Sufficiency Conditions for Constrained Extrema

VICTOR J. LAW
Tulane University, New Orleans, Louisiana
and
ROBERT H. FARISS
Monsanto Company, Springfield, Massachusetts

Sufficiency conditions for equality constrained extrema are formulated in terms of a transformed set of coordinates by means of a generalized inverse of the constraint Jacobian matrix. This procedure produces a computationally simple and efficient method for testing sufficiency of a previously established stationary point. Computation of Lagrange multipliers is also facilitated by the method.

Recently Schechter and Beveridge (7) have presented sufficiency conditions for constrained extrema in terms of constrained derivatives. Their work is an extension of that of Edelbaum (2), who gave only necessary conditions, and is almost identical to that presented by Wilde and Beightler (8). This paper presents a slightly different formulation of the sufficiency conditions and proposes an efficient method for computationally implementing them.

The constrained optimization problem under consideration here is as follows:

$$\text{extremize } q(\mathbf{x}) \tag{1}$$

subject to

$$f_j(\mathbf{x}) = 0; \qquad j = 1, 2, \cdots, m \tag{2}$$

Equation (2) includes equality constraints as well as any active inequality constraints when sufficiency conditions are considered.

NECESSARY CONDITIONS

Suppose that $\mathbf{x}^\circ$ is a point at which necessary conditions for an extremum are satisfied. One formulation of these necessary conditions is as follows [see Edelbaum (2)]:

$$dq = \mathbf{g}^T\,\mathbf{dx} \tag{3}$$

where $\mathbf{dx}$ is restricted by

$$\mathbf{df} = \mathbf{J}\,\mathbf{dx} = \mathbf{0} \tag{4}$$

and

$$\mathbf{f} = \mathbf{0} \tag{5}$$

The conditions given by Equations (3) to (5) are equivalent to the Lagrange multiplier formulation (2). The Lagrangian function is given by

$$q_L = q + \mathbf{f}^T \boldsymbol{\lambda} \tag{6}$$

An important function of the Lagrange multiplier formulation is that the $\mathbf{dx}$ vector is no longer explicitly constrained by Equation (4). Thus, Equations (3) and (4) may be replaced by the more straightforward necessary condition

$$\mathbf{g} + \mathbf{J}^T \boldsymbol{\lambda} = \mathbf{0} \tag{7}$$

SUFFICIENT CONDITIONS

Phipps (5) has given a concise representation of the sufficiency conditions for constrained extrema. In particular, he showed that the second differential of the Lagrangian function

$$d^2 q_L = \mathbf{dx}^T \mathbf{H}_L\,\mathbf{dx} \tag{8}$$

subject to the constraints of Equation (4), provides a convenient sufficiency test. That is, if $d^2 q_L > 0$ for all $\mathbf{dx}$ subject to the restriction of Equation (4), then $\mathbf{x}^\circ$ is a minimum point. Likewise, if $d^2 q_L < 0$, $\mathbf{x}^\circ$ is a maximum point. If the sign of the restricted $d^2 q_L$ is indeterminant, then $\mathbf{x}^\circ$ is a saddle point. If $d^2 q_L = 0$, then higher derivatives must be considered. However, in practice the latter situation does not often result. The above analysis could be extended to treat such situations.

A NEW SUFFICIENCY FORMULATION

The major shortcomings of previously proposed computational procedures for establishing sufficiency are twofold. First, they all require the evaluation of a sequence of determinants. Second, they require that the variables be rearranged so that the first n columns of $\mathbf{J}$ are independent. The first requirement leads to an excessive amount of computation. The second one is more serious, however, in that it completely ignores the possibility of redundant constraints. In practice, constraint redundancy (within computational accuracy) often occurs and is not merely a complexity of theoretical interest. The computational procedure to be described here avoids both of the shortcomings mentioned above.

Consider Equation (4), which is rewritten below for convenience:

$$\mathbf{J}\,\mathbf{dx} = \mathbf{0} \tag{9}$$

If a solution to this linear algebraic equation set is found by means of a normalized generalized inverse of $\mathbf{J}$ (see Appendix), there results

$$\mathbf{dx}^c = \mathbf{T}_{n\times p}\mathbf{dy}_p \tag{10}$$

where $\mathbf{dy}_p$ is arbitrary.

Substitution of Equation (10) into Equation (8) gives

$$d^2 q_L^c = \mathbf{dy}_p^T\,\mathbf{T}_{n\times p}^T \mathbf{H}_L\,\mathbf{T}_{n\times p}\mathbf{dy}_p \tag{11}$$

Now, since $\mathbf{dy}_p$ is arbitrary, the sufficiency test consists only of examining the sign of $d^2 q_L^c$. This is most easily accomplished computationally by investigating the properties of the matrix

$$\mathbf{H}_L^P = \mathbf{T}_{n\times p}^T \mathbf{H}_L\,\mathbf{T}_{n\times p} \tag{12}$$

This can be done by diagonalizing $\mathbf{H}_L^P$ via elementary row and column operations (see Appendix). That is, consider a transformation S such that

$$\mathbf{S}^T \mathbf{H}_L^P \mathbf{S} = \mathbf{D}^P \quad (13)$$

where $\mathbf{D}^P$ is a diagonal matrix. The sufficiency test can then be stated as follows:

$$\text{if } d_{ii}^P \begin{cases} >0 \\ <0 \\ > \text{ or } <0 \end{cases} \text{then } \mathbf{x}^\circ \text{ is a } \begin{matrix} \text{minimum point} \\ \text{maximum point} \\ \text{saddle point} \end{matrix}$$

where $i = 1, 2, \cdots, p$.

COMPARISON OF COMPUTATIONAL REQUIREMENTS

The computational requirements of the presently proposed method for sufficiency evaluation are nominal in that one $n \times n$ symmetric matrix ($\mathbf{J}^T\mathbf{J}$) and one $p \times p$ symmetric matrix ($\mathbf{H}_L^P$) must be diagonalized. Diagonalization requires about the same number of operations as does matrix inversion.

The method presented by Phipps and the constrained derivative method require the evaluation of p determinants, the smallest of which is $(m + m) \times (m + m)$ and the largest of which is $(m + n) \times (m + n)$. A further complication is that a suitable arrangement of the variables must be made so that a known submatrix of the Jacobian is nonsingular. The net computational requirements are therefore considerably more severe than for the method proposed here.

COMPUTATION OF LAGRANGE MULTIPLIERS

An outgrowth of the use of a normalized generalized inverse of $\mathbf{J}$ to find a solution of Equation (9) is that the Lagrange multipliers may also easily be determined. This is important since λ must be known in order to form $\mathbf{H}_L$.

Consider Equation (7) written in rearranged form:

$$\mathbf{J}^T \lambda = -\mathbf{g} \quad (14)$$

where here $\mathbf{J}$ and $\mathbf{g}$ are evaluated at $\mathbf{x}^\circ$. A consistent set of multipliers is then given by

$$\lambda = -(\mathbf{J}^n)^T \mathbf{g} \quad (15)$$

Example 1

Consider the problem

$$\text{extremize} \quad q = x_1 x_2$$

subject to

$$f_1 = x_1^2 + x_2^2 - 1 = 0$$
$$f_2 = 2x_1^2 + 2x_2^2 - 2 = 0$$

The second constraint is obviously redundant with the first. This has been done purposefully in order to illustrate the ability of the proposed method to handle constraint redundancy. There are four points at which necessary conditions for a constrained extremum are satisfied. Only one of these will be considered for purposes of this illustration. In particular

$$\mathbf{x}^\circ = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}$$

$$\mathbf{J} = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{bmatrix}$$

Further

$$\mathbf{J}^T \mathbf{J} = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

A transformation $\mathbf{T}$ which diagonalizes $\mathbf{J}^T\mathbf{J}$ is given by

$$\mathbf{T} = \begin{bmatrix} 1 & \vdots & -1 \\ 0 & \vdots & 1 \end{bmatrix} = [\mathbf{T}_{n \times r} \vdots \mathbf{T}_{n \times p}]$$

A normalized generalized inverse of $\mathbf{J}$ may then be calculated from Equation (A4) as

$$\mathbf{J}^n = \begin{bmatrix} \frac{\sqrt{2}}{4} & \frac{\sqrt{2}}{4} \\ 0 & 0 \end{bmatrix}$$

From Equation (15)

$$\lambda = \begin{bmatrix} -\frac{1}{4} \\ -\frac{1}{4} \end{bmatrix}$$

The Hessian of the Lagrangian function is then given by

$$\mathbf{H}_L = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

Then, according to Equation (12)

$$\mathbf{H}_L^P = [-1 \quad 1] \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -4$$

Therefore $\mathbf{x}^\circ$ is a maximum point.

Example 2

Consider the problem where

$$q = -1.2338\, x_1^2 - 0.0203\, x_2^2 - 0.0136\, x_3^2 - 0.0027\, x_4^2 + 0.0031\, x_3 x_4 + 2139\, x_1 + 135\, x_2 + 103\, x_3 + 19\, x_4$$

and the constraints are given by

$$3.21\, x_1 + 1.62\, x_2 + 0.38\, x_3 + 0.02\, x_4 = 8{,}050$$
$$1.63\, x_1 + 0.03\, x_2 + 0.06\, x_3 + 0.02\, x_4 = 1{,}000$$
$$6.47\, x_1 + 1.68\, x_2 + 0.50\, x_3 + 0.06\, x_4 = 10{,}050$$

A known stationary point is

$$\mathbf{x}^\circ = \begin{bmatrix} 423 \\ 3{,}454 \\ 2{,}760 \\ 2{,}988 \end{bmatrix}$$

Proceeding according to the proposed method

$$\mathbf{J}^T \mathbf{J} = \begin{bmatrix} 54.8219 & 16.1187 & 4.5526 & 0.4850 \\ 16.1187 & 5.4477 & 1.4574 & 0.1338 \\ 4.5526 & 1.4574 & 0.3980 & 0.0388 \\ 0.4850 & 0.1338 & 0.0388 & 0.0044 \end{bmatrix}$$

A transformation which diagonalizes $\mathbf{J}^T\mathbf{J}$ according to Equation (A5) is given by

$$\mathbf{T} = \begin{bmatrix} 1.000 & -0.2940 & -0.0337 & -0.0125 \\ 0 & 1.0 & -0.1677 & 0.0124 \\ 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$$

The resulting diagonal matrix is

$$\mathbf{D} = \begin{bmatrix} 54.8219 & 0 & 0 & 0 \\ 0 & 0.7085 & 0 & 0 \\ 0 & 0 & 3 \times 10^{-7} & 0 \\ 0 & 0 & 0 & 6 \times 10^{-8} \end{bmatrix}$$

It is clear that the last two diagonal elements of **D** are zero within machine accuracy. Therefore the rank of **J** is two and a constraint redundancy exists.

Since the constraints are linear, their second derivatives are all zero and the Lagrange multipliers are not needed for the sufficiency calculations. Thus

$$\mathbf{H}_L = \begin{bmatrix} -2.4676 & 0 & 0 & 0 \\ 0 & -0.0406 & 0 & 0 \\ 0 & 0 & -0.0271 & 0.0031 \\ 0 & 0 & 0.0031 & -0.0054 \end{bmatrix}$$

Computing $\mathbf{H}_L^P$ from Equation (12) yields

$$\mathbf{H}_L^P = \begin{bmatrix} -0.03115 & 0.0021 \\ 0.0021 & -0.0058 \end{bmatrix}$$

Upon diagonalizing $\mathbf{H}_L^P$ by the transformation

$$\mathbf{S} = \begin{bmatrix} 1.0 & 0.0688 \\ 0 & 1 \end{bmatrix}$$

there results

$$\mathbf{D}^P = \begin{bmatrix} -0.0311 & 0 \\ 0 & -0.0056 \end{bmatrix}$$

Since both diagonal elements of $\mathbf{D}^P$ are negative, then the stationary point is a maximum.

CONCLUSIONS

A formulation of sufficiency conditions for constrained extrema has been given which is computationally efficient and simple. Constraint redundancy does not hamper the method and only an algorithm for diagonalizing a symmetric matrix is required. Inequality constraints can also be considered in that any active inequalities can be treated simply as equalities for purposes of the sufficiency test. An outgrowth of the method allows easy determination of the Lagrange multipliers of the problem.

NOTATION

$\mathbf{D}$ = diagonal matrix
$\mathbf{D}^P$ = diagonal matrix
d_{ii} = i^{th} diagonal element of $\mathbf{D}$
d_{ii}^P = i^{th} diagonal element of $\mathbf{D}^P$
$\mathbf{f}$ = $m \times 1$ vector of constraint functions
$\mathbf{df}$ = $m \times 1$ vector of constraint first differentials
$\mathbf{d^2f}$ = $m \times 1$ vector of constraint second differentials
dq = first differential of q
d^2q_L = second differential of q
$d^2q_L^c$ = constrained second differential of the Lagrangian function
$\mathbf{dx}$ = $n \times 1$ vector of first-order differential perturbation
$\mathbf{dx}^c$ = $\mathbf{dx}$ vector subjected to the constraint of Equation (4)
$\mathbf{dy}$ = transformed $\mathbf{dx}$ vector = $\mathbf{T}^{-1}\mathbf{dx}$
$\mathbf{dy}_r$ = first r elements of $\mathbf{dy}$
$\mathbf{dy}_p$ = last p elements of $\mathbf{dy}$
$\mathbf{J}$ = $m \times n$ Jacobian matrix, $J_{ij} = \partial f_i/\partial x_j$
$\mathbf{g}$ = $n \times 1$ gradient vector, $\partial q/\partial \mathbf{x}$
$\mathbf{H}$ = $n \times n$ Hessian of q, $\partial^2 q/\partial \mathbf{x}^2$
$\mathbf{H}_L$ = $n \times n$ Hessian of q_L, $\partial^2 q_L/\partial \mathbf{x}^2$
$\mathbf{H}_L^P$ = $p \times p$ transformed Hessian
m = number of constraints
n = number of variables
$\mathbf{O}$ = null matrix
p = $n - r$
q = objective function to be extremized
q_L = Lagrangian function
r = rank of $\mathbf{J}$
$\mathbf{S}$ = $p \times p$ transformation matrix
$\mathbf{T}$ = $n \times n$ transformation matrix
$\mathbf{T}_{n\times r}$ = first r columns of $\mathbf{T}$
$\mathbf{T}_{n\times p}$ = last p columns of $\mathbf{T}$
$\mathbf{x}$ = $n \times 1$ vector of variables
$\mathbf{x}^\circ$ = constrained stationary point
λ = $m \times 1$ vector of Lagrange multipliers

Superscript

n = normalized generalized inverse of a matrix, for example, $\mathbf{J}^n$

LITERATURE CITED

1. Birkhoff, G., and S. MacLane, "Survey of Modern Algebra," p. 276, Macmillan, New York (1941).
2. Edelbaum, T. N., "Optimization Techniques," G. Leitmann, ed., Chap. 1, Academic Press, New York (1962).
3. Moore, E. H., *Bull. Am. Math. Soc.*, **26**, 394 (1920).
4. Penrose, R., *Proc. Cambridge Phil.Soc.*, **51**, 406 (1955).
5. Phipps, C. G., *Am. Math. Monthly*, **59**, 230 (1952).
6. Rhode, C. A., *J.S.I.A.M.*, **13**, 1033 (1965).
7. Schechter, R. S., and G. S. G. Beveridge, *Ind. Eng. Chem. Fundamentals*, **5**, 571 (1966).
8. Wilde, D. J., and C. S. Beightler, "Foundations of Optimization," p. 52, Prentice Hall, Englewood Cliffs, N.J. (1967).
9. *Ibid.*, p. 81.

APPENDIX

The concept of generalized inverses of singular or rectangular matrices has received considerable attention since first introduced by Moore (*3*) and later, independently, by Penrose (*4*). Rhode (*6*) has recently discussed several types of generalized inverses. A particularly useful one in the context of sufficiency conditions is the so-called normalized generalized inverse (NGI). For any matrix **A**, the NGI of **A** satisfies the following properties:

$$\mathbf{A}\mathbf{A}^n\mathbf{A} = \mathbf{A} \quad (A1)$$

$$\mathbf{A}^n\mathbf{A}\mathbf{A}^n = \mathbf{A}^n \quad (A2)$$

$$(\mathbf{A}\mathbf{A}^n)^T = \mathbf{A}\mathbf{A}^n \quad (A3)$$

It may be shown that an $\mathbf{A}^n$ can be computed by the relationship

$$\mathbf{A}^n = \mathbf{T}_{n\times r}\mathbf{D}_r^{-1}\mathbf{T}_{n\times r}^T\mathbf{A}^T \quad (A4)$$

where **T** is a transformation which diagonalizes the symmetric matrix $\mathbf{A}^T\mathbf{A}$. That is

$$\mathbf{T}^T\mathbf{A}^T\mathbf{A}\mathbf{T} = \mathbf{D} \quad (A5)$$

$\mathbf{A}^n$ may be used to solve linear algebraic equations of the form

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (A6)$$

In particular, a general solution to Equation (*A6*) is given by

$$\mathbf{x} = \mathbf{A}^n\mathbf{b} + \mathbf{T}_{n\times p}\mathbf{y}_p \quad (A7)$$

where $\mathbf{y}_p$ is arbitrary. An additional useful relationship is that

$$\mathbf{x} = \mathbf{T}\mathbf{y} \quad (A8)$$

The basic reason why the NGI is preferred for sufficiency condition calculations is that a symmetric matrix diagonalization algorithm is needed to investigate the properties of $\mathbf{H}_L^P$ [see Equation (13)]. The same algorithm may therefore be used to compute $\mathbf{J}^n$. Wilde and Beightler (*8*) describe one possible algorithm for diagonalizing a symmetric matrix. Birkhoff and MacLane (*1*) prove that **T** is nonsingular since it is constructed as a sequence of elementary row column operations.

Manuscript received January 1, 1969; revision received August 5, 1969; paper accepted August 28, 1969. Paper presented at AIChE New Orleans meeting.