Faculty Working Papers

GENERALIZED CONSTRAINED DERIVATIVES

Ronald J. Stern

#63

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

September 12, 1972

GENERALIZED CONSTRAINED DERIVATIVES

Ronald J. Stern

#63

To be submitted to the Journal of Operations Research

GENERALIZED CONSTRAINED DERIVATIVES

by

Ronald J. Stern*

ABSTRACT

In this paper we deal with a general equality constrained minimization problem. The concept of the constrained derivative as due to Wilde et al., [5] - [8] is extended by considering more general partitions of the vector of unknowns into state and decision components and utilizing the properties of the generalized matrix inverse. The main result is a generalization of a sufficient condition of Wilde, guaranteeing the existence of a local minimum.

*Department of Business Administration
University of Illinois
Urbana, Illinois 61801

GENERALIZED CONSTRAINED DERIVATIVES

In equality constrained minimization problems, a test for a local minimum is to verify the stationarity of a Lagrangian and the definiteness of a certain quadratic form. Using the concepts of state and decision variables and constrained derivatives Wilde et al. [5] - [8] showed that stationarity of the Lagrangian and definiteness of a quadratic form of reduced dimension is sufficient. In this paper we introduce the concept of a generalized constrained derivative and derive a generalization of Wilde's sufficient condition. The result obtained here does not depend on the local existence of a non-singular sub-Jacobian of the constraint Jacobian, but rather a sub-Jacobian of maximal full column rank.

1. NOTATION AND PRELIMINARIES

Let A be a real mxn matrix. We denote the <u>range</u> of A by R(A) and the <u>nullspace</u> of A by N(A). The <u>transpose</u> of A is denoted by A^*. The <u>generalized inverse</u> of A is denoted by A^+. If A is square and non-singular then $A^+ = A^{-1}$, the <u>inverse</u> of A.

For references on the generalized inverse the reader is referred to e.g. [1] and [2]. A^+ is uniquely defined for any (mxn) real matrix A, and computational methods for determining it may be found for example in [2]. Some particular properties of A^+ relevant to this paper are now given.

(1.1) $AA^+ = P_{R(A)}$, <u>the projection onto R(A)</u>.

(1.2) $I - A^+A = P_{N(A)}$, <u>the projection onto N(A)</u>.

Projection maps are discussed in [4].

In view of (1.1) and (1.2) we have the following result on linear equations: If $y \varepsilon R(A)$ then the general solution to $Ax = y$ is given by

(1.3) $\quad x = A^{+}y + (I-A^{+}A)z \quad$ for any $z \varepsilon R^n$.

We will make use of the following analytic notation. Let $h: R^n \to R^m$. Let $|| \quad ||$ be the Eucledian norm. Then

$$h(t) = o|| t ||^k \text{ means } \lim_{t \to 0} \frac{h(t)}{|| t ||^k} = 0.$$

The Jacobian of h at t_o, if it exists, will be denoted $\left(\frac{\partial h}{\partial t}\right)_{t_o}$. Here

$$h = \begin{pmatrix} h_1 \\ h_2 \\ \cdot \\ \cdot \\ \cdot \\ hm \end{pmatrix}$$

and the Jacobian is the (mxn) matrix $\left(\frac{\partial h}{\partial t}\right)_{t_o} = \left(\frac{\partial h_i(t_o)}{\partial t_j}\right)$ where $(i = 1,2,\ldots,m)$, $(j = 1,2,\ldots,n)$ and $t_o \varepsilon R^n$.

If $m = 1$ we write $\left(\frac{\partial h}{\partial t}\right)_{t_o} = \nabla_t h(t_o)$, the usual gradient notation.

Also, when $m = 1$, the Hessian of h at t_o, if it exists, is the (nxn) matrix $\left(\frac{\partial^2 h(t_o)}{\partial t_i \partial t_j}\right)$ where $(i = 1,2,\ldots n)$ and $(j = 1,2,\ldots,n)$.

2. CONSTRAINED DERIVATIVES

In this section a brief review of the concepts to be generalized will be presented. The problem which we will consider, denoted $\overline{P}$, is given by

$\overline{P}$ minimize $y(x)$

subject to $\quad f_i(x) = 0 \qquad (i = 1,\ldots,m)$

$\qquad x \varepsilon R^n$

The functions $y, f_1, \ldots, f_m$ are assumed to be twice differentiable on R^n. The constraints are also written $f(x) = 0$.

Partition the vector $x = \left(\frac{s}{d}\right)$, where $s \epsilon R^m$ is referred to as the <u>state vector</u> and $d \epsilon R^{n-m}$ is called the <u>decision vector</u>. Assume a feasible point x_o is such that the Jacobian $\left(\frac{\partial f}{\partial s}\right)_{x_o}$ is non-singular. By Taylor's theorem

$$(2.1) \quad \partial y = \nabla_d y(x_o) \partial d + \nabla_s y(x_o) \partial s + o||\partial x||$$

If x_o is a feasible point then by Taylor's theorem the feasibility of $x_o + \partial x$ is equivalent to

$$(2.2) \quad \left(\frac{\partial f}{\partial s}\right)_{x_o} \partial s + \left(\frac{\partial f}{\partial d}\right)_{x_o} \partial d + o||\partial x|| = 0$$

Since $\left(\frac{\partial f}{\partial s}\right)_{x_o}$ is invertible (2.2) yields

$$(2.3) \quad \partial s = - \left(\frac{\partial f}{\partial s}\right)^{-1}_{x_o} \left(\frac{\partial f}{\partial d}\right)_{x_o} \partial d - o||\partial x||$$

The <u>constrained derivative</u> of y at x_o with respect to the partition $\left(\frac{s}{d}\right)$ is defined by

$$(2.4) \quad \left(\frac{\delta y}{\delta d}\right)_{x_o} = \nabla_d y(x_o) - \nabla_s y(x_o) \left(\frac{\partial f}{\partial s}\right)^{-1}_{x_o} \left(\frac{\partial f}{\partial d}\right)_{x_o}$$

From (2.1) - (2.4) we obtain

$$(2.5) \quad \partial y = \left(\frac{\delta y}{\delta d}\right)_{x_o} \partial d + o||\partial x||$$

Since ∂d is free to range over R^{n-m} we have the following result: <u>A necessary condition for x_o to be a local minimum of $\overline{P}$ is that</u> $\left(\frac{\delta y}{\delta d}\right)_{x_o} = 0$.

Again consider problem $\overline{P}$ but now assume $y, f_1, \ldots, f_m$ are three times differentiable. Let H denote the Hessian of the objective function and let H_i denote the Hessian of the ith constraint at x_o. For any set of numbers $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ we have

$$(2.6) \qquad \partial y = \frac{\partial}{\partial x}\left[y - \sum_{i=1}^{m} \lambda_i f_i\right] \partial x + \frac{1}{2}\,\partial x^* \left(H - \sum_{i=1}^{m} \lambda_i H_i\right) \partial x + o||\partial x||^2$$

If $(\overline{\lambda}_1, \ldots \overline{\lambda}_m)$ are <u>Lagrange multipliers</u>, that is, satisfy

$$(2.7) \qquad \nabla y(x_o) = \sum_{i=1}^{m} \overline{\lambda}_i \nabla f_i(x_o)$$

then

$$(2.8) \qquad \partial y = \frac{1}{2}\,\partial x^* \left(H - \sum_{i=1}^{m} \overline{\lambda}_i H_i\right) \partial x + o||\partial x||^2 .$$

Notice that a unique vector of Lagrange multipliers is guaranteed to exist at x_o, by our assumption on $\left(\frac{\partial f}{\partial s}\right)_{x_o}$.

Write $P = H - \sum_{i=1}^{m} \overline{\lambda}_i H_i$. Appropriate partitioning P, we have

$$(2.9) \qquad \partial y = \frac{1}{2}\,(\partial d^* \; \partial s^*) \left(\begin{array}{c|c} P_{dd} & P_{ds} \\ \hline P_{ds} & P_{ss} \end{array}\right) \left(\begin{array}{c} \partial d \\ \hline \partial s \end{array}\right) + o||\partial x||^2 .$$

Upon defining

$$S_{dd} = P_{dd} - P_{ds} \left(\frac{\partial f}{\partial s}\right)^{-1}_{x_o} \left(\frac{\partial f}{\partial d}\right)_{x_o} - \left(P_{ds} \left(\frac{\partial f}{\partial s}\right)^{-1}_{x_o} \left(\frac{\partial f}{\partial d}\right)_{x_o}\right)^*$$

$$+ \left(\left(\frac{\partial f}{\partial s}\right)^{-1}_{x_o} \left(\frac{\partial f}{\partial d}\right)_{x_o}\right)^* P_{ss} \left(\frac{\partial f}{\partial s}\right)^{-1}_{x_o} \left(\frac{\partial f}{\partial d}\right)_{x_o}$$

and making use of (2.3) and (2.9) we obtain

$$(2.10) \qquad \partial y = \frac{1}{2}\,\partial d^* S_{dd} + o||\partial x||^2$$

which yields the following result: <u>A sufficient condition for x_o to be a local minimum of problem $\overline{P}$ is that the Lagrangian</u> $y - \sum_{i=1}^{m} \overline{\lambda}_i f_i$ <u>is stationary at x_o and S_{dd} is positive definite.</u>

All the material in section 2 can be found in [7].

3. GENERALIZED CONSTRAINED DERIVATIVES

Again consider problem $\overline{P}$, with the functions $y, f_1, \ldots, f_m$ being twice differentiable where x_o is a feasible point. Now take a partition $y = \begin{pmatrix} s \\ - \\ d \end{pmatrix}$ where <u>s is not necessarily an m-vector</u>.

An <u>auxiliary problem</u>, denoted $\overline{P}_{x_o}$, is now introduced:

$$\overline{P}_{x_o} \qquad \underset{x \in R^n}{\text{minimize}}\ y(x) \quad \text{subject to} \quad \left(\frac{\partial f}{\partial s}\right)_{x_o} \left(\frac{\partial f}{\partial s}\right)^+_{x_o} f(x) = 0$$

Assume that $\left(\frac{\partial f}{\partial s}\right)_{x_o}$ is a maximal full column rank submatrix of $\left(\frac{\partial f}{\partial x}\right)_{x_o}$.

Feasibility of $x_o + \partial_x$ for problem $\overline{P}_{x_o}$ is equivalent to

$$(3.1) \quad \left(\frac{\partial f}{\partial s}\right)_{x_o} \left(\frac{\partial f}{\partial s}\right)^+_{x_o} \left(\frac{\partial f}{\partial s}\right)_{x_o} \partial s + \left(\frac{\partial f}{\partial s}\right)_{x_o} \left(\frac{\partial f}{\partial s}\right)^+_{x_o} \left(\frac{\partial f}{\partial d}\right)_{x_o} \partial d$$

$$+ \left(\frac{\partial f}{\partial s}\right)_{x_o} \left(\frac{\partial f}{\partial s}\right)^+_{x_o} o||\partial x|| = 0$$

Since $\left(\frac{\partial f}{\partial s}\right)_{x_o} \left(\frac{\partial f}{\partial s}\right)^+_{x_o} = P_{R\left(\left(\frac{\partial f}{\partial s}\right)_{x_o}\right)}$ and by the above assumption on $\left(\frac{\partial f}{\partial s}\right)_{x_o}$, (3.1) may be rewritten as follows:

$$(3.2) \quad \left(\frac{\partial f}{\partial s}\right)_{x_o} \partial s + \left(\frac{\partial f}{\partial d}\right)_{x_o} \partial d + \left(\frac{\partial f}{\partial s}\right)_{x_o} \left(\frac{\partial f}{\partial s}\right)^+ o||\partial x|| = 0$$

Since all three terms in this equation are in $R\left(\frac{\partial f}{\partial s}\right)_{x_o}$ we can solve for ∂s and obtain

$$(3.3) \qquad \partial s = -\left(\frac{\partial f}{\partial s}\right)^{+}_{x_o} \left(\frac{\partial f}{\partial d}\right)_{x_o} \partial d - o\|\partial x\|$$

The <u>generalized constrained derivative</u> of y at x_o with respect to the partition $\left(\frac{s}{d}\right)$ is defined by

$$(3.4) \qquad \left(\frac{\delta y}{\delta d}\right)_{x_o} = \nabla_d y(x_o) - \nabla_s y(x_o) \left(\frac{\partial f}{\partial s}\right)^{+}_{x_o} \left(\frac{\partial f}{\partial d}\right)_{x_o}$$

Similarly to (2.5) we have

$$(3.5) \qquad \partial y = \left(\frac{\delta y}{\delta d}\right)_{x_o} \partial d + o\|\partial x\|$$

and the following fact:

<u>A necessary condition for x_o to be a local minimum of $\bar{P}_{x_o}$ is that</u> $\left(\frac{\delta y}{\delta d}\right)_{x_o} = 0.$

It is <u>not</u> necessarily true that the generalized constrained derivative = 0 at a local minimum of problem $\bar{P}$, however, as is evidenced by the following counterexample. Consider the problem

$$\text{minimize } y(x) = x_1^2 + x_2^2 + x_3$$

$$\text{subject to} \quad f_1(x) = x_1 - x_2^2 - x_3^2 = 0$$

$$f_2(x) = x_3^2 = 0$$

The solution obviously is (0,0,0) and $\left(\frac{\partial f}{\partial x}\right)_{x_o} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Upon setting $s = x_1$, $d = \begin{pmatrix} x_2 \\ x_3 \end{pmatrix}$ we obtain $\left(\frac{\delta y}{\delta d}\right) = (0,1)$.

This occurs for the following reason: In expression (2.2), when $\left(\frac{\partial f}{\partial s}\right)_{x_o}$ is not square, the terms $o\|\partial x\|$ need not be in $R\left(\frac{\partial f}{\partial s}\right)_{x_o}$, and one

cannot proceed to solve (2.2) for ∂s by using a generalized inverse. This, it can be verified, is precisely the case in the above counter-example. This difficulty cannot occur, of course, if the constraints in problem $\overline{P}$ are <u>linear</u>; that is, $f(x) = Ax-b$. The generalized constrained derivative for a problem with linear constraints is

$$\left(\frac{\overline{\delta y}}{\delta d}\right) = \nabla_d y(x_o) - \nabla_s y(x_o)\, A_s^{+} A_d$$

where $\quad A = (A_s \mid A_d) = \left(\frac{\partial f}{\partial x}\right).$

<u>For the case where $\overline{P}$ has linear constraints it is necessary that the generalized constrained derivative be 0 at a local minimum.</u> If $y(x)$ is linear and A_s is non-singular, then the last statement is equivalent to saying that the simplex criterion usually denoted by $z_j - c_j$ is 0 for variables x_j in the basis. Different partitions $\left(\frac{s}{d}\right)$ correspond to different vertices of the feasible region. The state variables are the basic variables and the decision variables are the variables not in the basis. (See e.g. [3]).

Observe that

(3.6) $\quad x_o$ is a local minimum solution of $\overline{P}_{x_o} \Longrightarrow x_o$ is a local minimum solution of $\overline{P}$

Thus a sufficient condition for x_o to be a local minimum of $\overline{P}_{x_o}$ is sufficient condition for $\overline{P}$ as well. Such a condition will now be derived.

Let x_o be a feasible point for problem $\overline{P}$. It is clearly then feasible for $\overline{P}_{x_o}$. Again assume that $\left(\frac{\partial f}{\partial s}\right)_{x_o}$ is a maximal full column rank submatrix of $\left(\frac{\partial f}{\partial x}\right)_{x_o}$. Also, assume now that the functions $y, f_1, \ldots, f_m$ are three times differentiable. Assume

(3.7) $\quad \nabla y(x_o) \in R \left(\frac{\partial f}{\partial x}\right)^*_{x_o}$

This implies the existence of a (not necessarily unique) vector of Lagrange multipliers $(\bar{\lambda}_1, \ldots, \bar{\lambda}_m)$.

Following the development from (2.6) - (2.10), using (3.3) in place of (2.3), and (3.6), we see that <u>a sufficient condition for</u> x_o <u>to be a local minimum of $\bar{P}$ is that</u> (3.7) <u>holds and</u> S'_{dd} <u>is positive definite. Here</u> S'_{dd} <u>is the square matrix</u>

$$S'_{dd} = P_{dd} - P_{ds}\left(\frac{\partial f}{\partial s}\right)^+_{x_o}\left(\frac{\partial f}{\partial d}\right)_{x_o} - \left(P_{ds}\left(\frac{\partial f}{\partial s}\right)^+_{x_o}\left(\frac{\partial f}{\partial d}\right)_{x_o}\right)^*$$
$$+ \left(\left(\frac{\partial f}{\partial s}\right)^+_{x_o}\left(\frac{\partial f}{\partial d}\right)_{x_o}\right)^* P_{ss}\left(\frac{\partial f}{\partial s}\right)^+_{x_o}\left(\frac{\partial f}{\partial d}\right)_{x_o}$$

<u>where</u> $P = H - \sum_{i=1}^{m} \bar{\lambda}_i H_i$ <u>is appropriately partitioned.</u>

4. CONCLUDING REMARKS

Upon locating a non-singular sub-Jacobian of $\left(\frac{\partial f}{\partial x}\right)_{x_o}$ and verifying that the Lagrangian of problem $\bar{P}$ is stationary at x_o for a Lagrange multiplier $\bar{\lambda}$, Wilde has defined a matrix of "constrained second derivatives," S_{dd}, whose definition involves $\bar{\lambda}$. S_{dd} is of dimension (n-m x n-m), where m is the number of constraints. The definiteness of S_{dd} guarantees that x_o is a local extremum of problem P.

In this paper we have treated the case where a non-singular sub-Jacobian may not exist, but we considered maximal full column-rank sub-Jacobians. By introducing an "artificial" problem, $\bar{P}_{\bar{x}_o}$, and using the properties of the generalized matrix inverse, we defined S'_{dd}, which is an

(n-p x n-p) analog of S_{dd}. Here $p \leq m$. Under stationarity of the Lagrangian of $\bar{F}$ the definiteness of S'_{dd} guarantees a local extremum. It can be seen, by following the arguments of section 3, that if $p = m$, then our result is precisely that of Wilde.

ACKNOWLEDGEMENT

The author wishes to thank Professor Adi Ben-Israel for discussions on this subject. Comments and corrections made by the preliminary referee were very helpful. The example in section 3 is due to his report.

REFERENCES

[1] A. Ben-Israel and A. Charnes, "Contributions to the Theory of Generalized Inverses," J. SIAM, 11, (1963), 667-699.

[2] A. Ben-Israel and D. Cohen, "On Iterative Computation of Generalized Inverses and Associated Projections," J. SIAM Num. An. , 3 (1966, 410-419.

[3] S. Gass, Linear Programming, McGraw-Hill, New York, 1964.

[4] P. Halmos, Finite-Dimensional Vector Spaces, Van Nostrand, Princeton, 1958.

[5] D. Wilde, "Differential Calculus in Nonlinear Programming," Operations Research, 10 (1962), 764-773.

[6] D. Wilde, "Jacobians in Constrained Nonlinear Optimization," Operations Research, 13 (1965), 848-856.

[7] D. Wilde & C. Beightler, Foundations of Optimization, Prentice Hall, Englewood Cliffs, 1967.

[8] D. Wilde & G. Reklaitis, "A Computationally Compact Sufficient Condition for Constrained Minima," Operations Research, 17 (1969), 425-235.