$ be defined as follows:

$$\bar{M}(\varepsilon) = \left[\begin{array}{c|ccc|ccc} \nabla_x^2 L & -\nabla_x^T g_1 & \cdots & -\nabla_x g_r^T & \nabla_x h_1^T & \cdots & \nabla_x h_p^T \\ \hline u_1 \nabla_x g_1 & g_1 & & & & & \\ \vdots & & \ddots & & & 0 & \\ u_r \nabla_x g_r & & & g_r & & & \\ \hline \nabla_x h_1 & & & & & & \\ \vdots & & 0 & & & 0 & \\ \nabla_x h_p & & & & & & \end{array}\right]$$

Then, rearranging the rows and columns corresponding to the last $m - r$ (nonbinding) inequality constraints, it follows that

$$M(\varepsilon) = \left[\begin{array}{cccc} \bar{M}(\varepsilon) & -\nabla_x g_{r+1}^T, \ldots, -\nabla_x g_m^T \\ 0 & g_{r+1} & & \\ \vdots & & \ddots & \\ 0 & & & g_m \end{array}\right]$$

Let $\bar{y}(\varepsilon) = (x(\varepsilon), \bar{u}(\varepsilon), w(\varepsilon))^T$ and $\bar{N}(\varepsilon) = (-(\nabla_{\varepsilon x}^2 L)^T, \ldots, -u_i \nabla_\varepsilon^T g_i, \ldots, -\nabla_\varepsilon^T h_j \ldots)^T$, where $i = 1,\ldots,r$, $j = 1,\ldots,p$, and $\bar{u}(\varepsilon) = (u_1(\varepsilon), \ldots, u_r(\varepsilon))^T$. Since $g_i(x(0),0) > 0$, $i = r+1,\ldots,m$, complementary slackness and continuity imply that $u_i(\varepsilon) = 0$, $i = r+1,\ldots,m$ for ε near 0. Hence, the corresponding components of $\nabla_\varepsilon y(\varepsilon)$ are zero, i.e., $\nabla_\varepsilon u_i = 0$, $i = r+1,\ldots,m$, and it follows that (5) may be reduced to solving the system

$$\bar{M}(\varepsilon) \nabla_\varepsilon \bar{y}(\varepsilon) = \bar{N}(\varepsilon) . \tag{9}$$

Let $\bar{g} \equiv (g_1, \ldots, g_r)^T$, $\bar{G} \equiv \text{diag}(g_i)$ and $\bar{U} \equiv \text{diag}(u_i)$, $i = 1,\ldots,r$, $\nabla_x \bar{g} = [\nabla g_1^T, \ldots, \nabla g_r^T]^T$ and $\nabla_x h = [\nabla h_1^T, \ldots, \nabla h_p^T]^T$. Since $\bar{G} = 0$, it follows that

$$\bar{M}(\varepsilon) = \begin{bmatrix} I & 0 & 0 \\ 0 & -\bar{U} & 0 \\ 0 & 0 & I \end{bmatrix} \cdot \begin{bmatrix} \nabla_x^2 L & -\nabla_x^T \bar{g} & \nabla_x^T h \\ -\nabla_x \bar{g} & 0 & 0 \\ \nabla_x h & 0 & 0 \end{bmatrix}$$

Thus

$$\bar{M}(\varepsilon)^{-1} = \begin{bmatrix} \nabla_x^2 L & -\nabla_x^T \bar{g} & \nabla_x^T h \\ -\nabla_x \bar{g} & 0 & 0 \\ \nabla_x h & 0 & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} I & 0 & 0 \\ 0 & -\bar{U}^{-1} & 0 \\ 0 & 0 & I \end{bmatrix}. \tag{10}$$

Thus far, we have assumed no special structure on the problem or the nature of $M(\varepsilon)$. However, in order to make further progress in calculating $\bar{M}(\varepsilon)^{-1}$, consider several special cases: (1) $\nabla_x^2 L^{-1}$ exists; (2) $\nabla_x^2 L = 0$; (3) there are n linearly independent constraint gradients; and (4) $r + p < n$, $\nabla_x^2 L \neq 0$ and $\nabla_x^2 L^{-1}$ does not exist. Let $P \equiv (-\nabla_x^T \bar{g}, \nabla_x^T h)^T$, an $(r+p) \times n$ matrix. Let $\hat{M}(\varepsilon)^{-1}$ denote the left matrix on the right-hand side of Equation (10), thus

$$\hat{M}(\varepsilon) = \begin{bmatrix} \nabla_x^2 L & P^T \\ P & 0 \end{bmatrix} \tag{11}$$

Now suppose that

$$\hat{M}(\varepsilon)^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{12}$$

Our task is to determine the A_{ij} for the various cases.

Case 1: $\nabla_x^2 L^{-1}$ exists.

It is easily shown that

$$A_{11} = \nabla_x^2 L^{-1}[I-P^T(P\nabla_x^2 L^{-1}P^T)^{-1}P\nabla_x^2 L^{-1}]$$
$$A_{12} = A_{21}^T = \nabla_x^2 L^{-1}P^T[P\nabla_x^2 L^{-1}P^T]^{-1} \qquad (13)$$
$$A_{22} = -[P\nabla_x^2 L^{-1}P^T]^{-1} .$$

Note that $[P\nabla_x^2 L^{-1}P^T]^{-1}$ exists by our assumptions.

Case 2: $\nabla_x^2 L = 0$.

There are two possible situations: there are $r + p < n$, or $r + p = n$ linearly independent binding constraint gradients. If there are less than n , Assumption A2 is violated and it is easily seen that $\hat{M}(\varepsilon)^{-1}$ does not exist. For example, this corresponds to the situation characterizing a degenerate solution in linear programming. In this case, $\nabla_\varepsilon \bar{y}(\varepsilon)$ may not be differentiable. We shall not pursue this possibility further here. When there are n linearly independent binding constraint gradients, we have a special instance of Case 3, which is developed below.

Case 3: There are n linearly independent binding constraint gradients.

The Jacobian of the n constraints with respect to the n variables is non-vanishing. It is easily shown that in (11), $(P^T)^{-1} = P^{-T}$ exists. Hence,

$$A_{11} = 0$$
$$A_{12} = A_{21}^T = P^{-1} \qquad (14)$$
$$A_{22} = -P^{-T}\nabla_x^2 L\, P^{-1} .$$

Note that if $\nabla_x^2 L^{-1}$ exists, (14) is computable from (13); however, here the existence of $\nabla_x^2 L^{-1}$ is not assumed. Also, the remaining Case 2 possibility mentioned above gives $A_{11} = A_{22} = 0$ and $A_{12} = A_{21}^T = P^{-1}$. Note also that the standard linear programming problem nondegenerate solution case falls into this latter category, with n linearly independent binding constraint gradients and $\nabla_x^2 L = 0$.

Case 4: $r + p < n$, $\nabla_x^2 L \neq 0$ and $\nabla_x^2 L^{-1}$ does not exist.

This represents the more general situation and is treated in detail here. In Equation (9) $\bar{M}(\varepsilon)$ is an $n+r+p$ square matrix. By Assumption A3, the $\nabla_x g_i$, $i = 1,\ldots,r$, and the $\nabla_x h_j$, $j = 1,\ldots,p$, are linearly independent. By Assumption A2 the second-order sufficient conditions for a local minimum are satisfied. Thus, the Hessian of the Lagrangian is positive definite with respect to those nonzero vectors orthogonal to the binding constraint gradients and hence

$$z^T \nabla_x^2 L \, z > 0 \text{ , for all } z \neq 0 \text{ such that}$$

$$P z = (-\nabla_x \bar{g}, \nabla_x h)^T z = 0 . \tag{15}$$

Under the assumption of linear independence of the binding constraint gradients, P has rank $r + p$, and without loss of generality, assume that the submatrix involving the first $r + p$ columns of P is nonsingular. Therefore, we partition P as follows:

$$P = (P_D, P_I) ,$$

where P_D is an $(r+p) \times (r+p)$ nonsingular matrix and P_I is an $(r+p) \times q$ matrix where $q = n-r-p$. It is easily shown that the matrix

$$D_{xx} = T^T \nabla_x^2 L \, T \text{ , where } T = \begin{bmatrix} -P_D^{-1} P_I \\ I \end{bmatrix}$$

is positive definite. This fact leads us to a representation of $\hat{M}(\varepsilon)^{-1}$.

Using the partitions defined above,

$$\hat{M}(\varepsilon) = \begin{bmatrix} \nabla_x^2 L & P_D^T \\ & P_I^T \\ P_D, P_I & 0 \end{bmatrix}, \tag{16}$$

and

$$\hat{M}^{-1}(\varepsilon) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where

$$\begin{aligned} A_{11} &= T\, D_{xx}^{-1}\, T^T , \\ A_{12} &= A_{21}^T = [I - A_{11}\nabla_x^2 L] \begin{bmatrix} P_D^{-1} \\ 0 \end{bmatrix} , \end{aligned} \tag{17}$$

and

$$A_{22} = -A_{21}\nabla_x^2 L \begin{bmatrix} P_D^{-1} \\ 0 \end{bmatrix} .$$

This representation of the inverse of $\bar{M}(\varepsilon)$ is due to McCormick (1975).

We have now developed analytical expressions for $\hat{M}(\varepsilon)^{-1}$ for all cases that can occur, and now turn to the calculation of $\nabla_\varepsilon \bar{y}(\varepsilon)$, using the block components A_{ij} of $\hat{M}(\varepsilon)^{-1}$.

Careful attention to the algebra yields the following expression for the first-order sensitivity of a Kuhn-Tucker triple as

$$\nabla_\varepsilon \bar{y}(\varepsilon) = \begin{bmatrix} \nabla_\varepsilon x(\varepsilon) \\ \hline \nabla_\varepsilon \bar{u}(\varepsilon) \\ \nabla_\varepsilon w(\varepsilon) \end{bmatrix} = \left[\begin{array}{c} -A_{11}\nabla_{\varepsilon x}{}^2 L + A_{12}\begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix} \\ \hline -A_{21}\nabla_{\varepsilon x}{}^2 L + A_{22}\begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix} \end{array}\right] = \hat{M}(\varepsilon)^{-1}\hat{N}(\varepsilon) \quad (18)$$

where $$\hat{N}(\varepsilon) = \begin{bmatrix} -\nabla_{\varepsilon x}{}^2 L \\ \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix}.$$

Equation (18) holds for the general problem whenever Equation (9) is well defined, e.g., when the conditions of Theorem 1 hold. For the particular cases treated earlier in this section, $\nabla_\varepsilon \bar{y}(\varepsilon)$ may be calculated from (18) by first evaluating the A_{ij} as given in (13), (14), or (17), depending on which conditions apply.

Equivalent results will be obtained in Section 4 using augmented Lagrangians.

3. Background on Augmented Lagrangians

As noted in the Introduction, Fiacco (1976) established a basis for estimating the sensitivity of a Kuhn-Tucker triple when the Kuhn-Tucker triple is not known but can be estimated using a penalty function. Penalty function algorithms belong to a more general class of algorithms whereby the constrained problem is transformed into a sequence of unconstrained problems by means of an auxiliary function.

Motivated in part by an effort either to "regularize" the usual Lagrangian (1) or to overcome the ill-conditioning typically associated with the traditional penalty function procedures, several other classes of related auxiliary functions have recently received considerable attention. These include exact penalty functions and generalized Lagrangians [Arrow, Gould and Howe (1973)] and augmented Lagrangians [Hestenes (1969) and Buys (1972)].

The functions are similar in structure and derive their different names largely from the viewpoint taken in their development. The following description gives a general idea of the different approaches applied to a "standard" NLP problem of the form $P(\varepsilon)$, i.e., where the parameter ε is not present.

Generalized Lagrangians generally refer to more general forms of the usual Lagrangian (1) associated with Problem P (suppressing ε throughout). Typically, f , g_i , h_j , u_i or w_j are replaced by functions of these quantities. The object is to structure the new function so that it behaves "like" a Lagrangian (e.g., in characterizing optimality) but acquires certain desirable properties (e.g., convexity or strict convexity) that (1) does not possess. A popular extension introduces generalized Lagrange multipliers that may also be functions of x . Augmented Lagrangians are usually formed by adding a penalty term to the usual Lagrangian, though this class can easily be enlarged by using the generalized Lagrangians described earlier in the paragraph. In this context, the term "Method of Multipliers" refers to a particular approach using a particular form of the augmented Lagrangian. The term is due to Hestenes (1969) who proposed an algorithm based on sequentially improving the estimates of the Lagrange multipliers.

A penalty function would be considered exact if, for particular values of certain parameters, an unconstrained (local) minimum of the function (locally) solves the given programming problem. Since the "optimal value" of these parameters is generally unknown, they must be estimated, and since these parameters often correspond to Lagrange multipliers, the result is that augmented Lagrangian and exact penalty function algorithms are quite similar in spirit.

It is much easier to deal with equality constraints in these methods. When considering inequality constraints, penalty functions may be introduced for which higher than first order differentiability is not inherited from the problem functions, and which, in some cases, are only piecewise differentiable.

Arrow, Gould, and Howe (1973) develop a saddle point theory for an "extended" Lagrangian, defining multiplier functions with certain limiting properties. Realizations of their extended Lagrangian yield many of the augmented Lagrangians and exact penalty functions found in the literature.

As indicated, a primary motivation for the study of augmented Lagrangians and exact penalty functions has been to overcome the problems of ill-conditioning associated with ordinary penalty functions. Because of this, considerable effort has been made to investigate the computational aspects of these algorithms [Bertsekas (1975), Miele et al (1972), Rockafellar (1973), and Rupp (1976)]. Buys (1972) provided perhaps the best detailed analysis of a dual approach, using an augmented Lagrangian first introduced by Rockafellar (1970). Recently, Buys and Gonin (1975) used the same approach to obtain sensitivity analysis results in terms of the same augmented Lagrangian formulation. Since our intent is to redevelop these sensitivity results, further descriptions of the computational approaches are not included.

In the interest of keeping the focus, the specific augmented Lagrangian used by Buys and Gonin (1975) will be utilized here. It is noted, however, that the approach is directly applicable to the general form of the extended Lagrangian given by Arrow, Gould, and Howe, thus encompassing most of the currently popular augmented Lagrangians and exact penalty functions.

4. Augmented Lagrangians and Sensitivity Analysis

Rather than use the dual approach of Buys, the sensitivity results which obtain using augmented Lagrangians follow directly by considering the equations which are satisfied at a solution point, as in Lemma 1. The key point in the following development is that the gradient of the augmented Lagrangian is equal to the gradient of the ordinary Lagrangian near a solution point, when the problem parameter ε is perturbed.

Let $c > 0$ be a constant, $J \equiv \{i: u_i - cg_i(x,0) \geq 0 \ , \ i=1,\ldots,m\}$ And $K \equiv \{i: u_i - cg_i(x,0) < 0, \ i=1,\ldots,m\}$. Assume that Assumptions

A1, A2, A3 and A4 hold for Problem P(ε). Let x^* be a local minimum of Problem P(0) with associated Lagrange multipliers u^* and w^*. Let $B^*(0) = \{i: g_i(x^*,0) = 0 , i=1,\ldots,m\}$. Then, for $i \in B^*(0)$, $u_i^* - cg_i(x^*,0) = u_i^* > 0$ and for $i = 1,\ldots,m \notin B^*(0)$, $g_i(x^*,0) > 0$ and hence $u_i^* - cg_i(x^*,0) = -cg_i(x^*,0) < 0$. Thus at the solution point, J is defined with strict inequality and corresponds to $B^*(0)$ and K contains all i such that $g_i(x^*,0) > 0$.

The augmented Lagrangian is defined as

$$\begin{aligned}\emptyset(x,u,w,\varepsilon,c) \equiv f(x,\varepsilon) - \sum_{i\in J} (u_i - \tfrac{1}{2}cg_i(x,\varepsilon))g_i(x,\varepsilon) \\ + \sum_{j=i}^{p} (w_j + \tfrac{1}{2}ch_j(x,\varepsilon))h_j(x,\varepsilon) - (1/2c) \sum_{i\in K} u_i^2 .\end{aligned} \tag{19}$$

The gradient of $\emptyset$ taken with respect to x is

$$\nabla_x \emptyset(x,u,w,\varepsilon,c) = \nabla_x L + \sum_{i\in J} cg_i \nabla_x g_i + \sum_{j=1}^{p} ch_j \nabla_x h_j . \tag{20}$$

At $(x,u,w,\varepsilon,c) = (x^*,u^*,w^*,0,c)$, since $g_i(x^*,0) = 0$, $i \in J$, and $h_j(x^*,0) = 0$ for all j, it follows that

$$\nabla_x \emptyset(x^*,u^*,w^*,0,c) = \nabla_x L(x^*,u^*,w^*,0) . \tag{21}$$

The augmented Lagrangian is twice continuously differentiable in x except at points where $u_i - cg_i(x,\varepsilon) = 0$. By Assumption A4, $u_i^* - cg_i(x^*,0) \neq 0$ for all i and hence, $\emptyset$ is twice continuously differentiable for (x,u,ε) near $(x^*,u^*,0)$. Differentiating (20) with respect to x yields

$$\begin{aligned}\nabla_x^2 \emptyset(x,u,w,\varepsilon,c) = \nabla_x^2 L + \sum_{i\in J} cg_i \nabla_x^2 g_i + \sum_{i\in J} c\nabla_x g_i^T \nabla_x g_i \\ + \sum_{j=1}^{p} ch_j \nabla_x^2 h_j + \sum_{j=1}^{p} c\nabla_x h_j^T \nabla_x h_j .\end{aligned} \tag{22}$$

Again, at $(x,u,w,\varepsilon,c) = (x^*,u^*,w^*,0,c)$,

$$\nabla_x^2 \emptyset(x^*,u^*,w^*,0,c) = \nabla_x^2 L^* + c \sum_{i\in J} \nabla_x g_i^{*T} \nabla_x g_i^* + c \sum_{j=1}^{p} \nabla_x h_j^{*T} \nabla_x h_j^* , \tag{23}$$

where the exponent * denotes evaluation of the given functions at $x^*,u^*,w^*,0$. Without loss of generality, assume that $J = B^*(0) = \{1,\ldots,r\}$. Then, recalling (see Section 2, just prior to Equation 11) that

$$P = \begin{bmatrix} -\nabla_x \bar{g} \\ \nabla_x h \end{bmatrix} ,$$

it follows that

$$\nabla_x^2 \emptyset = \nabla_x^2 L + cP^T P . \tag{24}$$

Since $\nabla_x^2 L$ is positive definite for all $z \neq 0$ such that $Pz = 0$ by Assumptions A_2 and A_4 , then for all c sufficiently large it follows easily that $\nabla_x^2 \emptyset$ is positive definite near $(x^*,u^*,w^*,0,c)$. Thus, there is a number $c^* > 0$ such that for $c > c^*$, $\nabla_x^2 \emptyset(x^*,u^*,w^*,0,c)$ is positive definite (and hence nonsingular). Assume that $c > c^*$ and the notation of Section 2 in the remainder of this section. Recall that the first r inequality constraints are assumed binding, the superbar indicating evaluation at $i = 1,\ldots,r$. For convenience we shall use a bar underscore to denote evaluation at $i = r + 1,\ldots,m$. The main result can be stated as follows.

Theorem 1. (Sensitivity results using an augmented Lagrangian for Problem P(ε)).

If Assumptions A1, A2, A3 and A4 hold for Problem P(ε), then for ε near 0 and $c > c^*$, there exists a unique, once continuously differentiable vector function $y_a(\varepsilon,c) = (x(\varepsilon,c),\bar{u}(\varepsilon,c),w(\varepsilon,c),\underline{u}(\varepsilon,c))^T$ satisfying

$$\nabla_x \emptyset(x,u,w,\varepsilon,c) = 0 , \tag{25}$$

$$u_i g_i(x,\varepsilon) = 0 , \quad i = 1,\ldots,r , \tag{26}$$

$$h_j(x,\varepsilon) = 0 , \quad j = 1,\ldots,p , \tag{27}$$

$$u_i g_i(x,\varepsilon) = 0 , \quad i = r+1,\ldots,m , \tag{28}$$

with $(x(\varepsilon,c),u(\varepsilon,c),w(\varepsilon,c))^T = (x(\varepsilon),u(\varepsilon),w(\varepsilon))^T \equiv y(\varepsilon)$ (the Kuhn-Tucker triple of Lemma 1) and such that for any ε near 0 and $c > c^*$, $x(\varepsilon,c)$ is a locally unique unconstrained local minimizing point of $\emptyset[x,u(\varepsilon,c),w(\varepsilon,c),\varepsilon,c]$ and $\nabla_x^2 \emptyset[x,\varepsilon,c]$ is positive definite for (x,u,w) near (x^*,u^*,w^*) .

Proof.

The Jacobian matrix of Equations (25) - (28) taken with respect to $(x,\bar{u},w,\underline{u})$, the precise analogy of the Jacobian $M(\varepsilon)$ of the Kuhn-Tucker system (2) - (4) , is

$$M_a = \begin{bmatrix} \nabla_x^2 \emptyset & -\nabla_x \bar{g}^T & \nabla_x h^T & -\nabla_x \underline{g}^T \\ \bar{U}\nabla_x \bar{g} & \bar{G} & 0 & 0 \\ \nabla_x h & 0 & 0 & 0 \\ \underline{U}\nabla_x \underline{g} & 0 & 0 & \underline{G} \end{bmatrix} ,$$

which, evaluated at $(x,\bar{u},w,\underline{u},\varepsilon,c) = (x^*,\bar{u}^*,w^*,\underline{u}^*,0,c)$, gives

$$M_a^* = \begin{bmatrix} \nabla_x^2 \emptyset^* & -\nabla_x \bar{g}^{*T} & \nabla_x h^{*T} & -\nabla_x \underline{g}^{*T} \\ \bar{U}^* \nabla_x \bar{g}^* & 0 & 0 & 0 \\ \nabla_x h^* & 0 & 0 & 0 \\ 0 & 0 & 0 & \underline{G}^* \end{bmatrix} . \tag{29}$$

Under the given assumptions, it follows that $M_a{}^{*-1}$ exists and hence, by the implicit function theorem, there exists a unique, once continuously differentiable vector function $y_a(\varepsilon,c) = (x(\varepsilon,c),\bar{u}(\varepsilon,c),w(\varepsilon,c),\underline{u}(\varepsilon,c))^T$ satisfying Equations (25) - (28) for ε near 0 and $c > c^*$.

As indicated above, $\nabla_x^2 \emptyset$ is positive definite at, and hence near $(x^*,u^*,w^*,0,c)$. It follows that $x(\varepsilon,c)$ is a locally unique local minimum of $\emptyset(x,u(\varepsilon,c),w(\varepsilon,c),\varepsilon,c)$ for ε near 0 and $c > c^*$.

Observing that $\nabla_x \emptyset \equiv \nabla_x L$ at $(x(\varepsilon,c),u(\varepsilon,c),w(\varepsilon,c))$ near $\varepsilon = 0$, a comparison of (25) - (28) and the Kuhn-Tucker system (2) - (4), and the uniqueness of the solutions of both systems, implies that $(x(\varepsilon,c),u(\varepsilon,c),w(\varepsilon,c)) \equiv (x(\varepsilon),u(\varepsilon),w(\varepsilon))$ for ε near 0, and the proof is complete.

It may also be noted in passing that $\emptyset \equiv f$ at $(x(\varepsilon,c),u(\varepsilon,c),w(\varepsilon,c))$ for ε near 0. Along with the conclusions given in the last part of the proof, this immediately implies the results that precisely parallel those given in Lemma 2 for the optimal value function and its first and second derivatives, where the augmented Lagrangian $\emptyset$ replaces the usual Lagrangian L in the given expressions.

Since the system (25) - (28) is once continuously differentiable and identically equal to zero for ε near 0 and $c > c^*$, it can be differentiated with respect to ε to yield

$$M_a(\varepsilon,c)\ \nabla_\varepsilon y_a(\varepsilon,c) = N_a(\varepsilon,c) \tag{30}$$

where $M_a(\varepsilon,c)$ is defined by Equation (29) for ε near 0 and $c > c^*$,

$$\nabla_\varepsilon y_a(\varepsilon,c) \equiv \begin{bmatrix} \nabla_\varepsilon x(\varepsilon,c) \\ \nabla_\varepsilon \bar{u}(\varepsilon,c) \\ \nabla_\varepsilon w(\varepsilon,c) \\ \nabla_\varepsilon \underline{u}(\varepsilon,c) \end{bmatrix}$$

and

$$N_a(\varepsilon,c) \equiv \begin{bmatrix} -\nabla_{\varepsilon x}{}^2 \phi \\ -\bar{U}\nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \\ 0 \end{bmatrix} .$$

Notice first that Equation (30) yields $\underline{G}\nabla_\varepsilon \underline{u}(\varepsilon,c) = 0$ and since $\underline{G} > 0$, it follows that $\nabla_\varepsilon \underline{u}(\varepsilon,c) = 0$ as expected since $u_i(\varepsilon,c) \equiv 0$ for $i \notin B^*(0)$. Let $\bar{M}_a(\varepsilon,c)$, $\nabla_\varepsilon \bar{y}_a(\varepsilon,c)$ and $\bar{N}_a(\varepsilon,c)$ be the portions of (30) excluding the nonbinding inequality constraints. Then with

$$\hat{M}_a(\varepsilon,c) = \begin{bmatrix} \nabla_x{}^2 \phi & P^T \\ P & 0 \end{bmatrix},$$

and letting

$$D_g = \begin{bmatrix} -\bar{U}^{-1} & 0 \\ 0 & I \end{bmatrix},$$

it follows that

$$\bar{M}_a(\varepsilon,c)^{-1} = \hat{M}_a(\varepsilon,c)^{-1} \begin{bmatrix} I & 0 \\ 0 & D_g \end{bmatrix}.$$

Letting

$$\hat{M}_a(\varepsilon,c)^{-1} \equiv \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \tag{31}$$

we have that

$$\bar{M}_a(\varepsilon,c)^{-1} = \begin{bmatrix} C_{11} & C_{12}D_g \\ C_{21} & C_{22}D_g \end{bmatrix} .$$

The elements C_{ij} are immediately determined using a result given in Section 2. Recall the computations for the general problem when $\nabla_x^2 L^{-1}$ exists (Equations (13)). The situation here is identical with $\nabla_x^2 \phi$ replacing $\nabla_x^2 L$ and we obtain

$$\begin{aligned} C_{11} &= \nabla_x^2\phi^{-1}(I - P^T(P\ \nabla_x^2\phi^{-1}P^T)^{-1}P\ \nabla_x^2\phi^{-1}) , \\ C_{12} &= C_{21}^T = \nabla_x^2\phi^{-1}P^T(P\ \nabla_x^2\phi^{-1}P^T)^{-1} , \\ C_{22} &= -(P\ \nabla_x^2\phi^{-1}P^T)^{-1} . \end{aligned} \tag{32}$$

Following the development in Section 2 (in particular, using Equation (18)), the estimate of the first order sensitivity of the Kuhn-Tucker triple for Problem $P(\varepsilon)$ is given by

$$\nabla_\varepsilon \bar{y}_a(\varepsilon,c) = \begin{bmatrix} \nabla_\varepsilon x(\varepsilon,c) \\ \hline \nabla_\varepsilon \bar{u}(\varepsilon,c) \\ \nabla_\varepsilon w(\varepsilon,c) \end{bmatrix} = \left[\begin{array}{c} -C_{11}\nabla_{\varepsilon x}^2\phi + C_{12}\begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix} \\ \hline -C_{21}\nabla_{\varepsilon x}^2\phi + C_{22}\begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix} \end{array}\right] . \tag{33}$$

5. Equivalence of Sensitivity Calculations

Since $\bar{y}_a(\varepsilon,c) \equiv \bar{y}(\varepsilon)$ near $\varepsilon = 0$, we must have $\nabla_\varepsilon y_a(\varepsilon,c) \equiv \nabla_\varepsilon y(\varepsilon)$ for ε near 0.

The structure of the original problem and the augmented Lagrangian can be used to demonstrate explicitly that $\nabla_\varepsilon y_a(\varepsilon,c) \equiv \nabla_\varepsilon y(\varepsilon)$. Recall that for the given Problem $P(\varepsilon)$,

$$\nabla_\varepsilon \bar{y}(\varepsilon) = \hat{M}(\varepsilon)^{-1} \hat{N}(\varepsilon) \tag{18}$$

where

$$\hat{N}(\varepsilon) = \begin{bmatrix} -\nabla_{\varepsilon x}{}^2 L \\ \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix}, \quad \hat{M}(\varepsilon)^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

and the components A_{ij} are defined by Equations (13), (14), or (17) depending on which conditions apply.

Differentiating Equation (20) with respect to ε yields

$$\nabla_{\varepsilon x}{}^2 \phi = \nabla_{\varepsilon x}{}^2 L + c \sum_{i=1}^{r} g_i \nabla_{\varepsilon x}{}^2 g_i + c \sum_{i=1}^{r} \nabla_x g_i{}^T \nabla_\varepsilon g_i$$
$$+ c \sum_{j=1}^{p} h_j \nabla_{\varepsilon x}{}^2 h_j + c \sum_{j=1}^{p} \nabla_x h_j{}^T \nabla_\varepsilon h_j .$$

These and the following equations are evaluated at $y_a(\varepsilon,c)$ for ε near 0, and hence, $g_i(x(\varepsilon,c),\varepsilon) = 0$, $i \in J = B^*(0)$ and $h_j(x(\varepsilon,c),\varepsilon) = 0$, $j = 1,\ldots,p$.

This yields

$$\nabla_{\varepsilon x}{}^2 \phi = \nabla_{\varepsilon x}{}^2 L - cP^T \begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix}$$

with P defined as before. An explicit expression for $\nabla_\varepsilon y_a(\varepsilon,c)$ can now be derived which simplifies (33). Since

$$C_{11}\nabla_{\varepsilon x}{}^2\emptyset = C_{11}\nabla_{\varepsilon x}{}^2 L - cC_{11}P^T \begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix} ,$$

and since (32) implies that

$$C_{11}P^T = 0 ,$$

it follows that

$$C_{11}\nabla_{\varepsilon x}{}^2\emptyset = C_{11}\nabla_{\varepsilon x}{}^2 L .$$

Similarly,

$$C_{21}\nabla_{\varepsilon x}{}^2\emptyset = C_{21}(\nabla_{\varepsilon x}{}^2 L - cP^T \begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix}) ,$$

and since (32) implies that

$$C_{21}P^T = I ,$$

we obtain

$$C_{21}\nabla_{\varepsilon x}{}^2\emptyset = C_{21}\nabla_{\varepsilon x}{}^2 L - c \begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix} .$$

The first order sensitivity of the Kuhn-Tucker triple may then be written as

$$\begin{aligned} \nabla_\varepsilon \bar{y}_a(\varepsilon,c) &= \begin{bmatrix} -C_{11}\nabla_{\varepsilon x}{}^2 L + C_{12} \begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix} \\ -C_{21}\nabla_{\varepsilon x}{}^2 L + (C_{22} + cI) \begin{bmatrix} \nabla_\varepsilon \bar{g} \\ -\nabla_\varepsilon h \end{bmatrix} \end{bmatrix} \\ &= \left[\hat{M}_a(\varepsilon,c)^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & cI \end{bmatrix} \right] \hat{N}(\varepsilon) \end{aligned} \qquad (34)$$

for ε near 0 .

Since it was previously concluded that $\nabla_\varepsilon \bar{y}(\varepsilon) \equiv \nabla_\varepsilon \bar{y}_a(\varepsilon,c)$ for ε near 0 and $c > c^*$, it must be true that

$$\hat{M}(\varepsilon)^{-1} \equiv \hat{M}_a(\varepsilon,c)^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & cI \end{bmatrix} . \tag{35}$$

To show explicitly that this relationship holds, note that

$$\begin{aligned}
\hat{M}_a(\varepsilon,c) &= \begin{bmatrix} \nabla_x^2 \phi & P^T \\ P & 0 \end{bmatrix} && \text{(definition)} \\
&= \begin{bmatrix} \nabla_x^2 L + cP^T P & P^T \\ P & 0 \end{bmatrix} && \text{(using (24))} \\
&= \begin{bmatrix} I & cP^T \\ 0 & I \end{bmatrix} \begin{bmatrix} \nabla_x^2 L & P^T \\ P & 0 \end{bmatrix} \\
&= \left(I + \begin{bmatrix} 0 & cP^T \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \nabla^2 L & P^T \\ P & 0 \end{bmatrix} \\
&= \left(I + \begin{bmatrix} \nabla_x^2 \phi & P^T \\ P & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & cI \end{bmatrix} \right) \hat{M}(\varepsilon) && \text{(using (11))} \\
&= \left([I + \hat{M}_a(\varepsilon,c)] \begin{bmatrix} 0 & 0 \\ 0 & cI \end{bmatrix} \right) \hat{M}(\varepsilon) .
\end{aligned}$$

Premultiplying the last equation by $\hat{M}_a(\varepsilon,c)^{-1}$ and post-multiplying by $\hat{M}(\varepsilon)^{-1}$ yields (35) .

6. Conclusions and Extensions

Under the given conditions, the formulas (32) and (34) obtained using the augmented Lagrangian may be viewed as an alternative computational device for calculating the first order sensitivity of a Kuhn-Tucker triple for Problem $P(\varepsilon)$, encompassing the formula (Equation (18)) obtained in Section 2 together with the formulas (13), (14) and (17) that apply in the various cases specified. That is, using the augmented Lagrangian instead of the usual Lagrangian, only Case 1 (Section 2) can arise, and, hence, only the one set of formulas (32) for the inverse of the Jacobian are needed under the given conditions. Note that both approaches require knowledge of the binding inequality constraint indices and in fact require the determination of a Kuhn-Tucker triple. In addition, Equations (32) and (34) require a value of c for which $\nabla_x^2 \emptyset$ is positive definite. Using either the usual Lagrangian or the augmented Lagrangian, the indicated information permits an exact calculation of the first order sensitivity of the Kuhn-Tucker triple.

The requirements for exact sensitivity information are rather severe and in effect result in "post-optimality" sensitivity analysis calculations. If inexact sensitivity information is considered, then one obvious possibility is to use estimates of the local solution and its associated optimal Lagrange multipliers in the given formulas. (The question of error-bounds arises, one that we do not pursue here. However, for certain important results is this connection, the interested reader is referred to Robinson (1973).) With the usual Lagrangian (Section 2) problem-oriented approach, any algorithm could be used that provides such estimates.

The augmented Lagrangian approach is already "algorithmic", involving first estimating the constant c and the optimal Lagrange multipliers $u(\varepsilon,c)$, $w(\varepsilon,c)$ for Problem $P(\varepsilon)$, and then minimizing $\emptyset$ over x . Of course, any valid unconstrained minimization algorithm could be used, as could any appropriate procedure for estimating the optimal multipliers. In this

regard, from Theorem 1 it may be observed that if $u(\varepsilon),w(\varepsilon)$ are the (locally) unique optimal Lagrange multipliers for Problem $P(\varepsilon)$ associated with $x(\varepsilon)$ and if ε is sufficiently close to 0, then

$(u(\varepsilon,c),w(\varepsilon,c)) = (u(\varepsilon),w(\varepsilon)) \to (u^*,w^*)$ as $\varepsilon \to 0$ and hence the locally unique local minimum $\hat{x}(\varepsilon,c)$ of $\emptyset(x,u(\varepsilon),w(\varepsilon),c)$ is given by $\hat{x}(\varepsilon,c) = x(\varepsilon,c) = x(\varepsilon)$ and $x(\varepsilon) \to x(0) = x^*$. Clearly, the first partial derivatives (with respect to ε) of $(x(\varepsilon,c),u(\varepsilon,c),w(\varepsilon,c))$ also converge (component by component) to the first partial derivatives of $(x(0),u(0),w(0)) = (x^*,u^*,w^*)$. Actually, any procedure that determines the optimal multipliers of Problem $P(0)$ as $\varepsilon \to 0$ can be used. Unconstrained minimization of the augmented Lagrangian (in the appropriate neighborhood) will then yield an estimate of the local solution $x(0)$ of Problem $P(0)$, and the formulas given in (32) and (34) can be used to calculate the corresponding estimates of the first partial derivatives of the Kuhn-Tucker triple $(x(0),u(0),w(0))$ of Problem $P(0)$.

An alternative to the above procedures for estimating the desired sensitivity information is the penalty function procedure mentioned in the Introduction and developed rather extensively by Fiacco (1976) and Armacost and Fiacco (1974), (1975), (1976a) in terms of a well-known logarithmic-quadratic penalty function. It is shown that that local unconstrained minimization of the penalty function yields an estimate of the Kuhn-Tucker triple under the same conditions assumed throughout this paper. Also, analogous to the augmented Lagrangian result, the penalty function Hessian is shown to be positive definite in the appropriate neighborhood, so that one set of formulas (analogous to (32) and (34)) suffice to calculate the partial derivatives of the Kuhn-Tucker triple. There is no need to make a prior calculation of the optimal Lagrange multipliers or of any other information, only the unconstrained minimizing points (having preset a scalar parameter) normally required by the algorithm, to estimate the triple and its derivatives. Thus, the required effort is comparatively modest, with respect to the Lagrangian and augmented Lagrangian calculations,

and makes this procedure appealing as a pre-optimality sensitivity analysis estimation technique. This appeal is somewhat offset by the typical ill-conditioning that characterizes the penalty function Hessian near a local solution of the given problem. Thus, all the indicated approaches involve compensating factors, each offering advantages and disadvantages.

As a final observation, the expression developed above for $\nabla_{\varepsilon} y_a(\varepsilon,c)$ in Equations (32) and (34) can be placed in precise correspondence with the sensitivity expressions obtained by Buys and Gonin (1975) in their Equations (11) and (12). Our method of proof is simpler, however, utilizing as shown a result previously obtained for the usual Lagrangian. Further, the relationship between the Lagrangian and augmented Lagrangian calculations is demonstrated explicitly and allows application of all prior sensitivity results involving the usual Lagrangian.

Although a specific augmented Lagrangian function was used above to obtain the sensitivity results, the analysis and analogous results obtain for a more general function, such as the Arrow, Gould and Howe (1973) "extended" Lagrangian, which encompasses most of the popular functions of the augmented type. A more general concept, an "Acceptable Sequential Algorithm," (ASA) was proposed by Fiacco and developed by Armacost (1976b). It is noted that most popular penalty and barrier functions qualify as an ASA, so that sensitivity results anologous to those developed here can be obtained and will provide estimates of the sensitivity of the optimal value function and the Kuhn-Tucker triple.

REFERENCES

[1] ARMACOST, ROBERT L. (1976a). Computational experience with optimal value function and Lagrange multiplier sensitivity in NLP. Technical Paper Serial T-335. Institute for Management Science and Engineering, The George Washington University.

[2] ARMACOST, ROBERT L. (1976b). Sensitivity analysis in parametric nonlinear programming. Doctoral Dissertation, School of Engineering and Applied Science, The George Washington University.

[3] ARMACOST, ROBERT L. and FIACCO, ANTHONY V. (1974). Computational experience in sensitivity analysis for nonlinear programming. Mathematical Programming. 6 301-326.

[4] ARMACOST, ROBERT L. and FIACCO, ANTHONY V. (1975). Second-order parametric sensitivity analysis in NLP and estimates by penalty function methods. Technical Paper Serial T-324. Institute for Management Science and Engineering, The George Washington University.

[5] ARMACOST, ROBERT L. and FIACCO, ANTHONY V. (1976a). NLP sensitivity for RHS perturbations: A brief survey and recent second-order extensions. Technical Paper Serial T-334. Institute for Management Science and Engineering, The George Washington University.

[6] ARMACOST, ROBERT L. and FIACCO, ANTHONY V. (1976b). Sensitivity analysis for parametric nonlinear programming using penalty methods. Technical Paper Serial T-340. Institute for Management Science and Engineering, The George Washington University.

[7] ARMACOST, ROBERT L. and MYLANDER, W. CHARLES (1973). A guide to a SUMT-Version 4 computer subroutine for implementing sensitivity analysis in nonlinear programming. Technical Paper Serial T-287. Institute for Management Science and Engineering, The George Washington University.

[8] ARROW, K. J., GOULD, F. J. and HOWE, S. M. (1973). A general saddle point result for constrained optimization. Mathematical Programming. 5 225-234.

[9] BERTSEKAS, D. P. (1975). Combined primal-dual and penalty methods for constrained minimization. SIAM Journal on Control. (3) 13.

[10] BUYS, J. D. (1972). Dual algorithms for constrained optimization problems. Dissertation, University of Leiden, Leiden.

[11] BUYS, J. D. and GONIN, R. (1975). The use of augmented Lagrangian functions for sensitivity analysis in nonlinear programming. CSIR Special Report WISK 174. National Research Institute for Mathematical Sciences, Pretoria, South Africa.

[12] FIACCO, ANTHONY V. (1976). Sensitivity analysis for nonlinear programming using penalty methods. Mathematical Programming. 10 287-311.

[13] FIACCO, ANTHONY V. and McCORMICK, GARTH P. (1968). Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York.

[14] HESTENES, M. R. (1969). Multiplier and gradient methods. Journal of Optimization Theory and Applications. 4 303-320.

[15] McCORMICK, GARTH P. (1975). Unpublished notes and personal communication.

[16] MIELE, A., MOSELEY, P. A., LEVY, A. V. and COGGINS, G. M. (1972). On the method of multipliers for mathematical programming problems. Journal of Optimization Theory and Applications. (1) 10.

[17] ROBINSON, S. M. (1974). Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms. Mathematical Programming. 7 1-16.

[18] ROBINSON, S. M. (1973). Computable error bounds for nonlinear programming. Mathematical Programming. 5 235-242.

[19] ROCKAFELLAR, R. T. (1970). New applications of duality in convex programming. Presented at the 7th International Symposium on Mathematical Programming, The Hague, 1970. Published in the Proceedings of the 4th Conference on Probability.

[20] ROCKAFELLER, R. T. (1973). The multiplier method of Hestenes and Powell applied to convex programming. Journal of Optimization Theory and Applications. (6) 12.

[21] RUPP, R. D. (1976). Quadratic multiplier method convergence. Journal of Optimization Theory and Applications. (1) 20.