1971

A Simplex Algorithm - Gradient Projection Method for Nonlinear Programming

L. Duane Pyle

Report Number:

71-055

CSD TR 55

"A SIMPLEX ALGORITHM - GRADIENT PROJECTION METHOD FOR NONLINEAR PROGRAMMING".

by

L. Duane Pyle

10 June 1971

ABSTRACT

Witzgall [7], commenting on the gradient projection methods of R. Frisch and J. B. Rosen, states: "More or less all algorithms for solving the linear programming problem are known to be modifications of an algorithm for matrix inversion. Thus the simplex method corresponds to the Gauss-Jordan method. The methods of Frisch and Rosen are based on an interesting method for inverting symmetric matrices. However, this method is not a happy one, considered from the numerical point of view, and this seems to account for the relative instability of the projection methods". This paper presents an implementation of the gradient projection method which uses a variation of the simplex algorithm.

The underlying (well-known) geometric idea is that the simplex algorithm for linear programming [1] provides a method for obtaining vectors along the "edges" [4] of the feasible region $\Lambda=\{x \mid Ax=b, x\geq\theta\}$ which lie in certain null spaces. This property is discussed in

detail in section §1., Geometric Analysis of the Simplex Method of Linear Programming.

In section §2., Projection on Faces of Λ of Higher Dimension, the geometric analysis of §1. is extended to obtain the orthogonal projection matrix P such that

$$\mathcal{R}(P) = N(A) \bigcap_{i=s+1}^{n} M(u^{(i)})^{\perp}$$

where $\mathcal{R}(P)$ is the range of P; N(A) is the null space of A; and $M(u^{(i)})^{\perp} = \{x | x_i = 0\}$.

The gradient projection method [6], [2] requires computations involving (1) an orthogonal projection matrix whose range is a certain null space; and (2) a related generalized inverse [3]. In section §3., Simplex Algorithm Implementation of the Gradient Projection Method, the developments given in §2. are combined with the simplex algorithm to provide the computational results required by the gradient projection method. Motivation for this approach may be found in [5].

In the approach given here, a representation of the projection matrix

$$P = (I - N_r N_r^+)$$

is generated using the simplex algorithm, whereas Rosen gives a method for obtaining $N_r N_r^+$ based on an algorithm involving $(N_r^T N_r)^{-1}$. (N_r is a matrix whose columns are normals to the "active" constraints.) If the dimension of $R(I - N_r N_r^+)$ is small compared to the dimension of $R(N_r N_r^+)$, as is the case when the current vector iterate lies on a face of Λ of low dimension, one would expect significant computational improvements. This expectation is further enhanced by the use of a variation of the product form of the inverse in computing the vectors which constitute the representation of the matrix P, and by the use of "simplex multipliers" and "relative cost factors" in the standard fashion of simplex algorithm technology.

§1. Geometric Analysis of the Simplex Algorithm of Linear Programming.

The notation used here is similar to that in [1], Chapters 5 and 8.

Consider the linear programming problem

Minimize (x,c) where

$$(1.01)\qquad Ax \equiv [P_1 P_2 \ldots P_n]\begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = b \equiv P_o$$

and $x \geq \theta$, where θ is a column vector of zeros. Reformulate this problem as

$$\text{Minimize } \left(\begin{bmatrix} z \\ x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}\right) = z \text{ where}$$

(1.02)

$$\left[\hat{P}_z \hat{P}_1 \ldots \hat{P}_n\right]\begin{bmatrix} z \\ x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \equiv \begin{bmatrix} -1 & c_1 & \ldots & c_n \\ \theta & P_1 & \ldots & P_n \end{bmatrix}\begin{bmatrix} z \\ x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ P_o \end{bmatrix} \equiv \hat{P}_o$$

$$x_i \geq 0 \quad (i = 1, \ldots, n),$$

z unrestricted.

Now, suppose x_1, x_2, ..., x_m and z are basic, feasible variables (we assume that A is m by n and that rank A = m; we further assume that index sets have convenient labels, such as (1,2,...,m), rather than using the correct, but ponderous $(j_1, j_2, \ldots, j_m)$).

(1.03) With $\hat{B} = [\hat{P}_z \hat{P}_1 \ldots \hat{P}_m]$, where $\hat{P}_z = \begin{bmatrix} -1 \\ \theta \end{bmatrix}$, after a sequence of pivotal reductions one obtains the canonical form

(1.04) $$\hat{B}^{-1}[\hat{P}_z \hat{P}_1 \ldots \hat{P}_m \hat{P}_{m+1} \ldots \hat{P}_n \hat{P}_0] = \left[\begin{matrix} -1 & \theta \\ \theta & I \end{matrix} \;\middle|\; \hat{\bar{P}}_{m+1} \ldots \hat{\bar{P}}_n \hat{\bar{P}}_0\right]$$

which we partition as

(1.05) $$\begin{bmatrix} -1 & \theta^T & \bar{c}_{m+1} \ldots \bar{c}_n & -\bar{z}_0 \\ \theta & I & \bar{P}_{m+1} \ldots \bar{P}_n & \bar{P}_0 \end{bmatrix}.$$

(1.06) From $\hat{\bar{P}}_j = \hat{B}^{-1} \hat{P}_j$ for $(j = 0, m+1, \ldots, n)$

(1.07) one obtains $\hat{P}_j = \hat{B}\hat{\bar{P}}_j$, thus

(1.08) $P_j = [P_1 \ldots P_m] [\bar{P}_j] \quad (i = 0, m+1, \ldots, n)$

(1.09) $c_j = [c_1 \ldots c_m] [\bar{P}_j] + \bar{c}_j \quad (j = m+1, \ldots, n)$

and

(1.10) $0 = [c_1 \ldots c_m] [\bar{P}_0] - \bar{z}_0$.

(1.08) states that

(1.11) $$[P_1 P_2 \ldots P_m P_{m+1} \ldots P_j \ldots P_n] \begin{bmatrix} \bar{P}_j \\ \theta_1 \\ -1 \\ \theta_2 \end{bmatrix} = \theta$$

where θ_1 and θ_2 are vectors of zeros; the scalar -1 is the "jth" component of the vector

(1.12) $$\begin{bmatrix} \bar{P}_j \\ \theta_1 \\ -1 \\ \theta_2 \end{bmatrix},$$ which thus lies in the null space of the matrix A.

If the basic solution corresponding to the canonical form (1.04) is non-degenerate, that is, if $\bar{P}_0 > \theta$, then the point

(1.13) $$\begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_m \\ x_{m+1} \\ \cdot \\ \cdot \\ \cdot \\ x_j \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} \bar{P}_0 \\ \text{------} \\ \theta \end{bmatrix} + \alpha \begin{bmatrix} -\bar{P}_j \\ \text{------} \\ \theta_1 \\ 1 \\ \theta_2 \end{bmatrix}$$

is feasible for problem (1.01) provided $\alpha > 0$, α sufficiently small.

The vector $\begin{bmatrix} -\bar{P}_j \\ \theta_1 \\ 1 \\ \theta_2 \end{bmatrix}$ is a "vector along an edge of $\Lambda = \{x | Ax = b, x \geq \theta\}$" [5]. The question is whether or not a path in the direction of this vector produces a decrease in (x,c). This will be the case for $\alpha > 0$, provided that

$$(1.14)\quad \left(\begin{bmatrix} -\bar{P}_j \\ \theta_1 \\ 1 \\ \theta_2 \end{bmatrix}, c \right) < 0.$$

Now, from relation (1.09)

we have that $\left(\begin{bmatrix} -\bar{P}_j \\ \theta_1 \\ 1 \\ \theta_2 \end{bmatrix}, c \right) = \bar{c}_j$, thus if $\bar{c}_j < 0$ a decrease in (x,c) results for all $\alpha > 0$.

§2. Projection on Faces of Λ of Higher Dimension.

From the analysis given in §1. it is apparent that a non-degenerate, basic feasible canonical form provides, essentially, the directions of the projections of the gradient of the function $f(x) = (x,c)$, denoted

$$\nabla f \equiv \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix},$$

when ∇f is projected orthogonally upon the various 1-dimensional faces of $\Lambda = \{x | Ax = b, x \geq \theta\}$ which intersect in the vertex of Λ associated with that canonical form. In order to obtain the projections of ∇f upon higher dimensional faces of Λ, consider the following:

Suppose the canonical form

$$\left[I \;\vdots\; \overline{P}_{m+1} \cdots \overline{P}_n \right] \quad \text{has been obtained.} \tag{2.01}$$

(Note the omission of the "z-row" and $\overline{P}_0$ column.)

Define

$$\tilde{\eta}^{(m+1)} = \begin{bmatrix} -\overline{P}_{m+1} \\ 1 \\ \theta \end{bmatrix} \text{ and } \eta^{(m+1)} = \frac{1}{\|\tilde{\eta}^{(m+1)}\|} \tilde{\eta}^{(m+1)} . \tag{2.02}$$

Then from §1 it follows that

$$\eta^{(m+1)} \in \mathcal{N}(A) \cap \mathcal{M}(u^{(m+2)})^{\perp} \cap \ldots \cap \mathcal{M}(u^{(n)})^{\perp} ,$$

where

$$\mathcal{N}(A) \equiv \text{the null space of } A = \{z | Az = \theta\}.$$

Now, adjoin $\tilde{\eta}^{(m+1)T}$ to (2.01) and complete the reduction to "canonical form"; i.e. using elementary row operations in the obvious way, reduce the form

(2.03) $$\left[\begin{array}{c|ccc} I & \overline{P}_{m+1} & \cdots & \overline{P}_n \\ \hline -\overline{P}^T_{m+1} & 1 & & \theta^T \end{array}\right]$$ to the form

(2.04) $$\left[\begin{array}{c|ccc} I & \overline{P}_{m+2} & \cdots & \overline{P}_n \end{array}\right]$$ where I is (m+1) by (m+1).

(2.05) Define $$\tilde{\eta}^{(m+2)} = \begin{bmatrix} -\overline{\overline{P}}_{m+2} \\ 1 \\ \theta \end{bmatrix} \quad \text{and} \quad \eta^{(m+2)} = \frac{1}{\|\tilde{\eta}^{(m+2)}\|} \tilde{\eta}^{(m+2)} .$$

As before, it follows that

(2.06) $$\eta^{(m+2)} \in \mathcal{N}(A) \cap \mathcal{M}(\eta^{(m+1)})^{\perp} \cap \mathcal{M}(u^{(m+3)})^{\perp} \cap \ldots \cap \mathcal{M}(u^{(n)})^{\perp} .$$

Continuing in this manner, obtain vectors

(2.07) $$\eta^{(m+1)}, \eta^{(m+2)}, \ldots, \eta^{(s)} \text{ where } (\eta^{(i)}, \eta^{(j)}) = \delta_{ij}$$
$$(s = m+1, m+2, \ldots, n)$$

and

(2.08) $$\eta^{(i)} \in \mathcal{N}(A) \cap \mathcal{M}(u^{(s+1)})^{\perp} \cap \ldots \cap \mathcal{M}(u^{(n)})^{\perp}$$

for $(i = (m+1), (m+2), \ldots, s)$ where $s \leq n$.

Thus, as is well-known,

(2.09) $$P = \sum_{i=m+1}^{s} \eta^{(i)} \eta^{(i)T}$$ is an orthogonal projection and we

we have

Theorem 2.1:

(2.10) $\mathcal{R}(P) = \mathcal{N}(A) \bigcap_{i=s+1}^{n} \mathcal{M}(u^{(i)})^{\perp}$, where $\mathcal{R}(P) \equiv$ Range of $P = \{y \mid \exists\, z \text{ with } y = Pz\}$.

Proof:

Relation (2.10) follows from the assumption that the vertex associated with the canonical form (2.01) is non-degenerate (see definitions of degeneracy and non-degeneracy given in [4]), for then the vectors $u^{(m+1)},\ldots,u^{(n)}$, together with a basis for the column space of A^T, $\{e^{(i)}\}$ $(i = q+1,\ldots,n)$ where $q = n-m$, form a set of n, linearly independent vectors. Under the assumption that A has full row rank, one could just as well take the $\{e^{(i)}\}$ to consist of the columns of A^T. The origin of the notation $\{e^{(i)}\}$ resides in developments given in [4] and [5]. Thus, by construction

$$\begin{aligned}\eta^{(m+1)} &\in \mathcal{M}\,(e^{(q+1)},\ldots,e^{(n)},u^{(m+2)},\ldots,u^{(n)})^{\perp} \\ &= \mathcal{M}(e^{(q+1)},\ldots,e^{(n)})^{\perp} \bigcap_{i=m+2}^{n} \mathcal{M}(u^{(i)})^{\perp} \qquad (2.11) \\ &= \mathcal{N}(A) \bigcap_{i=m+2}^{n} \mathcal{M}(u^{(i)})^{\perp}\end{aligned}$$

(2.12) where dimension of $\mathcal{M}(e^{(q+1)},\ldots,e^{(n)},u^{(m+2)},\ldots,e^{(n)})^{\perp} = 1$. But dimension of $\mathcal{R}(\eta^{(m+1)}\eta^{(m+1)T}) = 1$, together with

$$\mathcal{R}(\eta^{(m+1)}\eta^{(m+1)T}) \subset \mathcal{N}(A) \bigcap_{i=m+2}^{n} \mathcal{M}(u^{(i)})^{\perp}$$

implies

$$\mathcal{R}(\eta^{(m+1)}\eta^{(m+1)T}) = \mathcal{N}(A) \bigcap_{i=m+2}^{n} \mathcal{M}(u^{(i)})^{\perp} . \tag{2.13}$$

Similarly,

$$\eta^{(m+2)} \in \mathcal{N}(A) \bigcap_{i=m+3}^{n} \mathcal{M}(u^{(i)})^{\perp}$$

where dimension of $\mathcal{N}(A) \bigcap_{i=m+3}^{n} \mathcal{M}(u^{(i)})^{\perp} = 2$,

and so on. Thus

$$\mathcal{R}(P) \subset \mathcal{N}(A) \bigcap_{i=s+1}^{n} \mathcal{M}(u^{(i)})^{\perp} \tag{2.14}$$

where dim $\mathcal{R}(P) = (s+1) - (m+1) = s-m$ and

dim $\mathcal{N}(A) \bigcap_{i=s+1}^{n} \mathcal{M}(u^{(i)})^{\perp} = n-[(n-q) + (n-s)] = s-m$,

consequently

$$\mathcal{R}(P) = \mathcal{N}(A) \bigcap_{i=s+1}^{n} \mathcal{M}(u^{(i)})^{\perp} . \tag{2.15}$$

Remark:

When the set of unit vectors is exhausted, i.e., when s=n, then $\mathcal{R}(P) = \mathcal{N}(A)$ and

$$P = I-A^{+}A,$$

where A^{+} is the generalized inverse of A [3].

§3. Simplex Algorithm Implementation of the Gradient Projection Method.

In §1. the Simplex Algorithm of linear programming [1] was shown to provide a method for obtaining "vectors along the edges" [5] of the feasible region, Λ, which lie in certain null spaces. This idea was extended in §2 to provide a means of projecting on faces of higher dimension. Use of the gradient projection method [6] requires computations involving

(1) an orthogonal projection matrix whose range is a certain null space; and (2) a related generalized inverse [6],[2]. In this section we continue the discussion given in §2 to show how the simplex algorithm may be used to provide the computational results involving (1) and (2). Motivation for this approach may be found in [4].

Consider the transformation which must be applied to the matrix discussed in §2:

$$\left[\begin{array}{c|ccc} I & \bar{P}_{m+1} & \cdots & \bar{P}_n \\ \hline \bar{P}^T_{m+1} & 1 & & \theta^T \end{array}\right] \quad . \tag{3.01}$$

Let

$$\bar{P}_{m+1} = \begin{bmatrix} \alpha_{1,m+1} \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{m,m+1} \end{bmatrix} \qquad \text{and} \tag{3.02}$$

$$\sigma_{m+1} = 1 + \sum_{i=1}^{m} \alpha_{i,m+1}^2 \equiv 1 + (\bar{P}_{m+1}, \bar{P}_{m+1}). \tag{3.03}$$

If $\sigma_{m+1} = 1$ then no action is required. Otherwise $\sigma_{m+1} > 1$ and the transformation of (3.01) to canonical form

$$\left[I \;\; \bar{\bar{P}}_{m+2} \; \ldots \; \bar{\bar{P}}_n \right] \tag{3.04}$$

is well-defined. This is achieved by pre-multiplying the columns of (3.01) by

$$M_1 \equiv N_1^{(1)} N_2^{(1)} N_3^{(1)} \equiv \begin{bmatrix} I & -\bar{P}_{m+1} \\ \theta^T & 1 \end{bmatrix} \begin{bmatrix} I & \theta \\ \theta^T & \frac{1}{\sigma_{m+1}} \end{bmatrix} \begin{bmatrix} I & \theta \\ \bar{P}_{m+1}^T & 1 \end{bmatrix} \tag{3.05}$$

If $B^{-1} = [P_1 \ldots P_m]^{-1}$ is retained in product form, then the required composite product form transformation is

$$T_1 = M_1 \begin{bmatrix} B^{-1} & \theta \\ \theta^T & 1 \end{bmatrix} \tag{3.06}$$

It is easily verified that

$$\begin{aligned} T_1 \begin{bmatrix} P_j \\ -\alpha_{j,m+1} \end{bmatrix} &= u^{(j)} \qquad (j = 1,\ldots,m) \\ T_1 \begin{bmatrix} P_{m+1} \\ 1 \end{bmatrix} &= u^{(m+1)} \\ T_1 \begin{bmatrix} P_j \\ 0 \end{bmatrix} &= \bar{\bar{P}}_j \qquad (j = m+2,\ldots,n). \end{aligned} \tag{3.07}$$

(Note that $u^{(j)}, u^{(m+1)}$ and $\bar{\bar{P}}_j$ are vectors with m+1 elements).

To compute $M_1 y$ for some vector y, one need have stored only the m+1 values $\{\alpha_{i,m+1}\}$ and σ_{m+1}. When

(3.08) $$y = \begin{bmatrix} B^{-1} & \theta \\ \theta^T & 1 \end{bmatrix} \begin{bmatrix} P_j \\ 0 \end{bmatrix} = \begin{bmatrix} \bar{P}_j \\ 0 \end{bmatrix} \quad (j = m+2,\ldots,n)$$

then

(3.09) $$\bar{\bar{P}}_j \equiv M_1 y = \begin{bmatrix} \bar{P}_j \\ 0 \end{bmatrix} + \frac{(\bar{P}_{m+1}, \bar{P}_j)}{1+(\bar{P}_{m+1}, \bar{P}_{m+1})} \begin{bmatrix} -\bar{P}_{m+1} \\ 1 \end{bmatrix},$$

which follows from the representation of M_1 given below:

(3.10) $$M_1 = I - \frac{1}{\sigma_{m+1}} \begin{bmatrix} \bar{P}_{m+1}\bar{P}^T_{m+1} & \bar{P}_{m+1} \\ -\bar{P}^T_{m+1} & \bar{P}^T_{m+1}\bar{P}_{m+1} \end{bmatrix}.$$

The continuation of the process described above, yielding the vectors $\bar{\bar{\bar{P}}}_{m+3},\ldots,\bar{\bar{\bar{P}}}_n$, is performed using the transformation

(3.11) $$T_2 = M_2 \begin{bmatrix} T_1 & \theta \\ \theta^T & 1 \end{bmatrix}$$

where

(3.12) $$M_2 \equiv N_1^{(2)} N_2^{(2)} N_3^{(2)} = \begin{bmatrix} I & -\bar{\bar{P}}_{m+2} \\ \theta^T & 1 \end{bmatrix} \begin{bmatrix} I & \theta \\ \theta^T & \frac{1}{\sigma_{m+2}} \end{bmatrix} \begin{bmatrix} I & \theta \\ \bar{\bar{P}}^T_{m+2} & 1 \end{bmatrix}$$

(3.13) and $\sigma_{m+2} = 1 + (\overset{=}{P}_{m+2}, \overset{=}{P}_{m+2})$. In general, then, the transformation which yields the successive canonical forms is

(3.14) $$T_{k+1} = M_{k+1} \begin{bmatrix} T_k & \theta \\ \theta^T & 1 \end{bmatrix} \quad \text{for } (k = 0,1,2,\ldots,(n-m-1))$$

where

(3.15) $$M_{k+1} \equiv N_1^{(k+1)} N_2^{(k+1)} N_3^{(k+1)} = \begin{bmatrix} I & -\overset{k+1}{P}_{m+k+1} \\ \theta^T & 1 \end{bmatrix} \begin{bmatrix} I & \theta \\ \theta^T & \frac{1}{\sigma_{m+k+1}} \end{bmatrix} \begin{bmatrix} I & \theta \\ \overset{k+1}{P}{}^T_{m+k+1} & 1 \end{bmatrix}$$

$$\sigma_{m+k+1} = 1 + (\overset{k+1}{P}_{m+k+1}, \overset{k+1}{P}_{m+k+1}),$$

$T_0 = B^{-1}$ and $\overset{k+1}{P}_{m+k+1}$ is the "(m+k+1)st" column of the "(k+1)st" canonical form. Note that T_{k+1} is an (m+k+1) by (m+k+1) matrix. The relations analogous to (3.07) are as follows:

(3.16)

$$T_{k+1} \begin{bmatrix} P_j \\ -\alpha_{j,m+1} \\ -\alpha_{j,m+2} \\ \cdot \\ \cdot \\ \cdot \\ -\alpha_{j,m+k+1} \end{bmatrix} = u^{(j)} \quad (j = 1,2,\ldots,m)$$

$$T_{k+1} \begin{bmatrix} P_{m+1} \\ 1 \\ -\alpha_{m+1,m+2} \\ -\alpha_{m+1,m+3} \\ \cdot \\ \cdot \\ \cdot \\ -\alpha_{m+1,m+k+1} \end{bmatrix} = u^{(m+1)}$$

(3.16)

$$T_{k+1}\begin{bmatrix} P_{m+2} \\ 0 \\ 1 \\ -\alpha_{m+2,m+3} \\ -\alpha_{m+2,m+4} \\ \cdot \\ \cdot \\ \cdot \\ -\alpha_{m+2,m+k+1} \end{bmatrix} = u^{(m+2)}$$

$$\vdots$$

$$T_{k+1}\begin{bmatrix} P_{m+k+1} \\ \theta_k \\ 1 \end{bmatrix} = u^{(m+k+1)}$$

where the unit vectors $u^{(j)}$ are the columns of the (m+k+1) by (m+k+1) identity matrix, θ_k is a vector whose k elements are zeros, and

$$T_{k+1}\begin{bmatrix} P_j \\ \theta_{k+1} \end{bmatrix} = \frac{k+2}{P_j} \quad \text{for } (j = m+k+2,\dots,n).$$

Consider the nonlinear programming problem

(3.17) Maximize g(x) where $F(x) = \begin{bmatrix} f_1(x) \\ \cdot \\ \cdot \\ \cdot \\ f_t(x) \end{bmatrix} \geq \theta$.

If g(x) and the $f_i(x)$ $(i = 1,\dots,t)$ are differentiable and concave (see [6],[2]) then if $F(\hat{x}) \geq \theta$, $\hat{x}$ is a (global) constrained maximum of $g(x)$ if and only if there exists an n by r matrix of rank $r \leq p$ whose r columns provide a set of linearly independent gradients to the p constraints active at $\hat{x}$, such that

(3.18) $$(I - N_r N_r^+) \nabla g(\hat{x}) = \theta$$

(3.19) $$-N_r^+ \nabla g(\hat{x}) \geq \theta .$$

A constraint $f_i(x)$ is active at $\hat{x}$ provided $f_i(\hat{x}) = 0$; otherwise $f_i(\hat{x}) > 0$ and $f_i(x)$ is said to be inactive at $\hat{x}$.

Rosen's conditions (3.18) and (3.19), qualified as above, are equivalent to the Kuhn-Tucker conditions [2]. The proof given in [2] requires only minor modifications in order to support the following:

(3.20) Let $N_p = [\nabla f_{j_1}(\hat{x}) \ldots \nabla f_{j_p}(\hat{x})]$

be that matrix whose columns consist of the gradients to all p active constraints. Then in (3.18) and (3.19) r may be taken to be equal to the rank of N_p. Thus the matrix N_r is composed as follows:

(3.21) $$N_r = [\nabla f_{j_1}(\hat{x}) \ldots \nabla f_{j_r}(\hat{x})]$$

where $f_{j_1}(\hat{x}), \ldots, f_{j_r}(\hat{x})$ form a subset of the active constraints, active at $\hat{x}$, and $r = \text{rank } N_p$.

Now restrict consideration to the linearly constrained, nonlinear programming problem

(3.22) $$\text{Maximize } h(x) \text{ where } Ax = b, \quad x \geq \theta ,$$

and h(x) is concave, differentiable. In order to simplify the expositions we will continue to assume that A is m by n of rank m and, further, that all points in the set $\Lambda = \{x | Ax = b, \ x \geq \theta\}$ are nondegenerate in the sense defined in [5]. A second paper will take up techniques for handling deficient rank and degeneracy, together with a related topic: the use of an interior gradient projection method [5] when some of the constraints $f_i(x)$, other than the non-negativity constraints $x \geq \theta$, are nonlinear.

If problem (3.22) is replaced by the equivalent problem

(3.23) $$\text{Maximize} \quad \left(\begin{bmatrix} z \\ x_1 \\ \vdots \\ x_n \end{bmatrix} , \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right) = g(z, x_1, \ldots, x_m) = z$$

where

$$\begin{aligned} h(x) - z &= 0 \\ Ax &= b \\ x &\geq \theta \end{aligned}$$

and each equality constraint is represented by a pair of inequality constraints, then at a non-degenerate vertex $\tilde{x}$ of Λ, which corresponds to a canonical form

(3.24)
$$\left[\begin{array}{c|c|cccc} -1 & \theta^T & \bar{\sigma}_{m+1}(\tilde{x}) & \dots & \bar{\sigma}_n(\tilde{x}) & -\bar{z}_o \\ \hline \theta & I & \bar{p}_{m+1} & \dots & \bar{p}_n & \bar{p}_o \end{array}\right]$$

where

$$\bar{z}_o = h(\tilde{x})$$

and the $\bar{\sigma}_j(\tilde{x})$ are obtained by transforming the vector $\nabla h(\tilde{x})$ in the standard way [1], the matrix N_p is composed as follows:

(3.25)
$$\left[\left[\begin{array}{cc} -1 & \theta^T \\ \hline \theta & I \\ \hline \bar{\sigma}_{m+1} & \bar{p}^T_{m+1} \\ \vdots & \vdots \\ \bar{\sigma}_n & \bar{p}^T_n \end{array}\right], \left[\begin{array}{cc} -1 & \theta^T \\ \hline \theta & I \\ \hline \bar{\sigma}_{m+1} & \bar{p}^T_{m+1} \\ \vdots & \vdots \\ \bar{\sigma}_n & \bar{p}^T_n \end{array}\right], \left[\begin{array}{c} 0 \\ \hline \\ I \\ \\ \end{array}\right]\right].$$

Since $\tilde{x}$ non-degenerate implies rank $N_p = n+1$, it follows that the matrix N_r of (3.18), (3.19) consists of (n+1) columns; precisely (m+1) of the columns of N_r are taken from the first two submatrices of N_p, the remainder being taken from the third submatrix. If column j appears in N_r, column (j+m+1) does not, and conversely, for

$(j = 1,2,\ldots,m+1)$. In any event, N_r is nonsingular and thus

$$I - N_r N_r^+ = I - N_r N_r^{-1} = I - I = 0$$

and (3.18) is trivially satisfied. Put somewhat differently, the vertex $\tilde{x}$ of Λ lies in the 0-dimensional intersection of the active constraints.

Condition (3.19) is realized by solving the system

(3.26) $$N_r \, w = u^{(1)} \equiv \nabla g$$

($u^{(1)}$ here has (n+1) elements)

for its unique solution $w = \begin{bmatrix} w_0 \\ w_1 \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{bmatrix}$, as follows:

Case A: If column 1 of N_p appears in N_r,

$$w_0 = -1,\ w_1 = \ldots = w_m = 0,\ w_{m+1} = \bar{c}_{m+1},\ldots,w_n = \bar{c}_n.$$

Case B: If column m+2 of N_p appears in N_r,

$$w_0 = 1,\ w_1 = \ldots = w_m = 0,\ w_{m+1} = \bar{c}_{m+1},\ldots,w_n = \bar{c}_n.$$

Then we have

Theorem 3.1: The non-degenerate $\tilde{x}$ is a maximizing solution for (3.23) if and only if

(3.27) $\bar{c}_j(\tilde{x}) \leq 0 \quad (j = m+1,\dots,n)$, where the $\bar{c}_j(\tilde{x})$ are defined in (3.24).

<u>Proof</u>: If $\tilde{x}$ is a maximizing solution for (3.24), then (3.18) and (3.19) imply that the solution w of (3.26) is such that $w \leq \theta$. This means that Case B: above is impossible, thus Case A must hold, in which case $\bar{c}_j(\tilde{x}) \leq 0$ for $(j = m+1,\dots,n)$.

If $\bar{c}_j(\tilde{x}) \leq 0$ for $(j = m+1,\dots,n)$, then with w determined as in Case A, conditions (3.18) and (3.19) hold with any N_r (of the type specified following relation (3.25)) having as its first column the first column of N_p in (3.25). Thus $\tilde{x}$ is a maximizing solution for (3.23).

At a vertex of Λ the condition

$$w = N_r^+ \nabla g \leq \theta$$

is thus the well-known condition involving the "relative cost factors" $\bar{c}_j$ [1].

If instead of (3.22) the problem is formulated as

(3.28) Minimize f(x) where Ax = b, $x \geq \theta$, where f(x) is convex, differentiable, then the above conditions become those for maximizing -f(x) and we take h(x) = -f(x), in which case, assuming the $\bar{c}_j(\tilde{x})$ values are obtained at a vertex $\tilde{x}$ by suitably transforming the vector $\nabla f(\tilde{x})$, condition (3.19) becomes

$-\bar{c}_j(\tilde{x}) \leq 0$; that is,

(3.29) $$\bar{c}_j(\tilde{x}) \geq 0 \quad \text{for } (j = m+1,\ldots,n).$$

Since the developments in sections §1 and §2 follow Dantzig's practice of formulating problems in terms of minimization, whereas Rosen formulates problems in terms of maximization, it is impossible to continue notational agreement. In the remainder of the paper we will follow Dantzig's practice. Thus suppose Rosen's gradient projection method [6] is initiated at a vertex $\tilde{x}$ of Λ. Such a vertex could be obtained, for example, by executing "Phase I" of the simplex method [1]. Then, unless $\tilde{x}$ is an optimal solution for the problem:

(3.30) Minimize $f(x)$ where $Ax = b$, $x \geq \theta$, $f(x)$ convex, differentiable; it must be that $\bar{c}_j(\tilde{x}) < 0$ for at least one value of j. Defining s, which for convenience we label $m+1$, by letting

$$\bar{c}_s(\tilde{x}) = \underset{\bar{c}_j(\tilde{x})<0}{\text{minimum}} \ \bar{c}_j(\tilde{x}),$$

compute the direction of a vector $\eta^{(m+1)}$ along the associated edge of Λ, determining a path in that direction. Let $\tilde{\tilde{x}}$ be that feasible point located at the maximal distance from $\tilde{x}$ along this (linear) path. Then determine $\hat{x}$ such that $f(\hat{x}) \leq f(x)$ for all $x = \mu \tilde{x} + (1-\mu) \tilde{\tilde{x}}$ where $0 \leq \mu < 1$. If $\hat{x} = \tilde{\tilde{x}}$, the process then employed is a simplex iteration. If $\hat{x} \neq \tilde{\tilde{x}}$, then at $\hat{x}$

(3.31) $$\eta^{(m+1)}\eta^{(m+1)T} \nabla f(\hat{x}) = \theta,$$

since otherwise $\hat{\hat{x}}$ may be determined on the path with $f(\hat{\hat{x}}) < f(\hat{x})$. Thus $\nabla f(\hat{x})$ is orthogonal to the vector $\tilde{\tilde{x}} - \tilde{x}$. Note that, as usual, a 1-dimensional minimization problem must be solved; we assume that this presents no difficulties. The point $\hat{x}$ lies in the 1-dimensional intersection of the active constraints, $Ax = b$, $x_{m+2} = \cdots = x_n = 0$, precisely at the point where it intersects the hyperplane $(\eta^{(m+1)}, x) = (\eta^{(m+1)}, \hat{x})$.

Let

(3.32) $$\bar{N}_r = \left[\begin{bmatrix} -1 & \theta^T \\ \hline \theta & I \\ \hline \bar{c}_{m+2}(\hat{x}) & \bar{P}^T_{m+2} \\ \vdots & \vdots \\ \bar{c}_n(\hat{x}) & \bar{P}^T_n \end{bmatrix}, \begin{bmatrix} 0 \\ \\ I \end{bmatrix}\right]$$

where the values $\bar{c}_j(\hat{x})$ for $(j = m+2,\ldots,n)$ are obtained by transforming the vector $\nabla f(\hat{x})$. (A method for determining the successive vectors of $\bar{c}_j$ values, using a variation of the simplex algorithm with multipliers, is given at the end of this section.) Then we have

<u>Theorem 3.2</u>: If $\eta^{(m+1)}\eta^{(m+1)} \nabla f(\tilde{x}) = \theta$ then $\hat{x}$ is a minimizing solution of (3.28) if and only if $\bar{c}_j(\hat{x}) \geq 0$ for $(j = m+2,\ldots,n)$.

<u>Proof</u>: Suppose $\bar{c}_j(\hat{x}) < 0$ for some value j_o in the index set $(j = m+2,\ldots,n)$. Then $\hat{x}$ is not a minimizing solution of (3.28) since $f(x)$ can be decreased by treating $\hat{x}$ as a vertex of the convex set

$$\Lambda^{(1)} = \{x \mid Ax = b;\ (x,\eta^{(m+1)}) = (\hat{x},\eta^{(m+1)});\ x \geq \theta\}$$

by proceeding along an edge of $\Lambda^{(1)}$ determined by j_o. Thus $\hat{x}$ a minimizing solution of (3.28) implies $\bar{c}_j(\hat{x}) \geq 0$ for $(j = m+2,\ldots,n)$.

Now suppose that $c_j(\hat{x}) \geq 0$ for $(j = m+2,\ldots,n)$, and that $\hat{x}$ is not a minimizing solution for (3.28). Let

(3.33) $$N_r = \begin{bmatrix} 1 & \theta^T & \theta^T \\ \nabla f(\hat{x}) & A^T & u^{(m+2)} \ldots u^{(n)} \end{bmatrix}$$

and $A \equiv \begin{bmatrix} a^{(1)} \\ \vdots \\ a^{(m)} \end{bmatrix}$ where $a^{(i)}$ is row i of A.

Since $\eta^{(m+1)}\eta^{(m+1)T} \nabla f(\hat{x}) = \theta$, $\nabla f(\hat{x}) \in \mathcal{M}(\eta^{(m+1)})^{\perp}$.

$$\text{But } \mathcal{M}(\eta^{(m+1)})^{\perp} = \{(\mathcal{N}(A) \bigcap_{i=m+2}^{n} \mathcal{M}(u^{(i)})^{\perp}\}^{\perp}$$

$$= \{(\bigcap_{i=1}^{m} \mathcal{M}(a^{(i)T})^{\perp}) \cap (\bigcap_{i=m+2}^{n} \mathcal{M}(u^{(i)}))\}^{\perp}$$

$$= \mathcal{M}(a^{(1)T},\ldots,a^{(m)T},u^{(m+2)},\ldots,u^{(n)}),$$

thus there exists a (unique) solution v to the system

$$(3.34) \qquad [A^T\ u^{(m+2)} \ldots u^{(n)}]\ v = \nabla f(\hat{x})$$

where $v = \begin{bmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_{n-1} \end{bmatrix}$, and we have $\bar{N}_r \begin{bmatrix} 1 \\ -v \end{bmatrix} = \nabla g = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$

where $\bar{N}_r$ is defined in (3.33). Sure $\bar{N}_r$ has full column rank,

$$(3.35) \qquad \bar{N}_r^{+} = (\bar{N}_r^T \bar{N}_r)^{-1} \bar{N}_r^T$$

hence

$$(3.36) \qquad \bar{N}_r^{+} \nabla g = \begin{bmatrix} 1 \\ -v \end{bmatrix}$$

thus

$$\bar{N}_r \bar{N}_r^{+} \nabla g = \nabla g ,$$

that is,

(3.37) $$(I - \tilde{N}_r \tilde{N}_r^{+}) \nabla g = \theta.$$

Therefore, if $\hat{x}$ is not a minimizing solution for (3.28), v must have at least one element v_{k_0}, k_o in the index set $(k = 1,\dots,n-1)$, such that $-v_{k_0} < 0$, i.e. such that $v_{k_0} > 0$.

Now let

(3.38) $$\tilde{N}_p = \left[\begin{bmatrix} -1 & \theta^T \\ -\nabla f(\hat{x}) & A^T \end{bmatrix}, \; - \begin{bmatrix} -1 & \theta^T \\ -\nabla f(\hat{x}) & A^T \end{bmatrix}, \; \begin{bmatrix} \theta^T \\ \hline u^{(m+2)} \dots u^{(n)} \end{bmatrix} \right].$$

It follows that the conclusion of the preceding paragraph may be drawn for any $\tilde{N}_r$ composed of n linearly independent columns taken from $\tilde{N}_p$. As in the development following relation (3.25), precisely (m+1) of the columns of such an $\tilde{N}_r$ are taken from the first two submatrices of N_p, the remaining columns being taken from the third. If column j appears in $\tilde{N}_r$, column (j+m+1) does not, and conversely, for $(j = 1,2,\dots,m+1)$. $\tilde{N}_r$ here has rank (n-1).

By previous developments, $\tilde{\sigma}_j(\hat{x}) \geq 0$ for $(j = m+2,\dots,n)$ implies that $\hat{x}$ is a minimizing solution for the problem

(3.39) Minimize f(x) where
$$Ax = b$$
$$(\eta^{(m+1)}, x) = (\eta^{(m+1)}, \hat{x})$$
and
$$x \geq \theta.$$

Thus, defining

$$(3.40)\quad \tilde{N}_p = \left[\begin{bmatrix} -1 & \theta^T & 0 \\ -\nabla f(\hat{x}) & A^T & \eta^{(m+1)} \end{bmatrix}, - \begin{bmatrix} -1 & \theta^T & 0 \\ -\nabla f(\hat{x}) & A^T & \eta^{(m+1)} \end{bmatrix} \begin{bmatrix} & \theta^T & \\ \hline u^{(m+2)} & \dots & u^{(n)} \end{bmatrix} \right]$$

(3.18) and (3.19) imply that there exists a matrix $\tilde{N}_r$ of rank (n+1), with columns chosen as in the development following relation (3.25) (except that (m+2) columns are taken from the first two submatrices; as before, $\tilde{N}_r$ is nonsingular, thus (3.19) is trivially satisfied), and such that the unique solution z of

$$(3.41)\quad \tilde{N}_r\, z = \nabla g$$

satisfies

$$z \geq \theta, \text{ where } z = \begin{bmatrix} z_o \\ z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{bmatrix}.$$

Since $z_o = \pm 1$, it follows that $z_o = 1$, thus that column (m+3), $[\nabla f(\hat{x})]$, of $\tilde{N}_p$ appears in $\tilde{N}_r$. Designating $\tilde{N}_r$ as follows:

$$\tilde{N}_r = \begin{bmatrix} 1 & \theta^T & \begin{matrix} & \theta^T & \end{matrix} \\ \nabla f(\hat{x}) & D & \begin{matrix} u^{(m+2)} & \dots & u^{(n)} \end{matrix} \end{bmatrix}$$

where $D = [d^{(1)} \dots d^{(m+1)}]$ is composed of (m+1) linearly independent columns taken from $[A^T\ \eta^{(m+1)}$ and $-[A^T\ \eta^{(m+1)}]$, thus

$$z_o\ \nabla f(\hat{x}) + [D,\ u^{(m+2)} \dots u^{(n)}] \begin{bmatrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{bmatrix} = \theta,$$

where $z_i \geq 0$ for $(i = 1,2,\ldots,n)$ and $z_o = 1$,

or

$$[D,\ u^{(m+2)}\ldots\ u^{(n)}]\begin{bmatrix}-z_1\\ \cdot\\ \cdot\\ \cdot\\ -z_n\end{bmatrix} = \nabla f(\hat{x})$$

where $-z_i \leq 0$. Now in reference to (3.33), take

$$N_r' = \left[\begin{array}{cc|ccc} 1 & \theta^T & & \theta^T & \\ \hline \nabla f(\hat{x}) & D' & u^{(m+2)} & \ldots & u^{(n)} \end{array}\right]$$

where D' consists of the first (m+1) columns of D. As pointed out following relation (3.37), $(I-(N_r')(N_r')^{+})\nabla g=\theta$ and if $\hat{x}$ is not a minimizing solution for (3.28), the associated v', which solves the system analogous to (3.34), must have at least one element $v'_{k_o} > 0$. The uniqueness of the representation of $\nabla f(\hat{x})$ as a linear combination of the columns of $[D,u^{(m+2)}\ \ldots\ u^{(n)}]$ then implies $-z_1 = v'_1$, $-z_2 = v'_2$, $\ldots$, $-z_m = v'_m$, $-z_{m+1} = 0$, $-z_{m+2} = v'_{m+1},\ldots,-z_n = v'_{n-1}$; and $v_{k_o} > 0$, $-z_i \leq 0$ provides a contradiction, thus $\hat{x}$ is a minimizing solution of (3.28).

Let us review:

Starting at a vertex $\tilde{x}$ of $\Lambda = \{x|Ax = b,\ x \geq \theta\}$, a path along an edge of Λ is determined. This path is followed to the point $\hat{x}$, possibly an adjacent vertex of Λ. If that be the case, the process employed is the simplex algorithm. If $\hat{x}$ is not a vertex of Λ, then the values

$\bar{c}_j(\hat{x})$ are determined and if $\bar{c}_j(\hat{x}) \geq 0$ for $(j = m+2, \ldots, n)$, $\hat{x}$ is optimal for (3.28). Otherwise $\hat{x}$, a vertex of

$$\Lambda^{(1)} = \{x \mid Ax = b, (\eta^{(m+1)}, x) = (\eta^{(m+1)}, \hat{x}),\ x \geq \theta\}$$

is treated in the same fashion as was $\tilde{x}$, with the following exception: In general, a path along an edge of $\Lambda^{(1)}$ (in the feasible direction of the vector $\eta^{(m+2)}$ developed in §2) lies on a 2-dimensional face of Λ. If a point $\hat{\hat{x}}$, not a vertex of $\Lambda^{(1)}$, is determined on such a path, then $\eta^{(m+2)}\eta^{(m+2)T}\nabla f(\hat{\hat{x}}) = \theta$ as before, however this need not imply that $P\nabla f(\hat{\hat{x}}) = \theta$, where $P = \eta^{(m+1)}\eta^{(m+1)T} + \eta^{(m+2)}\eta^{(m+2)T}$. At such a point one first determines whether or not $P\nabla f(\hat{\hat{x}}) = \theta$. If $P\nabla f(\hat{\hat{x}}) \neq \theta$, a new 1-dimensional path on the related 2-dimensional face of Λ is followed. (Note that if $\hat{\hat{x}}$ is a vertex of $\Lambda^{(1)}$, the process employed is the simplex algorithm.) The procedure outlined above is continued until one of two situations occurs:

(3.42) The terminal point, $\bar{x}$, lies on the related 2-dimensional face of Λ, no additional non-negativity constraints having become active, and is such that

$$P\nabla f(\bar{x}) = \theta \text{ where } P = \eta^{(m+1)}\eta^{(m+1)T} + \eta^{(m+2)}\eta^{(m+2)T};$$

or

(3.43) $\bar{x}$ lies in the intersection of a set of active constraints which is not a subset of those containing the initiating vertex $\tilde{x}$. In case situation (3.42) occurs, $\bar{x}$ is treated as a vertex of the convex set

$$\Lambda^{(2)} = \{x \mid Ax = b, (\eta^{(m+1)}, x) = (\eta^{(m+1)}, \bar{x}), (\eta^{(m+2)}, x) = (\eta^{(m+2)}, \bar{x}), x \geq \theta\}$$

in a manner completely analogous to that already discussed. It is apparent that the proof of Theorem 3.2 need be modified only slightly to support the indicated generalization; both its formulation and the modified proof are left to the reader.

The generalization of this approach to higher dimensional faces of Λ is quite apparently analogous to our discussion of 0,1 and 2 dimensional faces. Until the author has had time to ponder possible simplifications in notations and proofs, the formal steps required are also left to the (dedicated) reader.

In case situation (3.43) occurs, there is then the requirement to project on a face of Λ having a different initiating vertex than $\tilde{x}$. There does not seem to be an easy way to reverse the transformations (3.14). This does not appear to be a serious matter, however, since it seems likely that the intersecting hyperplanes at such a point will normally include many of those which intersect in the last vertex of Λ used in the process. Thus if the B^{-1} associated with this last vertex has been retained, as would usually be the case, an appropriate vertex, "adjacent" to the current point may be obtained by application of a few "Phase I" simplex steps.

To obtain the $\bar{c}_j(\hat{x})$ values using the method of simplex multipliers [1], one solves, for example,

(3.44) $$\pi T_1^{-1} = \gamma = [\bar{c}_1(\hat{x}) \; \ldots \; \bar{c}_{m+1}(\hat{x})]$$

for the row vector

(3.45) $$\pi(\hat{x}) = [\pi_1(\hat{x}) \ldots \pi_{m+1}(\hat{x})]$$

where $\bar{c}_j(x) = \dfrac{\partial f(\hat{x})}{\partial x_j}$ for $(j = 1,\ldots,m+1)$. Then the $\bar{c}_j(\hat{x})$ values for $(j = m+2,\ldots,n)$ are obtained from the relation

(3.46) $$\bar{c}_j(\hat{x}) = c_j(\hat{x}) - \pi(\hat{x}) \begin{bmatrix} P_j \\ 0 \end{bmatrix} .$$

From (3.44) (see (3.12)) we have

(3.47) $$[c_1 c_2 \; \cdots \; c_{m+1}] \;\; N_1^{(1)} N_2^{(1)} N_3^{(1)}$$
$$= [c_1 + \alpha_{1,m+1} c'_{m+1}, \cdots, c_m + \alpha_{m,m+1} c'_{m+1}, c'_{m+1}]$$

where $c'_{m+1} = \dfrac{c_{m+1} - \sum_{i=1}^{m} \alpha_i c_i}{\sigma_{m+1}}$.

Thus

(3.48) $$[c_1 c_2 \cdots c_{m+1}]T_1 = [[c_1 \cdots c_m]B^{-1} + c'_{m+1} P^T_{m+1} B^{-1}, c'_{m+1}] = [\pi_1 \pi_2 \cdots \pi_{m+1}] .$$

A similar relation may be developed for subsequent transformations T_{k+1}.

Concluding remark: In the approach given here the simplex algorithm is used in generating the projection matrix $I-N_q N_q^+$, whereas Rosen gives a method for obtaining $N_q N_q^+$ [6] based upon an algorithm involving $(N_q^T N_q)^{-1}$. Since it appears that dim $\mathcal{R}(I-N_q N_q^+)$ will usually be much smaller than dim $\mathcal{R}(N_q N_q^+)$, one might expect significant computational improvements. This expectation is further enhanced by the potential use of a variation of the product form of the inverse in computing the vectors which constitute the required projection and by the use of "simplex multipliers" and "relative cost factors" [1] in the standard fashion of linear programming algorithm technology.

REFERENCES

(1) Dantzig, G.B., "Linear programming and extensions," Princeton U. Press, (1963).

(2) Mangasarian, O.L., "Equivalence in nonlinear programming," NRL Quarterly, Vol. 10, #4 (Dec. 1963), pp. 299-306.

(3) Penrose, R., "A generalized inverse for matrices," Proc. Cambridge Philos. Soc. 51 (1955), pp. 406-413.

(4) Pyle, L.D., "The generalized inverse in linear programming - Basic structure," to appear J. SIAM.

(5) Pyle, L.D., and R.E. Cline, "The generalized inverse in linear programming-Interior gradient projection methods," to appear J. SIAM.

(6) Rosen, J.B., "The gradient projection method for nonlinear programming. Part 1: Linear constraints," J. SIAM 8 (1960), pp. 181-217.

(7) Witzgall, C., "On the gradient methods of R. Frisch and J.B. Rosen," in Recent Advances in Math. Prog., ed. R.L. Graves and P. Wolfe, McGraw-Hill (1963), p. 87.