GENERALIZED CONSTRAINED DERIVATIVES

Ronald J. Stern

#63

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign
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 range of A by R(A) and the nullspace of A by N(A). The transpose of A is denoted by A*. The generalized inverse of A is denoted by A⁺. If A is square and non-singular then A⁺ = A⁻¹, the inverse 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) \[ AA⁺ = P_{R(A)}, \text{ the projection onto } R(A). \]

(1.2) \[ I - A⁺A = P_{N(A)}, \text{ the projection onto } N(A). \]

Projection maps are discussed in [4].
In view of (1.1) and (1.2) we have the following result on linear equations: If \( y \in \mathbb{R}(A) \) then the general solution to \( Ax = y \) is given by

\[
(1.3) \quad x = A^+ y + (I - A^+ A)z \quad \text{for any } z \in \mathbb{R}^n.
\]

We will make use of the following analytic notation. Let \( h : \mathbb{R}^n \rightarrow \mathbb{R}^m \). Let \( \| \cdot \| \) be the Euclidean norm. Then

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

The Jacobian of \( h \) at \( t_0 \), if it exists, will be denoted \( \left( \frac{\partial h}{\partial t} \right)_{t_0} \). Here

\[
h = \begin{pmatrix}
h_1 \\
h_2 \\
\vdots \\
h_m
\end{pmatrix}
\]

and the Jacobian is the \((m \times n)\) matrix \( \left( \frac{\partial h}{\partial t} \right)_{t_0} = \begin{pmatrix} \frac{\partial h_1(t_0)}{\partial t_i} \\ \frac{\partial h_2(t_0)}{\partial t_i} \\ \vdots \\ \frac{\partial h_m(t_0)}{\partial t_i} \end{pmatrix} \) where \( (i = 1, 2, \ldots, m) \), \( (j = 1, 2, \ldots, n) \) and \( t \in \mathbb{R}^n \).

If \( m = 1 \) we write \( \left( \frac{\partial h}{\partial t} \right)_{t_0} = \nabla_t h(t_0) \), the usual gradient notation.

Also, when \( m = 1 \), the Hessian of \( h \) at \( t_0 \), if it exists, is the \((n \times n)\) matrix \( \begin{pmatrix} \frac{\partial^2 h(t_0)}{\partial t_i \partial t_j} \end{pmatrix} \) 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 \( \mathcal{P} \), is given by

\[
\mathcal{P} : \quad \text{minimize } y(x) \quad \text{subject to } \quad \begin{array}{l}
f_i(x) = 0 \quad (i = 1, \ldots, m) \\
x \in \mathbb{R}^n
\end{array}
\]
The functions $y, f_1, \ldots, f_m$ are assumed to be twice differentiable on $\mathbb{R}^n$. The constraints are also written $f(x) = 0$.

Partition the vector $x = \begin{pmatrix} s \\ d \end{pmatrix}$, where $s \in \mathbb{R}^m$ is referred to as the state vector and $d \in \mathbb{R}^{n-m}$ is called the decision vector. Assume a feasible point $x_0$ is such that the Jacobian $\begin{pmatrix} \frac{\partial f}{\partial s} \end{pmatrix}_{x_0}$ is non-singular. By Taylor's theorem

$$(2.1) \quad \partial y = \nabla_d y(x_0) \partial d + \nabla_s y(x_0) \partial s + o|\partial x|$$

If $x_0$ is a feasible point then by Taylor's theorem the feasibility of $x_0 + \partial x$ is equivalent to

$$(2.2) \quad \begin{pmatrix} \frac{\partial f}{\partial s} \\ \frac{\partial f}{\partial d} \end{pmatrix}_{x_0} \partial s + \begin{pmatrix} \frac{\partial f}{\partial d} \end{pmatrix}_{x_0} \partial d + o|\partial x| = 0$$

Since $\begin{pmatrix} \frac{\partial f}{\partial s} \end{pmatrix}_{x_0}$ is invertible (2.2) yields

$$(2.3) \quad \partial s = - \begin{pmatrix} \frac{\partial f}{\partial s} \end{pmatrix}_{x_0}^{-1} \begin{pmatrix} \frac{\partial f}{\partial d} \end{pmatrix}_{x_0} \partial d - o|\partial x|$$

The constrained derivative of $y$ at $x_0$ with respect to the partition $\begin{pmatrix} s \\ d \end{pmatrix}$ is defined by

$$(2.4) \quad \begin{pmatrix} \frac{\delta y}{\delta d} \end{pmatrix}_{x_0} = \nabla_d y(x_0) - \nabla_s y(x_0) \begin{pmatrix} \frac{\partial f}{\partial s} \end{pmatrix}_{x_0}^{-1} \begin{pmatrix} \frac{\partial f}{\partial d} \end{pmatrix}_{x_0}$$

From (2.1) - (2.4) we obtain

$$(2.5) \quad \partial y = \begin{pmatrix} \frac{\delta y}{\delta d} \end{pmatrix}_{x_0} \partial d + o|\partial x|$$

Since $\partial d$ is free to range over $\mathbb{R}^{n-m}$ we have the following result:

A necessary condition for $x_0$ to be a local minimum of $\overline{F}$ is that $\begin{pmatrix} \frac{\delta y}{\delta d} \end{pmatrix}_{x_0} = 0$. 

-3-
Again consider problem $P$ 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 $i$th constraint at $x_o$. For any set of numbers $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ we have

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

If $(\lambda_1, \ldots, \lambda_m)$ are Lagrange multipliers, that is, satisfy

$$
(2.7) \quad \nabla y(x_o) = \sum_{i=1}^{m} \lambda_i \nabla f_i(x_o)
$$

then

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

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

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

$$
(2.9) \quad \partial y = \frac{1}{2} \left( \partial d^* \partial s^* \right) \left( \begin{array}{c} \frac{\partial d}{\partial d} \frac{\partial s}{\partial s} \\ \frac{\partial d}{\partial s} \frac{\partial s}{\partial s} \end{array} \right) \left( \begin{array}{c} \partial d \\ \partial s \end{array} \right)_{x_o} + o \| \partial x \|^2.
$$

Upon defining

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

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

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

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

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

3. GENERALIZED CONSTRAINED DERIVATIVES

Again consider problem $P$, 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 $s$ is not necessarily an $m$-vector.

An auxiliary problem, denoted $P_{x_o}$, is now introduced:

$$
\begin{align*}
\text{minimize } & y(x) \\
\text{subject to } & \left( \frac{\partial f}{\partial s} \right)_{x_o} \quad \left( \frac{\partial f}{\partial s} \right)^+_{x_o} \quad f(x) = 0 \\
& x \in \mathbb{R}^n
\end{align*}
$$

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 $x_0 + \delta x$ for problem $P_{x_o}$ is equivalent to

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

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

Since $\left( \frac{\partial f}{\partial s} \right)_{x_o} \quad \left( \frac{\partial f}{\partial s} \right)^+_{x_o} = \mathbb{R} \left( \frac{\partial f}{\partial s} \right)_{x_o}$ and by the above assumption on $\left( \frac{\partial f}{\partial s} \right)_{x_o}$,

$$(3.1) \text{ may be rewritten as follows:}$$

$$(3.2) \quad \left( \frac{\partial f}{\partial s} \right)_{x_o} \quad \delta s + \left( \frac{\partial f}{\partial d} \right)_{x_o} \quad \delta d + \left( \frac{\partial f}{\partial s} \right)_{x_o} \quad \left( \frac{\partial f}{\partial s} \right)^+_{x_o} \quad 0 \quad \|\delta x\| = 0$$
Since all three terms in this equation are in \( \mathbb{R} \left( \frac{\partial f}{\partial s} \right) \), we can solve for \( \delta s \) and obtain

\[
(3.3) \quad \delta s = - \left( \frac{\partial f}{\partial s} \right)_{x_0}^{+} \left( \frac{\partial f}{\partial d} \right)_{x_0} \delta d - 0 ||\delta x||
\]

The generalized constrained derivative of \( y \) at \( x_0 \) with respect to the partition \( \left(-\frac{s}{d}\right) \) is defined by

\[
(3.4) \quad \left( \frac{\delta y}{\delta d} \right)_{x_0} = \nabla_d y(x_0) - \nabla_s y(x_0) \left( \frac{\partial f}{\partial s} \right)_{x_0}^{+} \left( \frac{\partial f}{\partial d} \right)_{x_0}
\]

Similarly to (2.5) we have

\[
(3.5) \quad \delta y = \left( \frac{\delta y}{\delta d} \right)_{x_0} \delta d + 0 ||\delta x||
\]

and the following fact:

A necessary condition for \( x_0 \) to be a local minimum of \( P_{x_0} \) is that

\[
\left( \frac{\delta y}{\delta d} \right)_{x_0} = 0.
\]

It is not necessarily true that the generalized constrained derivative = 0 at a local minimum of problem \( P \), however, as is evidenced by the following counterexample. Consider the problem

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

subject to

\[
f_1(x) = x_1 - x_2^2 - x_3^2 = 0
\]

\[
f_2(x) = x_3^2 = 0
\]

The solution obviously is \((0,0,0)\) and \( \left( \frac{\partial f}{\partial x} \right)_{x_0} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \). Upon setting \( s = x_1, d = \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \) we obtain \( \left( \frac{\delta y}{\delta d} \right)_{x_0} = (0,1) \).

This occurs for the following reason: In expression (2.2), when \( \left( \frac{\partial f}{\partial s} \right)_{x_0} \) is not square, the terms \( 0 ||\delta x|| \) need not be in \( \mathbb{R} \left( \frac{\partial f}{\partial s} \right) \), and one
cannot proceed to solve (2.2) for \( \mathfrak{g} \) by using a generalized inverse. This, it can be verified, is precisely the case in the above counterexample. This difficulty cannot occur, of course, if the constraints in problem \( \overline{P} \) are linear; that is, \( f(x) = Ax - b \). The generalized constrained derivative for a problem with linear constraints is

\[
\left( \frac{\partial y}{\partial d} \right) = \nabla_d y(x_0) - \nabla_s y(x_0) A^+_s A_d
\]

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

For the case where \( \overline{P} \) has linear constraints it is necessary that the generalized constrained derivative be 0 at a local minimum. 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 \( \left( \begin{array}{c} s \\ d \end{array} \right) \) 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

\[(3.6) \quad x_0 \text{ is a local minimum solution of } \overline{P}_{x_0} \iff x_0 \text{ is a local minimum solution of } \overline{P} \]

Thus a sufficient condition for \( x_0 \) to be a local minimum of \( \overline{P}_{x_0} \) is sufficient condition for \( \overline{P} \) as well. Such a condition will now be derived.

Let \( x_0 \) be a feasible point for problem \( \overline{P} \). It is clearly then feasible for \( \overline{P}_{x_0} \). Again assume that \( \left( \frac{\partial f}{\partial s} \right) \) is a maximal full column rank submatrix of \( \left( \frac{\partial f}{\partial x} \right)_{x_0} \). Also, assume now that the functions \( y, f_1, \ldots, f_m \) are three times differentiable. Assume
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 a sufficient condition for \( x_o \) to be a local minimum of \( \overline{P} \) is that (3.7) holds and \( S_{dd}' \) is positive definite. Here \( S_{dd}' \) is the square matrix

\[
S_{dd}' = P_{dd} - P_{ds} \left( \begin{array} {c} \frac{\partial f}{\partial s} \bigg|_{x_o} \\ \frac{\partial f}{\partial d} \bigg|_{x_o} \end{array} \right) - \left[ P_{ds} \left( \begin{array} {c} \frac{\partial f}{\partial s} \bigg|_{x_o} \\ \frac{\partial f}{\partial d} \bigg|_{x_o} \end{array} \right)^* \right]^* + \left( \begin{array} {c} \frac{\partial f}{\partial s} \bigg|_{x_o} \\ \frac{\partial f}{\partial d} \bigg|_{x_o} \end{array} \right)^* P_{ss} \left( \begin{array} {c} \frac{\partial f}{\partial s} \bigg|_{x_o} \\ \frac{\partial f}{\partial d} \bigg|_{x_o} \end{array} \right)
\]

where \( P = H - \sum_{i=1}^{m} \bar{\lambda}_i H_i \) is appropriately partitioned.

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 \( \overline{P} \) 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 \( (n-m \times n-m) \), 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, \( \overline{P}_{x_o} \), 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 $p < m$. Under stationarity of the Lagrangian of $\mathcal{F}$ 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


