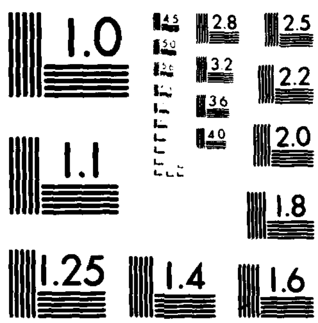
AD-A094 232 MARYLAND UNIV COLLEGE PARK DEPT OF COMPUTER SCIENCE F/G 12/1
A NEW METHOD FOR SOLVING LINEAR INEQUALITIES.(U)
NOV 80 G W STEWART N00014-76-C-0391
UNCLASSIFIED TR-970 NL

1 OF 1
AD A
094232

END
DATE FILMED
2-81
DTIC



MICROCOPY RESOLUTION TEST CHART

LEVEL

COMPUTER SCIENCE TECHNICAL REPORT SERIES

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742

6

12
T - 77

9
Technical Report 970 ✓

November 1980

12 12

6
A New Method for Solving Linear Inequalities

14
G. W. Stewart*

15 N00014-76-C-0391

Abstract

This paper describes a new method for finding nontrivial solutions of the inequality $Ax \geq 0$, where A is an $m \times n$ matrix of rank n. The method is based on the observation that a certain function f has a unique minimum if and only if the inequality fails to have a nontrivial solution. Moreover, if there is a solution, the direction of divergence of an attempt to minimize f will converge to a solution. The technique can also be used to solve inhomogeneous inequalities and hence linear programming problems, although no claims are made about competitiveness with existing methods.

*Department of Computer Science and Institute for Physical Sciences and Technology, University of Maryland. This work was supported in part by the Office of Naval Research under Contract No. N00014-76-C-0391.

46

A New Method

for Solving Linear

Inequalities

G.W. Stewart

1. Introduction

In this note we outline a method for solving linear inequalities of the form

$$(1.1)\qquad Ax \geq b,$$

where A is an $m\times n$ matrix of rank n. Since any linear programming problem may be case in this form, the method is also a new method for linear programming.

The heart of the method is a technique for either solving the homogeneous inequality

$$(1.2)\qquad Ax \geq 0, \qquad x \neq 0,$$

or determining that no solution exists. The underlying idea is simple. Consider the function

$$(1.3)\qquad f(x) = \underset{\sim}{1}^T \exp(-Ax)$$

where $\underset{\sim}{1} = (1,1,\ldots,1)^T$ and for any vector y

$$\exp(y) = (e^{y_1}, e^{y_2}, \ldots, e^{y_m}).$$

We shall show that one of two things must happen if f is minimized iteratively. If (1.2) has no solution, then $f(x)$ has a unique minimum to which the iteration must converge. If (1.2) has a solution, then the iterates will grow unboundedly in such a way that a solution can be computed from them.

The next two sections of this note will be devoted to the homogeneous inequality. In the last section we shall treat the general linear inequality (1.1).

2. Preliminaries

We shall use the following notation. Let a_i^T denote the i-th row of A. For any vector x set

$$P(x) = \{i: a_i^T x > 0\},$$
$$Z(x) = \{i: a_i^T x = 0\},$$
$$N(x) = \{i: a_i^T x < 0\}.$$

If x is a solution of (1.2), then $N(x) = \emptyset$. If x_1 and x_2 are solutions, then $x_1 + x_2$ is a solution and

$$P(x_1 + x_2) = P(x_1) \cup P(x_2),$$
$$Z(x_1 + x_2) = Z(x_1) \cap Z(x_2).$$

It follows that if there exists a solution to (1.2), then there exists a solution x^* for which the cardinality of $Z(x^*)$ is minimal. This minimally active solution need not be unique, but the sets $P^* = P(x^*)$ and $Z^* = Z(x^*)$ are.

A transformed version of the problem will be needed in the sequel. Suppose that (1.2) has a solution and let x^* be a minimally active solution. Without loss of generality, we may assume that the rows indexed by Z^* are the last rows of A; i.e.

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$$

where $A_1 x^* > 0$ and $A_2 x^* = 0$. Since $x^* \neq 0$, it follows that A_2 has a nontrivial null space. Let $V = (V_1\ V_2)$ be an orthogonal matrix with the columns of V_1 spanning the null space of A_2. If we set

$$B_{ij} = A_i V_j \tag{2.1}$$

and

$$u_i = V_i^T x,$$

then the inequality (1.2) becomes

$$Bu \equiv \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \geq 0. \tag{2.2}$$

Here both B_{11} and B_{22} have full column rank. Moreover, the minimally active solution u^* corresponding to x^* satisfies

$$B_{11}u_1^* > 0, \quad u_2^* = 0. \tag{2.3}$$

3. Homogeneous inequalities.

We begin by establishing some elementary facts about the function f defined by (1.3). Clearly f is bounded below by zero. Its gradient and Hessian are given by

$$f'(x) = -A^T \exp(-Ax)$$

and

$$f''(x) = A^T D(x) A,$$

where

$$D(x) = \mathrm{diag}(e^{a_1^T x}, e^{a_2^T x}, \ldots, e^{a_m^T x}).$$

Since $D(x)$ is positive definite and A is of full column rank, $f''(x)$ is positive definite. It follows that f is strictly convex and can have at most one local minimum, which, when it exists, is also a global minimum [2 , §3.4.6, §4.2.7]. Sufficient conditions for the existence of a minimum are contained in the following theorem.

<u>Theorem</u> 3.1 If (1.2) has no solution then f has a minimum.

<u>Proof</u>. It is sufficient to show that $f(x) \to +\infty$ as $||x|| \to \infty$ (any norm) [2,§4.3.3]. For any x with $||x|| = 1$ set

$$\phi(x) = \min \{a_i^T x: i \in N(x)\}.$$

Since (1.2) has no solution, $N(x)$ is nonempty and $\phi(x) < 0$. Clearly $\phi(x)$ is continuous. Hence

$$\theta = \sup_{||x||=1} \phi(x) < 0.$$

Now for any $x \neq 0$

$$f(x) \geq \sum_{i \in N(x)} e^{-a_i^T x} \geq e^{-\theta ||x||}$$

which establishes the theorem.

The condition in Theorem 3.1 is also necessary. However, for the purpose of this note we must take a more detailed look at the properties of f when (1.2) has a solution. This is most conveniently done in terms of the transformed system (2.2) and the associated function

$$\begin{aligned} g(u) &= \underset{\sim}{1}^T \exp (Bu) \\ &= \underset{\sim}{1}^T \exp (B_{11}u_1 + B_{12}u_2) + \underset{\sim}{1}^T \exp (B^{22}u^2) \\ &\equiv g_1(u) + g_2(u_2). \end{aligned}$$

<u>Lemma</u> 3.2. The system

$$B_{22}u_2 \geq 0 \tag{3.1}$$

has no nontrivial solution. Hence $g_2(u_2)$ has a unique minimum

$$\gamma = g_2(u_2'). \tag{3.2}$$

Proof. Suppose u_2 is a nontrivial solution of (3.1). Because B_{22} is of full rank, $B_{22}u_2 \neq 0$ and hence $B_{22}u_2$ has at least one positive component. From (2.3) it follows that there is an $\sigma > 0$ such that

$$\sigma B_{11}u_1^* + B_{12}u_2 > 0.$$

Hence the vector

$$\tilde{u} = \begin{bmatrix} \sigma u_1^* \\ u_2 \end{bmatrix}$$

is a solution of (1.2) with $Z(\tilde{u})$ a proper subset of $Z(u^*)$, which contradicts the minimality of $Z(u^*)$. The existence of a unique minimum now follows from Theorem 3.1.

Theorem 3.2. The function g satisfies

(3.3) $$g(u) > \inf g(v) = \gamma$$

where γ is defined by (3.2). Moreover, if $u^{(k)}$ is any sequence with $g(u^{(k)}) \to \gamma$ then

(3.4)
1. $u_2^{(k)} \to u_2'$
2. $B_{11}u_1^{(k)} > 0\{-\ln[g(u^{(k)}) - \gamma]\}$

Proof: For any vector u we have

(3.5) $$g(u) = g_1(u) + g_2(u_2) > g_2(u_2) \geq \gamma$$

On the other hand if we define

$$u_\alpha = \begin{bmatrix} \alpha u_1^* \\ u_2' \end{bmatrix}$$

Then

$$\lim_{\alpha\to\infty} g(u_\alpha) = \lim_{\alpha\to\infty} g_1(\alpha u) + g_2(u_2')$$

$$= 0 + \gamma = \gamma.$$

This establishes (3.3).

To establish (3.4), let $u^{(k)}$ be any sequence with $g(u^{(k)}) \to \gamma$. Then in view of (3.5) we must have $g_2(u_2^{(k)}) \to \gamma$ which implies (3.4.1). Now since we must have $g_1(u^{(k)}) \leq g(u^{(k)}) - \gamma$, it follows that

$$\exp(-B_{12}u_2^{(k)}) \odot \exp(-B_{11}u_1^{(k)}) \leq g(u^{(k)}) - \gamma$$

where $\odot$ denotes componentwise multiplication. Hence

$$\exp(-B_{11}u_1^{(k)}) \leq [g(u^{(k)}) - \gamma]\exp(B_{12}u_2^{(k)})$$

or

$$B_{11}u_1^{(k)} \geq - \ln[g(u^{(k)}) - \gamma] - B_{12}u_2^{(k)}. \tag{3.6}$$

Since $u_2^{(k)}$ is converging, (3.6) is equivalent to (3.4.2).

We are now in a position to describe a method for solving the system (1.2). Let ρ_k be a diverging sequence of radii and for each k let $x^{(k)}$ be the solution to

$$\text{minimize } f(x), \quad \text{subject to } \|x\|_2 \leq \rho_k, \tag{3.7}$$

where $\|\cdot\|_2$ denotes the usual Euclidean norm. The convexity of f and the constraint insure that $x^{(k)}$ is uniquely defined.

Now if (1.2) has no solution, theorem 3.1 assures us that for some finite k the solution $x^{(k)}$ will lie in the interior of the constraint.

On the other hand, if (1.2) has a solution, each $x^{(k)}$ will lie on the boundary of the constraint (i.e. $\|x^{(k)}\|_2 = p_k$). Moreover, from (3.4) it follows that the components of $Ax^{(k)}$ corresponding to P^* will diverge to

$+\infty$ while the components corresponding to N^* will converge. Once N^* has been recongnized, we can compute the transformation V to the u-coordinate system [cf.(2.1)]. For each k we then compute trial solutions of the form

$$\bar{x}^{(k)} = V_1 V_1^T x^{(k)} = V_1 u_1^{(1)}$$

(i.e. the vectors obtained by setting $u_2^{(k)} = 0$ so that $A_2\bar{x}^{(k)} = 0$). It is possible that initially the $\bar{x}^{(k)}$ may not be solutions, owing to the supression of the terms $B_{12}u_2^{(k)}$ in $A_1\bar{x}^{(k)}$; but (3.4.3) insures that ultimately $B_{11}u_2^{(k)} = A_1\bar{x}^{(k)} > 0$, and at that point $\bar{x}^{(k)}$ is a solution.

There are a number of observations to be made about this method.

Remark 1. The method produces a minimally active solution. This will be important in the next section.

Remark 2. We have not specified any particular algorithm for solving the constrained problem (3.7). However, we note that it is about as nice a problem as one could wish for. Within the constraint f is uniformly convex, and its first and second derivatives are easily computed. The constraint is not only convex, but the projection onto its boundary is trivially computed. Finally, in the passage from ρ_k to ρ_{k+1}, the vector $\rho_{k+1}x^{(k)}/\|x^{(k)}\|_2$ makes a natural starting point from which to find $x^{(k+1)}$.

Remark 3. Strictly speaking the method is not an iterative method, but a finitely terminating method with an inner calculation -- the solution of (3.7) -- that will almost certainly be done by iterative techniques. A true iteration may be obtained by observing that the vectors $x^{(k)}/\|x^{(k)}\|$ converge to a solution of (1.2). However, the solution thus obtained may not be minimally active, since there is nothing to prevent the components of $A_1x^{(k)}$ from diverging at different rates. The author conjectures that this will not happen if for the sequence $x^{(k)}$ generated by (3.7).

Remark 4. A possible alternative is to turn an unconstrained minimization method loose on f and see what happens. This will clearly be satisfactory when (1.2) has no solution, since most well constructed minimization methods are globally convergent when applied to convex functions. Unfortunately for the case where (1.2) has a solution, the divergence properties of minimization methods have not been well studied. Not only may the iteration produce a solution that is not minimally active, but one must deal with the possibility that it may produce values uniformly greater than the infimum γ. Nonetheless, the simplicity of the resulting class of methods makes the analysis well worth undertaking.

Historical note. The method described in this note has its origins in the maximum likelihood analysis of log-linear models. Specifically, one is given a vector $\underset{\sim}{n}$ of independent Poisson random variables with $E(\underset{\sim}{n}) = \exp(\underset{\sim}{U}\underset{\sim}{b})$, where $\underset{\sim}{U}$ is a matrix, often of formidable size. Up to a constant, the log-likelihood function for any estimate of $E(\underset{\sim}{n})$ is

$$\ell(\underset{\sim}{n}, \underset{\sim}{b}) = \underset{\sim}{n}^T \underset{\sim}{U}\underset{\sim}{b} - \underset{\sim}{1}^T \exp(\underset{\sim}{U}\underset{\sim}{b}), \tag{3.8}$$

which is seen to be an elaboration of our f. When some of the elements of $\underset{\sim}{n}$ are zero, the maximum likelihood estimate of the corresponding means may also be zero, in which case an iteration for maximizing (3.8) must diverge, since $\underset{\sim}{U}\underset{\sim}{b}$ estimates the logarithm of the means.

Haberman [1, Theorem 2.3] has given a necessary and sufficient geometric condition for the existence of a maximum of $\ell(\underset{\sim}{n},\underset{\sim}{b})$. In the course of work on the numerical analysis of log-linear models, the author reduced this conditon to the nonexistence of a solution of (1.2), where A will generally have dimensions much smaller than U. It was then an easy matter to recognize that the inequality corresponded to a trivial log-linear model with a log-likelihood function that is essentially $-f$. Thus the existence or nonexistence of a solution of (1.2) is equivalent to the nonexistence or existence of a minimum of f.

4. Coda — inhomogeneous inequalities.

It is easy to see that the inequality (1.1) is equivalent to the existence of a solution of the inequality

$$\begin{bmatrix} A & -b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ \eta \end{bmatrix} \gtrsim 0 \tag{4.1}$$

with $\eta > 0$, in which case $x = y/\eta$. Since the method of §3 produces minimally active solutions, when applied to (4.1) it will either produce a solution with $\eta > 0$, or it will produce one with $\eta = 0$ or indicate no solution, in which case (1.1) has no solution.

References

1. Haberman, S.J., The Analysis of Frequency Data, The University of Chicago Press, Chicago, 1974.

2. Ortega, J.M. and Rheinboldt, W.C., Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.


-8