AD 758 635

LIBRARY
TECHNICAL REPORT SECTION
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93940

AD

NAVAL PERSONNEL AND TRAINING RESEARCH LABORATORY

SAN DIEGO, CALIFORNIA 92152

TECHNICAL BULLETIN STB 73-4 APRIL 1973

A PRIMAL-DUAL METHOD FOR MINIMIZATION WITH LINEAR CONSTRAINTS

Gordon B. Hatfield

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

A PRIMAL-DUAL METHOD FOR MINIMIZATION WITH LINEAR CONSTRAINTS

By

Gordon B. Hatfield

April 1973

PF55.521.010.01.08
Research Bulletin STB 73-4

Submitted by

J. Silverman, Acting Director
Personnel Management Systems Research Department

Approved by

F. L. Nelson, Captain, USN
Commanding Officer

Approved for public release; distribution unlimited

Naval Personnel and Training Research Laboratory
San Diego, California 92152

A LABORATORY OF THE BUREAU OF NAVAL PERSONNEL

DEPARTMENT OF THE NAVY
NAVAL PERSONNEL AND TRAINING
RESEARCH LABORATORY
SAN DIEGO, CALIFORNIA 92152

IN REPLY REFER TO:

40:mjs
PF55.521.010.01.08
Ser 210
20 April 1973

From: Commanding Officer
To: Distribution List

Subj: Research on decision systems for personnel planning; forwarding of report on

Encl: (1) Technical Bulletin STB 73-4, A Primal-Dual Method for Minimization With Linear Constraints of April 1973 with Summary attached

1. Under the direction of the Chief of Naval Personnel (Pers-A3), this Laboratory is conducting subject research and is distributing as enclosure (1) copies of a report to addressees for information and retention. The Summary attached to each report gives a brief overview of the work and the final conclusions and recommendations which should be of use to those not having the time or need to read the entire report.

F L Nelson
F. L. NELSON

Distribution:
Chief of Naval Operations (OP-01)
(OP-098T)
(OP-099)
(OP-39)
(OP-964)
Commandant of the Marine Corps, G-1 Division, Code A01M
Chief of Naval Training (Code N-33)
Chief of Naval Material (MAT 03P)
Commander Antisubmarine Warfare Force, U. S. Atlantic Fleet (Code 70)
Office of Naval Research (Code 458) (2)
Naval Postgraduate School
Naval Examining Center
Center for Naval Analyses
Naval Personnel Research and Development Laboratory
Office of the Assistant Secretary of the Navy
(Manpower and Reserve Affairs)
Chief of Research and Development, U. S. Army
Army Research Institute for Behavioral and Social Sciences
Personnel Research Division, Lackland AFB
Personnel Management Development Office, OPO, U. S. Army
Human Factors Operations Research Laboratory, Bolling AFB
Science and Technology Divison, Library of Congress
Defense Documentation Center (Attn: DDC-TC) (12)

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

Naval Personnel and Training Research Laboratory
San Diego, California 92152

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

A Primal-Dual Method for Minimization With Linear Constraints

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

Gordon B. Hatfield

6. REPORT DATE

April 1973

7a. TOTAL NO. OF PAGES

62

7b. NO. OF REFS

13

8a. CONTRACT OR GRANT NO.

b. PROJECT NO. PF55.521.010.01.08

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

STB 73-4

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Chief of Naval Personnel (Pers-A3)
Navy Department
Washington, D. C. 20370

13. ABSTRACT

The purpose of this report is to develop a general algorithm for solving the class of nonlinear programming problems that have linear constraints. The constraints can be either equations or inequalities and the variables can be free or non-negative. The objective function is assumed to be continuously differentiable. The algorithm is an "effective" second-order method in that slow convergence is eliminated without requiring second partial derivatives. In addition it combines the desirable features of projection methods, conjugate gradient methods, and methods that solve LP problems to obtain feasible directions. Computational results on a wide variety of test problems are given. Some comments on the efficiency of the algorithm as compared to other algorithms is included.

DD FORM 1 NOV 65 1473 (PAGE 1)

0102-014-6600

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Nonlinear Programming						
Algorithm						
Primal-Dual Method						
Linear Constraints						
Convergence						
Feasible Direction						
Projection						
Conjugate Gradient Method						
Unconstrained Minimization						
Ellipsoid						
Operations Research						
Mathematical Programming						
Optimization						
Management Science						

DD FORM 1 NOV 65 1473 (BACK)

(PAGE 2)

UNCLASSIFIED

Security Classification

SUMMARY AND CONCLUSIONS

Problem

Increasingly, the research community--particularly that inhabited by management scientists and operations researchers--have become sensitive to the need for new or more realistic approaches to the solution of large, complex decision problems. This new interest shifted the focus of attention from operational, tightly-constrained, single-objective decisions at lower levels of organizations to planning and policy decisions at the upper level. The shift has not yet been rewarded with the same kind of dramatic success that was associated with highly-structured operational problems (e.g., the development of linear programming). Substantively, the impetus for this research was a problem in planning promotions for naval enlisted personnel. Upon investigation, it became clear that the Navy's enlisted advancement system sought to achieve multiple, conflicting objectives associated with authorized strength levels, meeting manpower requirements, minimizing cost, and maximizing promotion opportunity, among others. In attempting to develop an optimization technique to handle such problems, it was determined that an effective nonlinear programming algorithm would be required. As a result, it was necessary to address this problem before progress was possible in the case of the enlisted promotion planning problem. Specifically, this report documents the development of a general algorithm (and the theory that underlies it) for solving the class of nonlinear programming problems that have linear constraints. The linear constraints can be either equations or inequalities and the variables can be free or non-negative. The objective function is assumed to be continuously differentiable. This class of problems, i.e., those having linear constraints and a nonlinear objective, seems to be particularly appropriate for manpower management.

Background

Under the direction of the Chief of Naval Personnel, this Laboratory is conducting a research program in the area of enlisted personnel planning. The thrust of this program is toward the development of computer-assisted decision systems for more effective personnel planning. In the course of this research it became apparent that the ability to achieve personnel management objectives is heavily dependent on actions taken in the area of enlisted promotions. This awareness generated an investigation of the processes underlying the enlisted promotion system, resulting in the development of new techniques for planning promotions. Still unsolved, however, was the problem of achieving multiple personnel management objectives which involved drastic tradeoffs in planning the number of promotions in each pay grade of each rating over time. Existing techniques available for the solution of such problems, particularly those of a non-linear character, indicated some technical and logical difficulties in application.

Approach

During the last 10 to 15 years many algorithms have been developed for solving nonlinear programming problems. These algorithms range from those that apply to special classes of problems to some quite general algorithms applicable to broad classes of problems. The result of this development is a variety of methods, each with its particular strengths and weaknesses. It is expected that this trend in algorithm development will continue. However, while it is probable that no universal method will emerge that is clearly superior to all other methods for all problems, it seems quite reasonable to assume that new ideas and techniques will increase the variety of methods and algorithms available. Certain general techniques have been used in solving nonlinear programming problems and have been successful in practice. These techniques involve the following approaches: (1) solve a local linear programming problem for a feasible direction; (2) projection, i.e., project an infeasible direction onto the feasible set; and (3) solve a local conjugate gradient problem for curvature information. It is interesting to note that algorithms are usually developed in terms of only one of these techniques rather than a combination. However, there is a method that combines techniques (2) and (3). The approach of the method developed herein, called the primal-dual method, is to combine the desirable features of each of these techniques into a single algorithm. Thus, a large number of existing algorithms become special cases of the primal-dual algorithm. In addition to providing a partial unification of nonlinear programming algorithms, a very efficient algorithm results.

Findings, Conclusions, and Recommendations

The primal-dual algorithm is an "effective" second-order method in that slow convergence is eliminated without requiring second partial derivatives. It has additional advantages in that (1) zig-zagging due to highly eccentric ellipsoids is eliminated, (2) if the stationary point occurs at an interior point, no linear programming problems are solved and the solution is identified immediately, and (3) accurate bounded variable constraints are generated from the conjugate gradient solution. Computational results on a wide variety of small (in size) test problems indicate the validity of these findings. It is concluded that the algorithms of Rosen, Goldfarb, and a special case of Graves algorithm* can be viewed as special cases of the primal-dual method. A future report will describe the application of this method to the enlisted promotion problem.

---

*See References [3] through [6] and [12].

REPORT USE AND EVALUATION

Feedback from consumers is a vital element in improving products so that they better respond to specific needs. To assist the Chief of Naval Personnel in future planning, it is requested that the use and evaluation form on the reverse of this page be completed and returned. The page is preaddressed and franked; fold in thirds, seal with tape, and mail.

Department of the Navy

Postage and Fees Paid
Navy Department

Official Business

Commanding Officer
Naval Personnel and Training Research Laboratory
San Diego, California 92152

Report Title & No: A Primal-Dual Method for Minimization With Linear Constraints (STB 73-4)

1. Evaluation of Report. Please check appropriate column.

FACTORS	RATING LOW	AVE	HIGH	COMMENTS
Usefulness of Data				
Timeliness				
Completeness				
Technical Accuracy				
Validity of Recommendations				
Soundness of Approach				
Presentation and Style				
Other				

2. Use of Report. Please fill in answers as appropriate.

a. What are your main uses for the material contained in the report?

b. What changes would you recommend in report format to make it more useful?

c. What types of research would be most useful to you for the Chief of Naval Personnel to conduct?

d. Do you wish to remain on our distribution list?

e. Please make any general comments you feel would be helpful to us in planning our research program.

NAME: _____ CODE: _____

ORGANIZATION: _____

ADDRESS: _____

CONTENTS

		Page
I.	INTRODUCTION	1
II.	UNCONSTRAINED MINIMIZATION	3
III.	DETERMINATION OF A FEASIBLE DIRECTION	10
IV.	DETERMINATION OF THE DISTANCE MOVED	19
V.	CONVERGENCE OF THE ALGORITHM	22
VI.	DISCUSSION	25
VII.	NUMERICAL RESULTS	28
VIII.	TEST PROBLEMS	32
IX.	SOLUTION OF THE TEST PROBLEMS	42
	REFERENCES	52

A FIRST-ORDER, PRIMAL-DUAL METHOD FOR MINIMIZING A REAL-VALUED FUNCTION SUBJECT TO LINEAR CONSTRAINTS

I. INTRODUCTION

The purpose of this paper is to develop a general algorithm for solving the problem

subject to:

$$\begin{aligned} \sum_{j=1}^{n} a_{ij} y_j &= r_i & i &= 1 \ldots m_1 \\ \sum_{j=1}^{n} a_{ij} y_j &\leq r_i & i &= m_1+1, \ldots m \\ y_j &\geq 0 & j &= n_1+1, \ldots n \end{aligned} \tag{1}$$

minimize:

$$f(y_1, \ldots y_n)$$

where m_1 is the number of equations and n_1 is the number of free variables. The function $f(\underline{y})$ is assumed to be of class C^1, i.e., continuously differentiable.* The algorithm is a first order method in that only the first partial derivatives of the objective function are used. The conjugate gradient method is used to obtain an unconstrained stationary point when one exists. The particular conjugate gradient method used is quite efficient in that no linear searches are required.

In the vicinity of an unconstrained minimum the second-order terms in the Taylor series expansion of $f(\underline{y})$ dominate. However in the vicinity

---

*vectors are denoted by underlined lower case letters.

of a constrained minimum the first-order terms dominate. Thus the method should have quadratic convergence (without requiring the second partial derivatives) combined with the ability to obtain feasible directions. The algorithm presented here has the benefits of a second-order method in eliminating slow convergence without requiring the second partial derivatives. It also has additional advantages: (1) zig-zagging due to highly eccentric ellipsoids is eliminated; (2) if the stationary point occurs at an interior point, no linear programming (LP) problems are solved and the solution is identified immediately; and (3) accurate bounded variable constraints are generated from the conjugate gradient solution.

In order to present a complete theoretical development of the algorithm we begin by discussing unconstrained minimization and the conjugate gradient method. This is followed by the development of a first order accelerated conjugate gradient method. Next it is shown that a feasible direction for the constrained problem can be obtained by solving an LP problem and the method for determining the distance moved is given. Further, it is shown that the Kuhn-Tucker conditions are satisfied at the constrained minimum. A proof of convergence concludes the theoretical section. In the numerical section a number of test problems are presented. The algorithm is compared with existing algorithms by solving the test problems using total computation time as the basis of comparison.

II. UNCONSTRAINED MINIMIZATION

It is not difficult [8] to show the following.

Theorem 1: For $f \in C^2$ let $\underline{x}^*$ be a point for which $\nabla f(\underline{x}^*) = \underline{0}$; then $f(\underline{x})$ assumes a relative minimum at $\underline{x}^*$ if the quadratic form $(\underline{x}-\underline{x}^*)' H(\underline{x}^*)(\underline{x}-\underline{x}^*)$ is positive definite, or equivalently, if every eigenvalue of $H(\underline{x}^*)$ is positive.

Here $H = \| \partial^2 f / \partial x_i \partial x_j \|$ is the $n \times n$ Hessian matrix. It is also true that the quadric $(\underline{x}-\underline{x}^*)' H(\underline{x}^*)(\underline{x}-\underline{x}^*) = 1$ is an ellipsoid if all the eigenvalues of $H(\underline{x}^*)$ are positive. Consider the problem

$$\text{minimize } f(\underline{x})$$
$$\underline{x} \in E^n .$$

If $f \in C^2$ and $\underline{x}$ is an estimate of $\underline{x}^*$ the relative minimum then by Taylor's Theorem,

$$f(\underline{x}) = f(\underline{x}^*) + \tfrac{1}{2} \underline{y}' H(\underline{x}^*) \underline{y} + r(\underline{x}^*, \underline{y}),$$

where $\underline{y} = \underline{x} - \underline{x}^*$. When the error term is zero, minimizing $f(\underline{x})$ is equivalent to minimizing the quadratic form $\underline{y}' H(\underline{x}^*) \underline{y}$, which is an ellipsoid since all the eigenvalues of $H(\underline{x}^*)$ are positive. If $\underline{x} = \underline{x}_i$ is an estimate of $\underline{x}^*$ then

$$(\underline{x}_i - \underline{x}^*)' H(\underline{x}^*)(\underline{x}_i - \underline{x}^*) = k_i \qquad i = 1, 2, \ldots .$$

Thus we want to find a sequence of points $\{\underline{x}_i\}$ such that the sequence $\{k_i\} \to 0$. Geometrically this is equivalent to finding the center of the ellipsoid $\underline{y}' H(\underline{x}^*) \underline{y}$.

We can extend these ideas as follows: let $\underline{x}_1$ be an estimate of $\underline{x}^*$ with $f \in C^2$. Then by Taylor's Theorem

$$f(\underline{x}) = f(\underline{x}_1) + \nabla f(\underline{x}_1)\underline{y} + \frac{1}{2}\underline{y}' H(\underline{x}_1)\underline{y} + r(\underline{x}_1, \underline{y}) ,$$

where $\underline{y} = \underline{x} - \underline{x}_1$. Let

$$F(\underline{x}) = f(\underline{x}_1) + \nabla f(\underline{x}_1)\underline{y} + \frac{1}{2}\underline{y}' H(\underline{x}_1)\underline{y} .$$

We now show that $F(\underline{x})$ is an approximating ellipsoid at $\underline{x}_1$. If

$$\bar{f}(\underline{x}_1) = f(\underline{x}_1) - \nabla f(\underline{x}_1)\underline{x}_1 + \frac{1}{2}\underline{x}_1' H(\underline{x}_1)\underline{x}_1 ,$$

then

$$F(\underline{x}) = \bar{f}(\underline{x}_1) + \left[\nabla f(\underline{x}_1) - \underline{x}_1' H(\underline{x}_1)\right]\underline{x} + \frac{1}{2}\underline{x}' H(\underline{x}_1)\underline{x} .$$

Now for any $\underline{\alpha}$

$$F(\underline{x}) = \bar{f}(\underline{x}_1) - \frac{1}{2}\underline{\alpha}' H(\underline{x}_1)\underline{\alpha} + \left[\nabla f(\underline{x}_1) - \underline{x}_1' H(\underline{x}_1) + \underline{\alpha}' H(\underline{x}_1)\right]\underline{x}$$

$$+ \frac{1}{2}(\underline{x} - \underline{\alpha})' H(\underline{x}_1)(\underline{x} - \underline{\alpha}) .$$

Since a necessary condition for $H(\underline{x}_1)$ to be positive definite is that $\det H(\underline{x}_1) > 0$, it follows that $H(\underline{x}_1)^{-1}$ exists [11]. Suppose we choose $\underline{\alpha}$ such that

$$\nabla f(\underline{x}_1) - \underline{x}_1' H(\underline{x}_1) + \underline{\alpha}' H(\underline{x}_1) = \underline{0} .$$

Then

$$\underline{\alpha} = \underline{x}_1 - H(\underline{x}_1)^{-1} \nabla f(\underline{x}_1)' ,$$

and

$$F(\underline{x}) = \bar{f}(\underline{x}_1) - \frac{1}{2} \underline{\alpha}' H(\underline{x}_1) \underline{\alpha} + \frac{1}{2} (\underline{x} - \underline{\alpha})' H(\underline{x}_1) (\underline{x} - \underline{\alpha})$$

is an approximating ellipsoid at $\underline{x}_1$. Since $H(\underline{x}_1)$ is positive definite

$$(\underline{x} - \underline{\alpha})' H(\underline{x}_1) (\underline{x} - \underline{\alpha}) \geq 0 ,$$

with equality holding when $\underline{x} = \underline{\alpha}$. Thus the minimum of $F(\underline{x})$ occurs at $\underline{x} = \underline{\alpha}$ and is

$$\min F(\underline{x}) = \bar{f}(\underline{x}_1) - \frac{1}{2} \underline{\alpha}' H(\underline{x}_1) \underline{\alpha} .$$

The following has been established.

Theorem 2: If $H(\underline{x}_i)$ is positive definite then the approximating ellipsoid at $\underline{x}_i$,

$$F(\underline{x}) = \bar{f}(\underline{x}_i) + \left[\nabla f(\underline{x}_i) - \underline{x}_i' H(\underline{x}_i) \right] \underline{x} + \frac{1}{2} \underline{x}' H(\underline{x}_i) \underline{x} ,$$

has the minimum value

$$\bar{f}(\underline{x}_i) - \frac{1}{2} \underline{x}_{i+1}' H(\underline{x}_i) \underline{x}_{i+1} ,$$

at the center

$$\underline{x}_{i+1} = \underline{x}_i - H(\underline{x}_i)^{-1} \nabla f(\underline{x}_i)' .$$

Theorem 2 is an alternate derivation of Newton's method which finds the center of the approximating ellipsoid in one step. At this point we present a version of the conjugate gradient (cg) method which finds the center of the approximating ellipsoid in n steps. However the cg method does not require the Hessian or its inverse. The algorithm given here was developed by Hestenes [10].* Its numerical behavior on the standard test functions has also been investigated [8].

A description of the conjugate gradient method follows.

<u>Starting routine</u>: Select an initial point $\underline{x}_1$ and compute

$$g_1 = -\nabla f(\underline{x}_1) \quad , \quad p_1 = g_1 \ .$$

<u>Main routine</u>: Given $\underline{x}_k, g_k$ and p_k choose

$$\hat{\underline{x}}_{k+1} = \underline{x}_1 + \sigma_k p_k \ , \qquad (\sigma_k > 0 \ , \ \sigma_k \text{ arbitrary}) \ ,$$

and compute

$$\hat{g}_{k+1} = -\nabla f(\hat{\underline{x}}_{k+1}) \quad , \quad \underline{s}_k = (g_1 - \hat{g}_{k+1})/\sigma_k \ ,$$

$$d_k = p_k' \underline{s}_k \quad , \quad c_k = p_k' g_k \ , \quad a_k = c_k/d_k \ ,$$

$$\underline{x}_{k+1} = \underline{x}_k + a_k p_k \quad , \quad g_{k+1} = g_k - a_k \underline{s}_k \ ,$$

$$b_k = -\frac{\underline{s}_k' g_{k+1}}{d_k} \quad , \quad p_{k+1} = g_{k+1} + b_k p_k \ .$$

The basic idea of accelerating the conjugate gradient method** stems from the fact that the ray connecting the centers of two successive

---

*The conjugate gradient method was developed by Hestenes in 1952 [9].

**This idea was suggested to me by Professor M. Hestenes, UCLA.

approximating ellipsoids generally is a good direction toward the minimum. To develop this method let $\underline{x}^*$ be a relative minimum and let $\underline{x}_0$, $\underline{x}_1$ and $\underline{x}_2$ be successive estimates of $\underline{x}^*$. Consider the Taylor's expansion of $\underline{x}_2$ about $\underline{x}_1$, where $\underline{y} = \underline{x}_2 - \underline{x}_1$,

$$f(\underline{x}_2) = f(\underline{x}_1) + \nabla f(\underline{x}_1)\underline{y} + \frac{1}{2}\underline{y}' H(\underline{x}_1)\underline{y} + r(\underline{x}_1, \underline{y}) .$$

We also require the Taylor's expansion of $\underline{x}^*$ about $\underline{x}_2$,

$$f(\underline{x}^*) = f(\underline{x}_2) + \nabla f(\underline{x}_2)\underline{z} + \frac{1}{2}\underline{z}' H(\underline{x}_2)\underline{z} + r(\underline{x}_2, \underline{z}) ,$$

where $\underline{z} = \underline{x}^* - \underline{x}_2$. If $\underline{x}_2$ is obtained from $\underline{x}_1$ by the cg method then by Theorem 2

$$\underline{y} = -H(\underline{x}_1)^{-1}\nabla f(\underline{x}_1)' .$$

We can now prove the following,

Theorem 3: Let $\underline{x}^*$ be the minimum of a convex function $f \in C^2$. If $\underline{x}_0$, $\underline{x}_1$, and $\underline{x}_2$ are successive estimates of $\underline{x}^*$ where $\underline{x}_2$ is obtained from $\underline{x}_1$ by the cg method then

$$f(\underline{x}^*) > f(\underline{x}_1) + \nabla f(\underline{x}_1)\underline{y} + \nabla f(\underline{x}_2)\underline{z}$$

is a lower bound of $f(\underline{x}^*)$, and if $\underline{y} \cdot \underline{z} > 0$,

$$\hat{\underline{x}} = \underline{x}_2 + \frac{\underline{y} \cdot \underline{z}}{\|\underline{y}\|^2} \underline{y}$$

is the minimum of $f(\underline{x})$ along the line $\underline{x} = \underline{x}_2 + k\underline{y}$.

Proof: If f is convex then by standard properties of convex functions

$$f(\underline{x}^*) \geq f(\underline{x}_2) + \nabla f(\underline{x}_2)\underline{z} \geq f(\underline{x}_1) + \nabla f(\underline{x}_1)\underline{y} + \nabla f(\underline{x}_2)\underline{z}$$

$$= f(\underline{x}_1) - \nabla f(\underline{x}_1) H(\underline{x}_1)^{-1} \nabla f(\underline{x}_1)' + \nabla f(\underline{x}_2)\underline{z}$$

If $\underline{x}_2 \neq \underline{x}^*$ then

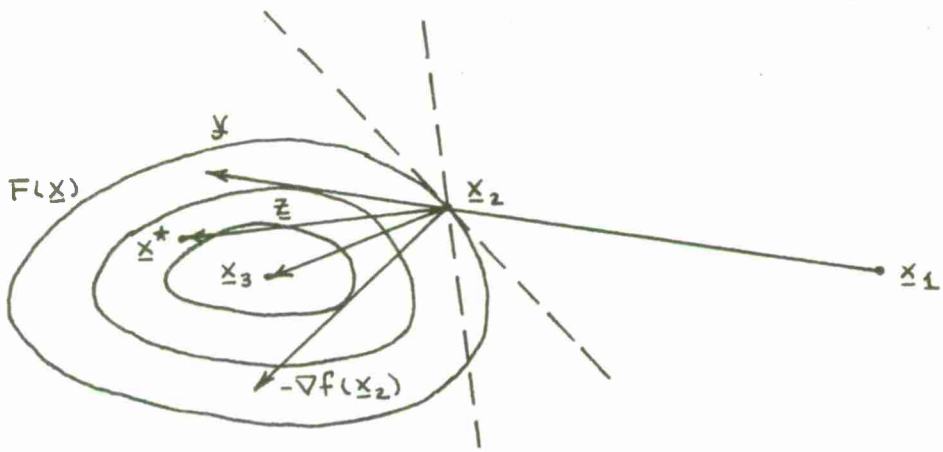
$$\frac{f(\underline{x}_2 + \lambda \underline{z}) - f(\underline{x}_2)}{\lambda} < 0 , \qquad \lambda > 0 ,$$

and

$$\lim_{\lambda \to 0} \frac{f(\underline{x}_2 + \lambda \underline{z}) - f(\underline{x}_2)}{\lambda} < 0 , \qquad \lambda > 0 ,$$

or

$$-\nabla f(\underline{x}_2)\underline{z} > 0 .$$



If $\underline{y} \cdot \underline{z} > 0$ and θ is the angle between $\underline{y}$ and $\underline{z}$ then $\cos\theta > 0$ and f can be reduced along the line $\underline{x} = \underline{x}_2 + k\underline{y}$. Let $\hat{\underline{x}} = \underline{x}_2 + \hat{k}\underline{y}$ be such that

$$\underline{y} \cdot (\underline{x}^* - \hat{\underline{x}}) = 0 .$$

Then

$$\hat{k} = \frac{\underline{y} \cdot \underline{z}}{\|\underline{y}\|^2} > 0 ,$$

since $\underline{y} \cdot \underline{z} > 0$.

This theorem is of no direct use computationally since $\underline{z}$ is unknown. However if the acceleration step is done by a linear search along $\underline{y}$ then we can restrict the search to positive values of k. Limited computational experience has shown that when acceleration can be achieved it is usually quite effective.*

---

*Acceleration works very well on Powells function of 4 variables but was completely ineffective on Rosenbrock's parabolic valley [8].

III. DETERMINATION OF A FEASIBLE DIRECTION

In the constrained problem (1), let $\underline{x}^*$ be the unconstrained stationary point if one exists. Let $\underline{y}^\circ$ (feasible) be an estimate of the constrained minimum, and $\Delta\underline{y}$ a direction emanating from $\underline{y}^\circ$. Thus we are interested in points of the form $\underline{y} = \underline{y}^\circ + \Delta\underline{y}$ where $\Delta\underline{y}$ is a feasible direction. Now $\underline{y}^\circ$ is either on the boundary of the feasible set or is an interior point. To obtain $\Delta\underline{y}$ when $\underline{y}^\circ$ is on the boundary consider the Taylor's expansion of the objective function,

$$f(\underline{y}) = f(\underline{y}^\circ) + \nabla f(\underline{y}^\circ)\Delta\underline{y} + r(\underline{y}^\circ, \Delta\underline{y}) \ . \tag{2}$$

We also require the distance function

$$d(\underline{y}) = \| \underline{x}^* - \underline{y} \|^2 = (x_1^* - y_1)^2 + \cdots + (x_n^* - y_n)^2 ,$$

and its Taylor's expansion

$$d(\underline{y}) = d(\underline{y}^\circ) + \nabla d(\underline{y}^\circ)\Delta\underline{y} + r(\underline{y}^\circ, \Delta\underline{y}) ,$$

where

$$\nabla d(\underline{y}^\circ) = -2(\underline{x}^* - \underline{y}^\circ)' \ .$$

This suggests we consider the convex combination

$$(1 - \alpha_B)\nabla f(\underline{y}^\circ)\Delta\underline{y} - 2\alpha_B(\underline{x}^* - \underline{y}^\circ)'\Delta\underline{y} , \tag{3}$$

where $0 \le \alpha_B \le 1$. This leads to a stepwise procedure where the direction $\Delta\underline{y}$ is obtained by solving the linear programming problem

subject to:

$$\begin{aligned} \sum_{j=1}^{n} a_{ij} y_j &= r_i & i &= 1 \ldots m_1 \\ \sum_{j=1}^{n} a_{ij} y_j &\le r_i & i &= m_1+1, \ldots m \\ \sum_{j=1}^{n} y_j &\le B_u & & \\ y_j &\ge 0 & j &= n_1+1, \ldots n \\ y_j &\ge -B_\ell & j &= 1 \ldots n_1 \end{aligned} \tag{4}$$

minimize

$$(1-\alpha_B)\nabla f(\underline{y}^\circ)\underline{y} - 2\alpha_B(\underline{x}^* - \underline{y}^\circ)'\underline{y} \quad .$$

In order to assure convergence to a stationary point we require $\alpha_B \to 0$. The upper bound is found by choosing B_u so that

$$\sum_{j=1}^{n} y_j \le B_u$$

is satisfied when $\underline{y} = \underline{y}^1$ (the first feasible solution) and $\underline{y} = \underline{x}^*$. The lower bound B_ℓ is found similarly. Note that for any solution $\underline{y}$ of the LP problem $\Delta\underline{y}$ is a feasible direction since the feasible set is convex.

The LP algorithm used here is the primal-dual algorithm of Graves [5].* The possible terminal states of the LP problems are:

(1) Finite optimal solution $\underline{y}$, $(\Delta\underline{y} = \underline{y} - \underline{y}^\circ)$;

(2) Inconsistent constraint (terminate because problem is inconsistent); or

(3) Unbounded solution (does not occur due to bounded variable constraints).

---

*It is particularly efficient in that no slack variables or artificial variables are required.

At the constrained minimum none of the bounded variable constraints can hold as equalities. If this condition occurs we simply increase B_u or B_ℓ and use the current feasible point as a new starting point.

In practice α_B is decreased to zero quite rapidly. When $\alpha_B = 0$ and $\underline{y}^o$ is on the boundary zig-zagging can still arise when solving (4). This is due to the eccentricity of the approximating ellipsoids. When this occurs we solve the local conjugate gradient problem to estimate the curvature of $f(\underline{y})$ in the vicinity of $\underline{y}^o$. The direction obtained is then projected onto the subspace determined by the binding hyperplanes to yield a feasible direction. The technique of projection is similar to that developed by Rosen [12]. However instead of projecting the gradient we project one of the local conjugate gradient directions $\underline{p}_k$ (see the conjugate gradient method). To develop these ideas let B be the set of binding hyperplanes at $\underline{y}^o$, i.e.,

$$B = \{ \underline{a}_i \underline{y} \le r_i \mid \underline{a}_i \underline{y}^o = r_i \text{ and } -\underline{e}_i \underline{y}^o = 0 \} .$$

Let $\underline{B}'$ be a matrix of rank h whose rows are a set of h linearly independent hyperplanes from B. Thus $\underline{B}$ is an $(n \times h)$ matrix of rank h. The subspace S generated by the columns of $\underline{B}$ is the same as that generated by the rows of $\underline{B}'$.

Now

$$E^n = S + O(S) ,$$

where $O(S)$ is the subspace which is the orthogonal complement of S. So, any vector in E^n can be written uniquely as the sum of a vector in S and

a vector in $O(S)$. If $\underline{d}$ is the direction obtained from the local conjugate gradient problem then

$$\underline{d} = \Delta \underline{y}^p + \underline{w} \ ,$$

where $\Delta \underline{y}^p \in O(S)$ and $\underline{w} \in S$. Now

$$\underline{w} = \sum_{j=1}^{h} \lambda_j b_{\cdot j} = \underline{B}\underline{\lambda} \ ,$$

giving

$$\underline{d} = \Delta \underline{y}^p + \underline{B}\underline{\lambda} \ .$$

Since $\Delta \underline{y}^p \in O(S)$,

$$\underline{B}'\underline{d} = \underline{B}'\underline{B}\underline{\lambda} \ .$$

Since $(\underline{B}'\underline{B})$ is a nonsingular matrix,

$$\underline{\lambda} = (\underline{B}'\underline{B})^{-1}\underline{B}'\underline{d} \ ,$$

giving

$$\underline{w} = \underline{B}(\underline{B}'\underline{B})^{-1}\underline{B}'\underline{d}$$

and

$$\Delta \underline{y}^p = \left[\underline{I} - \underline{B}(\underline{B}'\underline{B})^{-1}\underline{B}' \right] \underline{d} = \underline{P}\underline{d} \ ,$$

where

$$\underline{P} = \underline{I} - \underline{B}(\underline{B}'\underline{B})^{-1}\underline{B}' \ .$$

The matrix $\underline{P}$ is called a projection matrix since it projects $\underline{d}$ into $O(S)$. In practice we calculate

$$\hat{r}_k = \min_k \{ r_k = \nabla f(\underline{y}^o) \underline{P}\, \underline{p}_k \mid r_k < 0 \} ,$$

and take $\underline{d} = \hat{\underline{p}}_k$ where $\hat{\underline{p}}_k$ corresponds to $\hat{r}_k$. Actually, any $\underline{p}_k$ such that $\Delta \underline{y}^P = \underline{P}\underline{p}_k \neq \underline{0}$ and

$$\nabla f(\underline{y}^o) \underline{P}\, \underline{p}_k < 0$$

yields a direction in which to move. It is not apriori evident that $\hat{\underline{p}}_k$ is the best direction in which to move. Now since $\underline{p}_1 = -\nabla f(\underline{y}^o)'$, if we take $\underline{d} = \underline{p}_1$ and if $\underline{P}\underline{p}_1 \neq \underline{0}$ then

$$\nabla f(\underline{y}^o) \Delta \underline{y}^P = -\underline{d}' \Delta \underline{y}^P = -\Delta \underline{y}^{P'} (\Delta \underline{y}^P + \underline{\omega}) = -\Delta \underline{y}^{P'} \Delta \underline{y}^P < 0 .$$

Thus the set $\{ r_k \mid r_k < 0 \}$ is not empty and $\hat{r}_k$ exists.

Note that the calculation of the projection matrix $\underline{P}$ requires a matrix inversion. In practice this is done internally by the LP algorithm and is quite efficient.

If $\underline{y}^o$ is an interior point then $\Delta \underline{y}$ is locally unrestricted. In this case no LP problem is required. To motivate the expression for $\Delta \underline{y}$ when $\underline{y}^o$ is an interior point note that when $\alpha_B = 0$ (3) becomes

$$\nabla f(\underline{y}^o) \Delta \underline{y} .$$

If $\Delta \underline{y} = -\nabla f(\underline{y}^o)'$ then

$$\nabla f(\underline{y}^o) \Delta \underline{y} = -\| \nabla f(\underline{y}^o) \|^2 < 0 .$$

When $\alpha_B = 1$, (3) becomes

$$-2(\underline{x}^* - \underline{y}^\circ)' \Delta \underline{y} \ .$$

If $\Delta \underline{y} = (\underline{x}^* - \underline{y}^\circ)$ then

$$-2(\underline{x}^* - \underline{y}^\circ)' \Delta \underline{y} = -2 \| \underline{x}^* - \underline{y}^\circ \|^2 < 0 \ .$$

This suggests we take

$$\Delta \underline{y} = -(1 - \alpha_I) \nabla f(\underline{y}^\circ)' + \alpha_I (\underline{x}^* - \underline{y}^\circ) \ , \tag{5}$$

where $0 \leq \alpha_I \leq 1$. Substituting (5) into (3) we obtain

$$-(1-\alpha_B)(1-\alpha_I) \| \nabla f(\underline{y}^\circ) \|^2 - 2 \alpha_B \alpha_I \| \underline{x}^* - \underline{y}^\circ \|^2$$

$$+ \left[\alpha_I (1 - \alpha_B) + 2 \alpha_B (1 - \alpha_I) \right] \nabla f(\underline{y}^\circ)(\underline{x}^* - \underline{y}^\circ) \ . \tag{6}$$

The first two terms of (6) are negative and the sign of the third term depends on the angle between $\nabla f(\underline{y}^\circ)$ and $(\underline{x}^* - \underline{y}^\circ)$. Thus in order to assure a gain from an interior point when $\alpha_B = 0$ we require $\alpha_I \to 0$. With $\alpha_I = 1$ we can eliminate slow convergence due to highly eccentric ellipsoids.* When the unconstrained stationary point is not identified as optimal on the first step then the constrained minimum occurs on the boundary. So from an interior point we want to move back onto the boundary.

In practice we set $\alpha_I = 1$ and return to the boundary if a decrease in $f(\underline{y})$ is achieved on the boundary. Otherwise we decrease α_I until $\alpha_I = 0$. If $f(\underline{y})$ can not be decreased on the boundary for $\alpha_I = 0$ then a

---

*A good example of this is given in the test problems.

linear search is performed along $-\nabla f(\underline{y}^{\circ})'$. If a problem has multiple local minima and some are feasible and some are infeasible then an interesting possibility occurs. Suppose $\underline{x}^{*}$ occurs at one of the infeasible local minima. If $\underline{y}^{\circ}$ is close enough to one of the feasible local minima then the algorithm switches from trying to find a point that reduces the distance to $\underline{x}^{*}$ and finds the feasible local minima.

We now show that the Kuhn-Tucker conditions are satisfied at the constrained minimum $\underline{y}^{*}$. As an auxiliary result the classical Lagrange multipliers are given in the final LP tableau. To show this observe that the Lagrangian function of (1) is

$$F(\underline{y},\underline{\lambda}) = f(\underline{y}) + \sum_{i=1}^{m} \lambda_i \left[r_i - \sum_{j=1}^{n} a_{ij} y_j \right],$$

where the λ_i are called Lagrange multipliers. The necessary conditions for $\underline{y}^{*}$ to be a constrained minimum (Kuhn-Tucker conditions) are (see Hadley [7], chapter 6)

$$\sum_{i=1}^{m} \lambda_i^{*} a_{ij} = \frac{\partial f(\underline{y}^{*})}{\partial y_j}, \qquad j = 1 \ldots n_1$$

$$\sum_{i=1}^{m} \lambda_i^{*} a_{ij} \leq \frac{\partial f(\underline{y}^{*})}{\partial y_j}, \qquad j = n_1+1, \ldots n \tag{7}$$

$$\lambda_i^{*} \leq 0, \qquad i = m_1+1, \ldots m \tag{8}$$

$$\sum_{i=1}^{m} \lambda_i^{*} \left[r_i - \sum_{j=1}^{n} a_{ij} y_j^{*} \right] = 0, \tag{9}$$

and

$$\sum_{j=1}^{n} y_j^{*} \left[\frac{\partial f(\underline{y}^{*})}{\partial y_j} - \sum_{i=1}^{m} \lambda_i^{*} a_{ij} \right] = 0. \tag{10}$$

When $\alpha_B = 0$ and $\underline{y}^{\circ} = \underline{y}^{*}$ we have solved the LP problem

subject to:

$$\sum_{j=1}^{n} a_{ij} y_j = r_i \qquad i = 1, \ldots m_1$$

$$\sum_{j=1}^{n} a_{ij} y_j \leq r_i \qquad i = m_1+1, \ldots m$$

$$y_j \geq 0 \qquad j = n_1+1, \ldots n$$

minimize:

$$\nabla f(\underline{y}^*) \underline{y} \quad .$$

The dual of this problem is

subject to:

$$\sum_{i=1}^{m} a_{ij} x_i = \frac{\partial f(\underline{y}^*)}{\partial y_j} \qquad j = 1, \ldots n_1$$

$$\sum_{i=1}^{m} a_{ij} x_i \leq \frac{\partial f(\underline{y}^*)}{\partial y_j} \qquad j = n_1+1, \ldots n$$

$$x_i \leq 0 \qquad i = m_1+1, \ldots m$$

maximize:

$$\underline{r}\,\underline{x} \quad .$$

The necessary conditions (7) and (8) are satisfied by the dual constraints and we see that the dual variables are the Lagrange multipliers. Condition (9) says that when an inequality constraint does not hold exactly its corresponding Lagrange multiplier $\lambda_i^* = 0$. But this is equivalent to the fact that the dual variable $x_i = 0$ when its corresponding primal constraint does not hold exactly. Condition (10) can be rewritten

$$\sum_{j=1}^{n} y_j^* \frac{\partial f(\underline{y}^*)}{\partial y_j} = \sum_{i=1}^{m} \lambda_i^* \left(\sum_{j=1}^{n} a_{ij} y_j^* \right) .$$

Since $\lambda_i^* = 0$ when constraint i does not hold exactly we have

$$\sum_{j=1}^{n} y_j^* \frac{\partial f(\underline{y}^*)}{\partial y_j} = \sum_{i=1}^{m} \lambda_i^* r_i \quad ,$$

or

$$\nabla f(\underline{y}^*)\, \underline{y}^* = \underline{r}\, \underline{\lambda}^* \quad .$$

But from the dual theorem of linear programming

$$\nabla f(\underline{y}^*)\, \underline{y} = \underline{r}\, \underline{x} = \underline{r}\, \underline{\lambda}^* \quad ,$$

and since $\underline{y} = \underline{y}^\circ + \Delta \underline{y} = \underline{y}^* + \Delta \underline{y}$ we have

$$\nabla f(\underline{y}^*)\, \underline{y}^* + \nabla f(\underline{y}^*)\, \Delta \underline{y} = \underline{r}\, \underline{\lambda}^* \quad .$$

It will be shown later that the condition for termination of the algorithm is

$$| \nabla f(\underline{y}^\circ)\, \Delta \underline{y} | < \epsilon \quad .$$

From these considerations it is clear why the algorithm is called a primal-dual method. The feasible direction is determined from the primal problem, and the identification of the constrained minimum (by the Kuhn-Tucker conditions) follows from the dual problem. In this framework Wolfe's algorithm for quadratic programming would be considered a dual method.

IV. DETERMINATION OF THE DISTANCE MOVED

We now determine the distance moved along the direction $\Delta \underline{y}$. Consider the Taylor series expansion of constraint i about the feasible point $\underline{y}^{\circ}$,

$$g_i(\underline{y}) = g_i(\underline{y}^{\circ}) + \nabla g_i(\underline{y}^{\circ}) \Delta \underline{y} \ ,$$

or

$$\underline{a}_i \underline{y} = \underline{a}_i \underline{y}^{\circ} + \underline{a}_i \Delta \underline{y} \leq r_i \ .$$

Let

$$\Delta \underline{y} = k \Delta \underline{u} \quad \text{where} \quad \Delta \underline{u} = \frac{\Delta \underline{y}}{\| \Delta \underline{y} \|} \quad \text{and} \quad k > 0 \ .$$

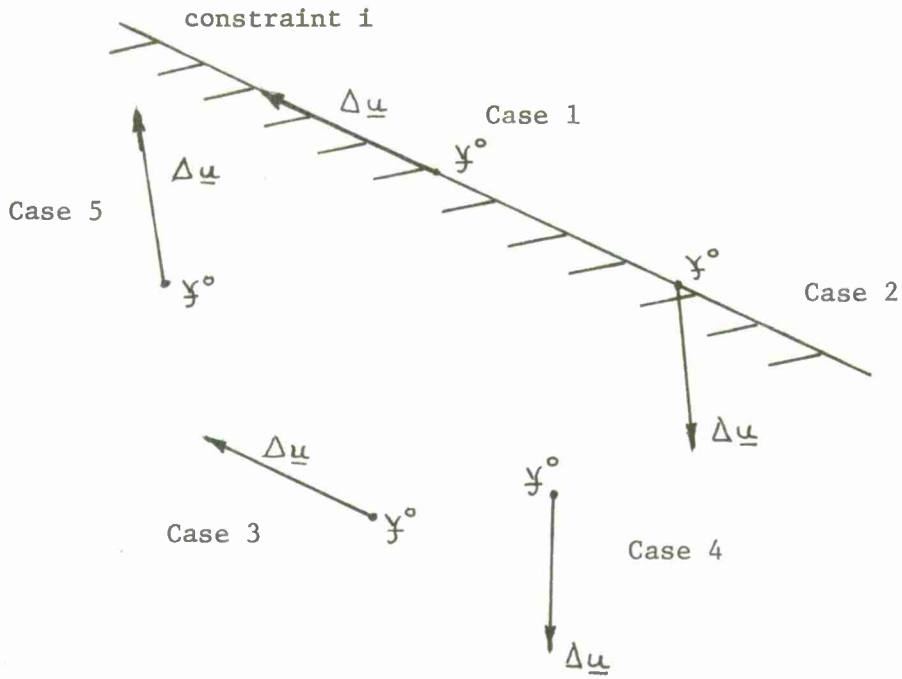
Then

$$k \underline{a}_i \Delta \underline{u} \leq r_i - \underline{a}_i \underline{y}^{\circ} \ .$$

Since $\underline{y}^{\circ}$ is feasible $r_i - \underline{a}_i \underline{y}^{\circ} \geq 0$. Thus we have five cases to consider for determining the restrictions on k :

1	$r_i - \underline{a}_i \underline{y}^{\circ} = 0, \quad \underline{a}_i \Delta \underline{u} = 0$	k unrestricted
2	$r_i - \underline{a}_i \underline{y}^{\circ} = 0, \quad \underline{a}_i \Delta \underline{u} < 0$	k unrestricted
3	$r_i - \underline{a}_i \underline{y}^{\circ} > 0, \quad \underline{a}_i \Delta \underline{u} = 0$	k unrestricted
4	$r_i - \underline{a}_i \underline{y}^{\circ} > 0, \quad \underline{a}_i \Delta \underline{u} < 0$	k unrestricted
5	$r_i - \underline{a}_i \underline{y}^{\circ} > 0, \quad \underline{a}_i \Delta \underline{u} > 0$	$k \leq r_i - \underline{a}_i \underline{y}^{\circ} / \underline{a}_i \Delta \underline{u}$

These cases can be represented schematically as :



In case 5 let*

$$\hat{k} = \min_i \frac{r_i - \underline{a}_i \underline{y}^\circ}{\underline{a}_i \Delta \underline{u}} .$$

Thus the restriction $k \leq \hat{k}$ ensures us that the point $\underline{y} = \underline{y}^\circ + k \Delta \underline{u}$ remains feasible. We actually determine k by a search along $\Delta \underline{u}$. The search algorithm given here is also used for the acceleration step in the cg method.

---

*The implicit constraints $-\underline{e}_i \underline{y} \leq 0$ must be considered in calculating $\hat{k}$.

Search Algorithm --

<u>Starting routine</u>: Set $\beta = \hat{k}$, $n = 1$ and calculate

$$\underline{z}^1 = \underline{y}^0 + \frac{1}{2}\beta\,\Delta\underline{u}\,, \quad \text{and} \quad \delta = f(\underline{z}^1) - f(\underline{y}^0)\,.$$

If $\delta < 0$ set $n = 2$ and go to the general routine .

If $\delta \geq 0$ set $\beta = \| \underline{z}^1 - \underline{y}^0 \|$.

If $\beta \geq \epsilon$ set $n = 1$ and continue .

If $\beta < \epsilon$ stop $k = 0$.

<u>General routine</u>:

Caluclate $\underline{z}^n = \underline{y}^0 + \left(1 - \frac{1}{2^n}\right)\beta\,\Delta\underline{u}$, and $\delta = f(\underline{z}^n) - f(\underline{z}^{n-1})$.

If $\delta < 0$ set $n = n+1$ and continue .

For $\delta \geq 0$:

If $\beta - \| \underline{z}^n - \underline{y}^0 \| \geq \epsilon$ set $n = 1$, $\beta = \| \underline{z}^n - \underline{y}^0 \|$

and continue.

If $\beta - \| \underline{z}^n - \underline{y}^0 \| < \epsilon$ and $f(\underline{z}^{n-1}) < f(\underline{y}^0)$

then $\underline{y} = \underline{z}^{n-1}$.

V. CONVERGENCE OF THE ALGORITHM

When $\alpha_B = 0$ and none of the bounded variable constraints hold exactly the LP problem (4) is equivalent to the primal problem

subject to:

$$\sum_{j=1}^{n} a_{ij} \Delta y_j = r_i - \underline{a}_i \underline{y}^\circ \qquad i = 1, \ldots m_1$$

$$\sum_{j=1}^{n} a_{ij} \Delta y_j \leq r_i - \underline{a}_i \underline{y}^\circ \qquad i = m_1+1, \ldots m$$

$$\Delta y_j \geq -y_j^\circ \qquad j = n_1+1, \ldots n$$

minimize:

$$\nabla f(\underline{y}^\circ) \Delta \underline{y} \; .$$

The dual of this problem is

subject to:

$$\sum_{i=1}^{m} a_{ij} x_i = \frac{\partial f(\underline{y}^\circ)}{\partial y_j} \qquad j = 1 \ldots n_1$$

$$\sum_{i=1}^{m} a_{ij} x_i - z_j = \frac{\partial f(\underline{y}^\circ)}{\partial y_j} \qquad j = n_1+1, \ldots n$$

$$x_i \leq 0 \qquad i = m_1+1, \ldots m$$

$$z_j \leq 0 \qquad j = n_1+1, \ldots n$$

maximize:

$$\sum_{i=1}^{m} (r_i - \underline{a}_i \underline{y}^\circ) x_i + \sum_{j=n_1+1}^{n} y_j^\circ z_j$$

From (2) and the dual theorem of linear programming, i.e.,

$$\nabla f(\underline{y}^\circ) \Delta \underline{y} = \sum_{i=1}^{m} (r_i - \underline{a}_i \underline{y}^\circ) x_i + \sum_{j=n_1+1}^{n} y_j^\circ z_j$$

we have

$$f(\underline{y}) = f(\underline{y}^{\circ}) + \sum_{i=1}^{m} (r_i - \underline{a}_i \underline{y}^{\circ}) x_i + \sum_{j=n_1+1}^{n} y_j^{\circ} z_j + r(\underline{y}^{\circ}, \Delta \underline{y}). \qquad (11)$$

Now

$$\sum_{i=1}^{m} (r_i - \underline{a}_i \underline{y}^{\circ}) x_i + \sum_{j=n_1+1}^{n} y_j^{\circ} z_j$$

since

$$x_i \leq 0 \ , \qquad i = m_1+1, \ldots m$$

and

$$\sum_{j=1}^{n} a_{ij} \Delta y_j = r_i - \underline{a}_i \underline{y}^{\circ} = 0 \ , \quad i = 1, \ldots m_1 \ .$$

Global or local convexity of $f(\underline{y})$ implies that $r(\underline{y}^{\circ}, \Delta \underline{y}) \geq 0$ in (11). Thus the condition for a constrained stationary point and termination of the algorithm is

$$\sum_{i=1}^{m} (r_i - \underline{a}_i \underline{y}^{\circ}) x_i + \sum_{j=n_1+1}^{n} y_j^{\circ} z_j = \nabla f(\underline{y}^{\circ}) \Delta \underline{y} \geq -\epsilon \qquad (12)$$

Note that this condition was also required by the Kuhn-Tucker conditions. This leads to the following

Theorem 4: (Convergence Theorem)* If $\alpha_B = 0$ then a stationary point is located in a finite number of steps.

---

*This theorem is a special case of a more general convergence theorem due to Graves [6].

Proof: If (12) is not satisfied, then by the Approximation Theorem (Buck [2] page 243)

$$f(\underline{y}^{\circ}+k\,\Delta\underline{u}) = f(\underline{y}^{\circ}) + \nabla f(\underline{y}^{\circ})\,k\,\Delta\underline{u} + r(\underline{y}^{\circ}, k\,\Delta\underline{u}) \,, \tag{13}$$

where

$$\lim_{k\Delta\underline{u}\to 0} \frac{r(\underline{y}^{\circ}, k\,\Delta\underline{u})}{k} = 0 \,, \quad \text{uniformly} \,.$$

Thus for any $\underline{y}^{\circ}$ there exists a $\delta > 0$ such that for $k\|\Delta\underline{u}\| = k \leq \delta$,

$$r(\underline{y}^{\circ}, k\,\Delta\underline{u}) < \frac{k\cdot\epsilon}{2\|\Delta\underline{y}\|} \leq \frac{\epsilon\cdot\delta}{2\|\Delta\underline{y}\|} \,.$$

From (12) and (13)

$$f(\underline{y}) = f(\underline{y}^{\circ}) + \frac{k}{\|\Delta\underline{y}\|}\nabla f(\underline{y}^{\circ})\,\Delta\underline{y} + r(\underline{y}^{\circ}, k\,\Delta\underline{u})$$

$$< f(\underline{y}^{\circ}) - \frac{\epsilon\cdot\delta}{\|\Delta\underline{y}\|} + \frac{\epsilon\cdot\delta}{2\|\Delta\underline{y}\|} = f(\underline{y}^{\circ}) - \frac{\epsilon\cdot\delta}{2\|\Delta\underline{y}\|} \,.$$

So if (12) is not satisfied then we can reduce $f(\underline{y})$ by at least $\epsilon\cdot\delta/2\|\Delta\underline{y}\|$ at each step until (12) is satisfied. Thus a stationary point is located in a finite number of steps.

VI. DISCUSSION

It has been known for some time that if $\underline{y}^{\circ}$ (an estimate of the constrained minimum) is not close to $\underline{y}^{*}$ then no local method can ever be devised which obtains a good direction from $\underline{y}^{\circ}$ to $\underline{y}^{*}$. In light of this fact it seems advisable to give up the search for better local directions and adopt a radically different criterion, such as that embraced in the primal-dual method. In the initial steps we focus on reducing the distance from the starting point to the constrained minimum. The direction from $\underline{y}^{\circ}$ to the unconstrained minimum serves as the surrogate direction. Upon reaching the boundary we hope to be close enough to the constrained minimum so that by solving (4) with $\alpha_B \to 0$ we obtain $\underline{y}^{*}$. Clearly problems can be constructed where this approach does not give a point on the boundary close to $\underline{y}^{*}$. But it is not apriori evident that we have moved further "out of the way" than we might by some other method. Although we may reach the boundary at a point quite close to $\underline{y}^{*}$, zig-zagging can still arise due to the eccentricity of the approximating ellipsoids of $f(\underline{y})$. Here it is evident that second order information on the curvature of $f(\underline{y})$ is absolutely necessary in order to proceed efficiently to $\underline{y}^{*}$. The method presented herein solves the local conjugate gradient problem for the complete set of conjugate directions and projects the "best" direction onto the subspace determined by the binding hyperplanes. Since the calculation of the projection matrix requires considerable matrix operations we take as many projection steps as possible, i.e., until the $\underline{B}'$ matrix must be changed.

The burden of calculation in the primal-dual method falls principally upon the LP algorithm and the conjugate gradient algorithm. The LP algorithm solves (4) for the feasible direction and inverts the $\underline{B}'\underline{B}$ matrix. The conjugate gradient algorithm solves the unconstrained problem and obtains the complete set of local conjugate directions for a projection step. Neither the LP algorithm of Graves [5] nor the conjugate gradient algorithm of Hestenes [10] are well known or widely used. However, it would appear that both these methods are the most efficient of their class. Thus it is hoped that the primal-dual method presented herein will prove to be the most effective general algorithm for non-linear programming problems with linear constraints. It is "effectively" a second-order method without requiring second partial derivatives. In addition it combines the desirable features of projection methods, conjugate gradient methods and methods that solve LP problems to obtain feasible directions.

In comparing the primal-dual method with other methods we will consider (1) a special case (that of linear constraints) of Graves' general nonlinear algorithm, (2) Goldfarb's conjugate gradient method, and (3) Rosen's gradient projection method for linear constraints. The algorithm of Graves is a first order method when only the first partial derivatives are used. The approach of Graves [6] is to attempt to solve all problems in this manner. When slow convergence is observed in the first order algorithm the second order conditions are appended as explicit constraints provided, of course, the second partial derivatives can be obtained. This approach expands the size of the problem and requires the second partial derivatives.

The conjugate gradient method of Goldfarb is effective on problems with eccentricity in the approximating ellipsoids of $f(\underset{\sim}{y})$. An additional advantage, as indicated by Goldfarb [3], is that a minimum of function evaluations are required. This is important in some problems in optimal control. However, many problems are difficult to solve just due to size alone (many constraints and/or many variables) even without additional complications. Since Goldfarb's method requires even more matrix operations than that of Rosen it may be quite slow due to the considerable amount of matrix operations on large matrices. In addition there is a large class of problems where $f(\underset{\sim}{y})$ simply decreases (unconstrained) to $-\infty$,i.e., there is no finite unconstrained minimum. When the conjugate gradient method is applied to a function of this class it can become unstable quite rapidly. Since Goldfarb's method requires the conjugate directions it may also become unstable for these functions. In the primal-dual method if instability is observed when solving for the unconstrained minimum we set $\alpha_B = 0$ and $\alpha_I = 0$ permanently. We then solve (4) to reach the constrained minimum $\underset{\sim}{y}^*$. If slow convergence is observed and $f(\underset{\sim}{y})$ is convex a direction can be obtained by solving the local conjugate gradient problem. Experience has shown that this procedure works quite well.

The gradient projection method of Rosen [12] is not as efficient as Goldfarb's method (see the numerical results in [3]). In addition it is probably not as efficient as Graves' algorithm due to the amount of matrix operations required.

VII. NUMERICAL RESULTS

The task of comparing nonlinear programming algorithms is at best a difficult job. For a variety of test problems it would seem that total computation time provides a universal norm for comparison. However, the complexities of modern computer operating systems do not allow the same test problem to be run twice with the same running time. Thus computation time is a random variable and the probability density function is unknown. Comparing computation times reported on different computers is even more difficult. For example, the test problems solved by Goldfarb [3] in 1968 were run on a CDC 6600--the CDC 6600 being approximately ten times faster (depending on the operating system and other factors) than the IBM System 360 Model 65 used here. The test problems solved by Graves [6] in 1966 were run on an IBM 7044/7094 direct couple system. He reports that the time sharing feature of that system could increase total time by as much as a factor of four. His times include loading and system times as well as execution time.

Comparing algorithms on the basis of total iterations (steps) is equally unsatisfactory. For example the Goldfarb algorithm generally solves a problem in a small number of steps although each step requires considerable calculation. On the other hand the Graves algorithm can take far more steps in the same amount of time since each step requires fewer calculations. In the primal-dual algorithm the situation is much more complicated since a number of different kinds of steps are possible. In solving for the unconstrained minimum the conjugate gradient steps

give an approximate second order Newton step, i.e.,

$$\Delta \underline{y} = -H(\underline{y}^{\circ})^{-1} \nabla f(\underline{y}^{\circ})' .$$

In solving for the constrained minimum we have ,

(1) LP steps:

$$\Delta \underline{y} = \underline{y} - \underline{y}^{\circ} ,$$

(2) interior steps:

$$\Delta \underline{y} = -(1-\alpha_I) \nabla f(\underline{y}^{\circ})' + \alpha_I (\underline{x}^{*} - \underline{y}^{\circ}) \text{ , and}$$

(3) projection steps:

$$\Delta \underline{y}^{P} = \underline{P}\, \underline{d} .$$

Obviously interior steps require the least amount of calculation and projection steps the most. The art of constructing an efficient algorithm is in providing the proper balance among the different kinds of steps.

The algorithms of Graves, Goldfarb, and Rosen can be classified in terms of the type of step used. For example, the Graves' algorithm takes LP steps with $\alpha_B = 0$ and interior steps with $\alpha_I = 0$. Rosen's algorithm takes projection steps and interior steps with $\alpha_I = 0$. In the projection step Rosen projects the gradient, which is always the first conjugate direction calculated. Recall that in the primal-dual method the complete set of conjugate directions are calculated and the "best" one projected. Goldfarb's algorithm takes projection steps and cg steps. In the projection step Goldfarb apparently projects the first conjugate direction which gives a gain, i.e., $\nabla f(\underline{y}^{\circ}) \underline{P}\, \underline{p}_K < 0$. So it is clear that each of these algorithms can be obtained as a special case of the primal-dual algorithm. If we choose not to solve for the unconstrained minimum the primal-dual algorithm becomes a local method and we have an alternate version (primal-dual II). At a boundary point we set $\alpha_B = 0$ and take either LP steps or projection steps. From an interior point we take conjugate gradient steps.

For a wide variety of problems (both large and small) the two primal-dual methods are probably the same in overall effectiveness. Of course for any particular problem one method is more efficient than the other. These classifications can be conveniently summarized in the following table:

Method or Algorithm	Type of Step*	
	Boundary Point	Interior Point
Graves	1) $\alpha_B = 0$	2) $\alpha_I = 0$
Rosen	3) $\underline{d} = -\nabla f(\underline{y}^0)' = \underline{p}_1$	2) $\alpha_I = 0$
Goldfarb	3) $\underline{d} = \underline{p}_K$ where $\nabla f(\underline{y}^0)\underline{P}\,\underline{p}_K < 0$	4)
Primal-Dual	1) or 3) $\underline{d} = \hat{\underline{p}}_K$ (see page 14)	2)
Primal-Dual II	1) $\alpha_B = 0$ or 3) $\underline{d} = \hat{\underline{p}}_K$ (see page 14)	4)

*1) LP step, 2) interior step, 3) projection step, and 4) cg step.

The test problems given in the next section have been devised to include most of the undesirable phenomena that arise in constrained minimization with linear constraints. In solving these problems we will report total computation time and total iterations (steps). The eight test problems are:

(1) constrained ellipse with $e \to 1$,

(2) Rosenbrock's parabolic valley with a linear constraint,

(3) Powell's function of 4 variables with linear constraints,

(4) Fletcher & Powell's helical valley with a range restricted variable,

(5) Colville problem #1,

(6) Colville problem #2,

(7) Chemical equilibrium, and

(8) constrained hyperbola with a saddle point.

The solution of test problems (5) and (6) by the Graves algorithm as a first order method are reported in [6]. The solution of test problems (5), (6), and (7) by the Goldfarb algorithm are reported in [3].

VIII. TEST PROBLEMS

(1). Constrained ellipse with $e \to 1$.

subject to:

$$\begin{aligned} y_1 \quad\quad\quad &\leq 8 \\ -y_1 \quad\quad\quad &\leq -3 \\ y_2 &\leq 10 \\ y_1, y_2 &\geq 0 \end{aligned}$$

minimize:

$$f(\underset{\sim}{y}) = \frac{(y_1 - 10)^2}{1000} + (y_2 - 5)^2$$

start at (1,1).

The eccentricity, e, of an ellipse is always between 0 and 1. As $e \to 0$ the ellipse approaches a circle and as $e \to 1$ the ellipse approaches a straight line. In this case the value of e is 0.9995. This problem is designed to create zig-zagging in ordinary first-order gradient methods.

(2). Rosenbrock's parabolic valley with a linear constraint.

subject to:

$$y_1 + y_2 \leq 1.9$$

minimize:

$$f(\underset{\sim}{y}) = 100(y_2 - y_1^2)^2 + (1 - y_1)^2$$

start at (-1,2,1).

The gradient and the Hessian are

$$\nabla f(\underline{y}) = (-400(y_2 - y_1^2)\,y_1 - 2(1-y_1),\ 200(y_2 - y_1^2)) ,$$

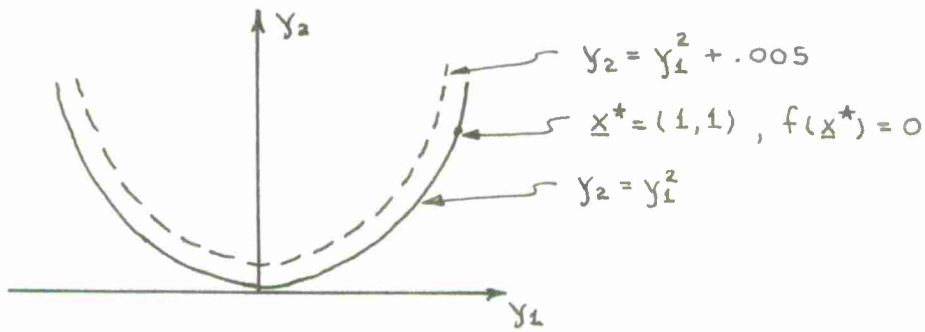
$$H(\underline{y}) = 200 \begin{bmatrix} 2(3y_1^2 - y_2 + .005) & -2y_1 \\ -2y_1 & 1 \end{bmatrix} .$$

Now $H(\underline{y})$ is singular when column 1 is a linear combination of column 2, i.e.,

$$2(3y_1^2 - y_2 + .005) = 4y_1^2 ,$$

or

$$y_2 - y_1^2 = .005 .$$



Along the curve $y_2 - y_1^2 = .005$ the eigenvalues of $H(\underline{y})$ are given by $(H(\underline{y}) - \lambda \underline{I})\underline{y} = \underline{0}$ or

$$\begin{vmatrix} 4y_1^2 - \lambda & -2y_1 \\ -2y_1 & 1-\lambda \end{vmatrix} = 0 ,$$

giving

$$(4y_1^2 - \lambda)(1-\lambda) - 4y_1^2 = 0 ,$$

and $\lambda_1 = 0, \lambda_2 = 1 + y_1^2 > 0$. Since $\lambda_1 = 0$ the approximating ellipsoids along the curve $y_1^2 - y_2 = -.005$ degenerate to parallel straight lines. This problem is designed to foul a second order method which uses the inverse Hessian and to create zig-zagging in a first order gradient method.

(3). Powell's function of 4 variables with linear constraints subject to:

$$-y_1 - y_2 - y_3 - y_4 \leq -1$$

$$y_i \leq 20 \qquad i = 1, \ldots 4$$

minimize:

$$f(\underline{y}) = (y_1 + 10y_2)^2 + 5(y_3 - y_4)^2 + (y_2 - 2y_3)^4 + 10(y_1 - y_4)^4$$

start at (3,-1,0,1) .

The gradient and the Hessian are

$$\nabla f(\underline{y}) = (2(y_1 + 10y_2) + 40(y_1 - y_4)^3,\ 20(y_1 + 10y_2) + 4(y_2 - 2y_3)^3,$$

$$10(y_3 - y_4) - 8(y_2 - 2y_3)^3,\ -10(y_3 - y_4) - 40(y_1 - y_4)^3),$$

$$h_{11} = 2 + 120(y_1 - y_4)^2, \qquad h_{12} = h_{21} = 20,$$

$$h_{22} = 200 + 12(y_2 - 2y_3)^2, \qquad h_{13} = h_{31} = 0,$$

$$h_{33} = 10 + 48(y_2 - 2y_3)^2, \qquad h_{14} = h_{41} = -120(y_1 - y_4)^2,$$

$$h_{44} = 10 + 120(y_1 - y_4)^2, \qquad h_{23} = h_{32} = -24(y_2 - 2y_3)^2,$$

$$h_{24} = h_{42} = 0,$$

$$h_{34} = h_{43} = -10.$$

The unconstrained minimum of Powell's function is obviously zero at (0,0,0,0) and at this point

$$H(\underline{0}) = \begin{bmatrix} 2 & 20 & 0 & 0 \\ 20 & 200 & 0 & 0 \\ 0 & 0 & 10 & -10 \\ 0 & 0 & -10 & 10 \end{bmatrix} .$$

Thus $H(\underline{0})$ has rank two and is singular. The unconstrained problem is designed to create slow convergence and this effect carries over to the constrained problem.

(4). Fletcher and Powell's helical valley with a range restricted variable.

subject to:

$$0.2 \leq y_1 \leq 0.8$$

minimize:

$$f(\underline{y}) = 100[(y_3 - 10\theta)^2 + (r-1)^2] + y_3^2 ,$$

where $y_1 = r\cos 2\pi\theta$ and $y_2 = r\sin 2\pi\theta$,

start at (-1,0,0) .

Since $y_1 = r\cos 2\pi\theta$ and $y_2 = r\sin 2\pi\theta$ we have on squaring and adding

$$r = (y_1^2 + y_2^2)^{1/2} .$$

Also ,

$$\theta = \frac{1}{2\pi} \tan^{-1}(u) \qquad \text{where } u = y_2/y_1 .$$

Since

$$\frac{\partial f}{\partial y_1} = \frac{\partial f}{\partial \theta}\frac{d\theta}{du}\frac{\partial u}{\partial y_1} + \frac{\partial f}{\partial r}\frac{\partial r}{\partial y_1} ,$$

we have

$$\frac{\partial f}{\partial y_1} = \frac{1000(y_3 - 10\theta)y_2}{\pi\left[1 + \left(\frac{y_2}{y_1}\right)^2\right]y_1^2} + \frac{200(r-1)y_1}{r} .$$

Similarly ,

$$\frac{\partial f}{\partial y_2} = \frac{-1000(y_3 - 10\theta)}{\pi\left[1 + \left(\frac{y_2}{y_1}\right)^2\right]y_1} + \frac{200(r-1)y_2}{r} ,$$

and

$$\frac{\partial f}{\partial y_3} = 200(y_3 - 10\theta) + 2y_3 .$$

Since $u = y_2/y_1$, appropriate restrictions on θ are required so that $y_1 \neq 0$. If $r \neq 0$ then $y_1 = 0$ when $\theta = -\frac{1}{4}, \frac{1}{4}, \frac{3}{4}$ so that we require

$$\frac{1}{4} < \theta < \frac{3}{4} , \qquad \text{or} \qquad \frac{\pi}{2} < 2\pi\theta < \frac{3\pi}{2} ,$$

and

$$-\frac{1}{4} < \theta < \frac{1}{4} , \qquad \text{or} \qquad -\frac{\pi}{2} < 2\pi\theta < \frac{\pi}{2} .$$

(5). Colville problem #1.

subject to:

$$\sum_{j=1}^{5} a_{ij} y_j \geq b_i \qquad i = 1, \ldots 10$$

$$y_j \geq 0 \qquad j = 1, \ldots 5$$

minimize:

$$f(y) = \sum_{j=1}^{5} e_j y_j + \sum_{i=1}^{5} \sum_{j=1}^{5} c_{ij} y_i y_j + \sum_{j=1}^{5} d_j y_j^3$$

start at (0,0,0,0,1) .

The coefficients are given as

	$i \backslash j$	1	2	3	4	5
e_j		-15	-27	-36	-18	-12
c_{ij}	1	30	-20	-10	32	-10
	2	-20	39	- 6	-31	32
	3	-10	- 6	10	- 6	-10
	4	32	-31	- 6	39	-20
	5	-10	32	-10	-20	30
d_j		4	8	10	6	2

a_{ij} (i \ j)	1	2	3	4	5	b_i
1	-16.	2.	0.0	1.	0.0	-40.
2	0.0	-2.	0.0	0.4	2.	- 2.
3	- 3.5	0.0	2.	0.0	0.0	- 0.25
4	0.0	-2.	0.0	-4.	-1.	- 4.
5	0.0	-9.	-2.	1.	-2.8	- 4.
6	2.	0.0	-4.	0.0	0.0	- 1.
7	-1.	-1.	-1.	-1.	-1.	-40.
8	-1.	-2.	-3.	-2.	-1.	-60.
9	1.	2.	3.	4.	5.	5.
10	1.	1.	1.	1.	1.	1.

This problem is designed to create considerable zig-zagging on the boundary of the constraint set in ordinary first order gradient methods.

(6). Colville problem #2.

subject to:

$$-10 \le y_i \le 10 \qquad i = 1, \ldots 4$$

minimize:

$$f(\underline{y}) = 100(y_2 - y_1^2)^2 + (1 - y_1)^2 + 90(y_4 - y_3^2)^2 + (1 - y_3)^2$$

$$+ 10.1\left[(y_2 - 1)^2 + (y_4 - 1)^2\right] + 19.8(y_2 - 1)(y_4 - 1)$$

start at (-3,-1,-3,-1) .

This problem is an example of the case where the unconstrained minimum is an interior point and hence is the optimal solution. When started from the given starting point the solution path leads through (the four-space equivalent of) a narrow, steep-sided, curving valley. In addition, a true stationary point lies in the path. The combination of these effects is to create zig-zagging near the stationary point which leads to instability.

(7). Chemical equilibrium.

subject to:

$$y_1 + 2y_2 + 2y_3 + y_6 + y_{10} = 2$$

$$y_4 + 2y_5 + y_6 + y_7 = 1$$

$$y_3 + y_7 + y_8 + 2y_9 + y_{10} = 1$$

$$y_i \geq 0 \qquad i = 1, \ldots 10$$

minimize:

$$f(\underline{y}) = \sum_{i=1}^{10} c_i y_i + \sum_{i=1}^{10} \left[y_i \log\left(\frac{y_i}{\sum_j y_j} \right) \right]$$

start at (0.1, 0.35, 0.5, 0.1, 0.35, 0.1, 0.1, 0.1, 0.1, 0.1) .

The coefficients C_i are :

	C_i		
1	- 6.089	6	-14.986
2	-17.164	7	-24.100
3	-34.054	8	-10.708
4	- 5.914	9	-26.662
5	-24.721	10	-22.179

The gradient is easily calculated to be

$$\frac{\partial f}{\partial y_i} = c_i + \log\left(\frac{y_i}{\sum_j y_j}\right) \qquad i = 1, \ldots 10 \quad .$$

A detailed discussion of the physical aspects of the chemical equilibrium problem is given in [1] pp. 46-49. Since log 0 is undefined a numerical problem arises in the primal-dual algorithm. The constraints $y_i \geq 0 \quad i = 1, \ldots 10$ are in reality $y_i > 0 \quad i = 1 \ldots 10$ but LP algorithms cannot handle the strict inequality. To circumvent this problem we add the auxiliary constraints

$$-y_i \leq -\epsilon \qquad i = 1, \ldots 10 \quad .$$

This problem is an example of the case where there is no finite unconstrained minimum yet $f(\underline{y})$ is convex.

(8). Constrained hyperbola with a saddle point.

subject to:

$$3y_1 + 2y_2 \leq 9$$

$$0.5y_1 + y_2 \leq 4$$

$$y_1, y_2 \geq 0$$

minimize:

$$f(\underline{y}) = -y_1 y_2$$

start at (1,1)

This problem was designed to trap the primal-dual method on the saddle point at (0,0). It is not difficult to show that the constrained minimum occurs at $(3/2, 9/4)$ and $f(3/2, 9/4) = -27/8$. Now the contours of $f(\underline{y})$ are hyperbolas and the conjugate gradient methods finds the center of hyperboloids as well as ellipsoids. The obvious solution is to add the constraint

$$-y_1 - y_2 \leq -\epsilon \ .$$

IX. SOLUTION OF THE TEST PROBLEMS

(1). Constrained ellipse with $e \to 1$.

	Steps	Type	Running Time (sec)
	2	cg steps	0.2
	5	LP steps	
	1	interior steps	0.5
	0	projection steps	—
Total	8		0.7

	Value
unconstrained minimum (U.M.) (10, 5)	0.0
constrained minimum (C.M.) (8, 5)	.004

Lagrange multipliers

1	-0.4×10^{-2}
2	0
3	0

Comments: The conjugate gradient method solved the unconstrained problem in 1 step. The 2nd step was for verification. Three of the five LP steps were required to drive $\alpha_B \to 0$. Generally, the number of steps required to drive $\alpha_B \to 0$ is given by min(m,n). Each test problem is considered to be solved when $\nabla f(\underline{y}^0)\Delta\underline{y} \geq -1 \times 10^{-6}$.

(2). Rosenbrock's parabolic valley with a linear constraint.

Steps	Type	Running Time (sec)
20	cg steps	1.7
5	LP steps	
1	interior steps	0.5
0	projection steps	—
Total 26		2.2

		Value
U.M.	(1, 1)	0.0
C.M.	(.96632704, .93367296)	$.1135190 \times 10^{-2}$

Lagrange multipliers

$1 \quad -0.2299606 \times 10^{-1}$

Comments: The conjugate gradient method has a tendency at times to take too large a step and become unstable. To counteract this we have introduced an under-relaxation technique [8]. Of the 20 steps required to solve the unconstrained problem, 4 included under-relaxation.

(3). Powell's function of 4 variables with linear constraints.

	Steps	Type	Running Time (sec)
	6	cg steps	
	3	acceleration steps	1.1
	6	LP steps	
	0	interior steps	10.3
	31	projection steps	
Total	46		11.4

		Value
U.M.	(0, 0, 0, 0)	0.0
C.M.	(.50332384, -.045560064, .23582560, .30641062)	0.11378385

Lagrange multipliers

1	- .4010324
2	0.
3	0.
4	0.
5	0.

Comments: Since the projection steps were consecutive only one projection matrix was calculated, thereby making these steps quite efficient.

(4). Fletcher & Powell's helical valley with a range restricted variable.

	Steps	Type	Running Time (sec)
	16	cg steps	
	3	acceleration steps	2.4
	5	LP steps	
	1	interior steps	0.3
	0	projection steps	—
Total	25		2.7

		Value
U.M.	(1, 0, 0)	0.0
C.M.	(0.8, 0, 0)	4.0

Lagrange multipliers

1	-40.
2	0.

(5) Colville problem #1.

Steps	Type	Running Time (sec)
6	cg steps	1.2
8	LP steps	
0	interior steps	2.6
2	projection steps	
Total 16		3.8

		Value
U.M.	(.5242668, .8826745, 1.258478, .7411507, .3355687)	-61.44833
C.M.	(0.3, 0.3334676, 0.4, .42831015, .22396490)	-32.348679

Lagrange multipliers

1	0.	6	-11.839545
2	0.	7	0.
3	-5.1740399	8	0.
4	0.	9	- .10389623
5	-3.0611088	10	0.

Comments: This problem was solved by both Graves and Goldfarb. Graves' algorithm, as a first order method, did not do well because of zig-zagging. It would appear that the primal-dual method solved this problem roughly twice as fast as Goldfarb's algorithm.

(6). Colville problem #2.

Steps	Type	Running Time (sec)
Total 41	cg steps	6.6

		Value
U.M. (C.M.)	(1, 1, 1, 1)	0.0

Comments: This problem was also solved by both Graves and Goldfarb. Graves' algorithm, as a first order method, was trapped at the non-optimal stationary point. Goldfarb's algorithm reached the global minimum in 105 steps. Although it was considerably slow in the vicinity of the stationary point it remained stable. The primal-dual algorithm became unstable at the 14th step, in the vicinity of the stationary point, However, this step resulted in movement to a point from which the global minimum was reached in 27 steps.

The results of this test problem should not be taken too seriously. The conjugate gradient method has the desirable property of moving away from true stationary points, but if forced to calculate in the immediate vicinity, it tends to become unstable.

(7). Chemical equilibrium.

	Steps	Type	Running Time (sec)
	< 1	cg step	0.3
	11	LP steps	
	0	interior steps	57.4
	54	projection steps	____
Total	65		57.7

No finite unconstrained minimum

	C.M.		Best Known Solution [13]
1	.040587	1	.040668
2	.146323	2	.147730
3	.785045	3	.783153
4	.001409	4	.001414
5	.485317	5	.485247
6	.000692	6	.000693
7	.027265	7	.027399
8	.017867	8	.017947
9	.036919	9	.037314
10	.095984	10	.096872
Value	-47.761079		-47.761106

Lagrange multipliers

1	- 9.789526
2	-12.97198
3	-15.22707

Comments: This is the only test problem of the eight that provided any particular difficulty for the primal-dual method. The difficulty probably arises from the fact that although $f(\underline{y})$ is convex, it decreases (unconstrained) to $-\infty$. The best known solution could have been achieved if a smaller zero level had been chosen. It is interesting that Goldfarb's algorithm solved this problem with relative ease (about 8 times faster than the primal-dual algorithm). It is likely that the Goldfarb algorithm generates different conjugate directions than that given here.*

*However, theoretical work for a quadratic function, indicates that, Hestenes and Fletcher-Powell both generate the same set of conjugate directions.

(8). Constrained Hyperbola with a Saddle Point.

	Steps	Type	Running Time (sec)
	1	cg step	0.1
	4	LP steps	
	0	interior steps	0.1
	0	projection steps	—
Total	5		0.2

	Value
saddle point $(-.470924 \times 10^{-11}, -.470924 \times 10^{-11})$	0.
constrained saddle point (0, 0)	0.

In terms of some of the standards described by Ignizio,* some additional information on the numerical part of this study is warranted; namely:

Date of computation--July 1972 to October 1972

Computer utilized--IBM System 360/Model 65

Programming language--FORTRAN IV

Total number of problems attempted--In addition to the test problems reported, a number of other small problems were solved. Since they exhibited no interesting mathematical properties and were solved quite rapidly they were not reported. Each test problem is considered to be solved when $\nabla f(\underset{\sim}{y}^{o}) \Delta \underset{\sim}{y} \geq -1 \times 10^{-6}$.

Computation time--The running times were calculated by the FORTRAN external function ITIME. The timer was set up to be the first and last executable statements in the program. The times reported are CPU times only.

---

*Ignizio, J., "On the Establishment of Standards for Comparing Algorithm Performance," TIMS Interfaces, 2(1), Nov. 1971, pp. 8-11.

REFERENCES

[1] Bracken, J. and McCormick, G., Selected Applications of Nonlinear Programming (John Wiley & Sons, New York, 1968).

[2] Buck, R., Advanced Calculus (McGraw Hill, New York, 2nd Ed., 1965).

[3] Goldfarb, D. and Lapidus, L., "Conjugate Gradient Method for Nonlinear Programming Problems with Linear Constraints," I&EC Fundamentals, Vol. 7, 142-151 (1968).

[4] Goldfarb, D., "Extension of Davidon's Variable Metric Method to Maximization Under Linear Inequality and Equality Constraints," SIAM J. Appl. Math., Vol. 17, No. 4 (July 1969).

[5] Graves, G., "A Complete Constructive Algorithm for the General Mixed Linear Programming Problem," Naval Res. Logistics Quart., Vol. 12, No. 1, 1-34 (March 1965).

[6] Graves, G. and Whinston, A., "The Application of a Nonlinear Programming Algorithm to a Second Order Representation of the Problem," Cahier de Centre d'Etudes de Recherche Operationelle Bruxelles, Belgique, Vol. 11, No. 2 (1969).

[7] Hadley, G., Nonlinear and Dynamic Programming (Addison-Wesley, Reading, Massachusetts, 1964).

[8] Hatfield, G., "Unconstrained Minimization of a Real-Valued Function," unpublished report, UCLA, Nov. 1971.

[9] Hestenes, M. and Stiefel, E., "Methods of Conjugate Gradients for Solving Linear Systems," Journal of Research of the National Bureau of Standards, Vol. 49, No. 6 (1952).

[10] Hestenes, M., "Multiplier and Gradient Methods," Journal of Optimization Theory and Applications, Vol. 4, No. 5, 303-320 (Nov. 1969).

[11] Noble, B., Applied Linear Algebra (Prentice-Hall, 1969).

[12] Rosen, J., "The Gradient Projection Method for Nonlinear Programming, Part 1. Linear Constraints," SIAM J. Appl. Math., Vol. 8, 181-217 (1960).

[13] White, W., Johnson, S., and Dantzig, G., "Chemical Equilibrium in Complex Mixtures," J. Chem. Phys., Vol. 28, 751-755 (1958).

DISTRIBUTION LIST

Chief of Naval Operations (OP-01)
(OP-098T)
(OP-099)
(OP-39)
(OP-964)
Commandant of the Marine Corps, G-1 Division, Code A01M
Chief of Naval Training (Code N-33)
Chief of Naval Material (MAT 03P)
Chief of Naval Personnel (Pers-11b)
(Pers-A12) (5)
(Pers-A3) (25)
Commander Antisubmarine Warfare Force, U. S. Atlantic Fleet (Code 70)
Office of Naval Research (Code 458) (2)
Naval Postgraduate School
Naval Examining Center
Center for Naval Analyses
Naval Personnel Research and Development Laboratory
Office of the Assistant Secretary of the Navy
(Manpower and Reserve Affairs)
Chief of Research and Development, U. S. Army
Army Research Institute for Behavioral and Social Sciences
Personnel Research Division, Lackland AFB
Personnel Management Development Office, OPO, U. S. Army
Human Factors Operations Research Laboratory, Bolling AFB
Science and Technology Division, Library of Congress
Defense Documentation Center (Attn: DDC-TC) (12)

U152077