AN ALGORITHM FOR QUADRATIC PROGRAMMING

Marguerite Frank and Philip Wolfe[1]
Princeton University

A finite iteration method for calculating the solution of quadratic programming problems is described. Extensions to more general non-linear problems are suggested.

1. INTRODUCTION

The problem of maximizing a concave quadratic function whose variables are subject to linear inequality constraints has been the subject of several recent studies, from both the computational side and the theoretical (see Bibliography). Our aim here has been to develop a method for solving this non-linear programming problem which should be particularly well adapted to high-speed machine computation.

The quadratic programming problem as such, called PI, is set forth in Section 2.

We find in Section 3 that with the aid of generalized Lagrange multipliers the solutions of PI can be exhibited in a simple way as parts of the solutions of a new quadratic programming problem, called PII, which embraces the multipliers. The maximum sought in PII is known to be zero. A test for the existence of solutions to PI arises from the fact that the boundedness of its objective function is equivalent to the feasibility of the (linear) constraints of PII.

In Section 4 we apply to PII an iterative process in which the principal computation is the simplex method change-of-basis. One step of our "gradient and interpolation" method, given an initial feasible point, selects by the simplex routine a secondary basic feasible point whose projection along the gradient of the objective function at the initial point is sufficiently large. The point at which the objective is maximized for the segment joining the initial and secondary points is then chosen as the initial point for the next step.

The values of the objective function on the initial points thus obtained converge to zero; but a remarkable feature of the quadratic problem is that in some step a secondary point which is a solution of the problem will be found, insuring the termination of the process.

A simplex technique machine program requires little alteration for the employment of this method. Limited experience suggests that solving a quadratic program in n variables and m constraints will take about as long as solving a linear program having m + n constraints and a "reasonable" number of variables.

Section 5 discusses, for completeness, some other computational proposals making use of generalized Lagrange multipliers.

Section 6 carries over the applicable part of the method, the gradient-and-interpolation routine, to the maximization of an arbitrary concave function under linear constraints (with one qualification). Convergence to the maximum is obtained as above, but termination of the process in an exact solution is not, although an estimate of error is readily found.

In Section 7 (the Appendix) are accumulated some facts about linear programs and concave functions which are used throughout the paper.

[1]Under contract with the Office of Naval Research.

2. THE QUADRATIC PROBLEM

Our primary concern is with the problem of maximizing the concave quadratic function

$$f(x) = \sum_{j=1}^{n} p_j x_j - \sum_{j,k=1}^{n,n} x_j C_{jk} x_k \tag{2.1}$$

subject to the linear constraints

$$\begin{aligned} & x_j \geq 0 && (j = 1, \ldots, n), \\ & \sum_{j=1}^{n} A_{ij} x_j \leq b_i && (i = 1, \ldots, m). \end{aligned} \tag{2.2}$$

Matrix notation will be used exclusively below. x is the $n \times 1$ matrix (i.e., column vector) of variables $x_1, \ldots, x_n$. A, C, p, and b are, respectively, $m \times n$, $n \times n$, $1 \times n$, and $m \times 1$ matrices. $A_{i.} = (A_{i1}, \ldots, A_{in})$ will denote the i^{th} row of A, and likewise $A_{.j}$ the j^{th} column. e_j is the j^{th} coordinate (column) vector of n dimensions. The symbol $'$ denotes matrix transposition. For any function $f(x)$, the gradient $1 \times n$ matrix $\left(\frac{\partial f}{\partial x_1}, \ldots\right)$ is denoted by ∂f. C may without loss of generality be supposed symmetric.

In matrix terms, we may rewrite the quadratic problem as

<u>PI</u>: Maximize $f(x) = px - x'Cx$ subject to

$$\text{(I)} \quad \begin{cases} x \geq 0 \\ Ax \leq b. \end{cases}$$

An x satisfying (I) will be called <u>feasible</u>. The set of all feasible x is the <u>constraint</u> set. The problem PI is called <u>feasible</u> if the constraint set is not empty, that is, if there exists a feasible x. If the objective function has a supremum on the constraint set, this supremum will be denoted by M. An x for which $f(x) = M$ is called a <u>solution</u> of PI and the set of all solutions will be called the <u>solution set</u>. It will be assumed in this and the following section that PI is feasible, but not that solutions necessarily exist.

The postulated concavity of the objective function is equivalent to the requirement that C be the matrix of a positive semidefinite quadratic form [App. d]. A function is said to be concave if interpolation never overestimates its values: that is, if

$$f(\alpha x + \beta y) \geq \alpha f(x) + \beta f(y) \quad \text{whenever} \quad \alpha, \beta \geq 0, \; \alpha + \beta = 1. \tag{2.3}$$

It follows that any local maximum point x of f on a convex constraint set is also a global maximum: for otherwise a y such that $f(y) > f(x)$ could be found, and then any point on the segment joining x to y, no matter how close to x, would yield a higher value for f than would x. Similarly for a <u>convex</u> function, any local <u>minimum</u> on a convex set is a global minimum. It is this fact which makes maximizing a concave function (or minimizing a convex function) one of the simpler of the non-linear programming problems.

Concavity of the objective function implies that a "mixture" of two equally good sets of activities can never result in a poorer set of activities. Conversely, the worst possible program is always "extreme" in that, roughly speaking, it involves as little mixing as is possible. In certain cases it may happen that an "extreme" program is also the best program; but this may be considered as not a typical case, although not a pathological case either. It appears that non-linear programming problems in which (2.3) does not hold — e.g., contract award programs if discounts are given for large orders — will require substantially different techniques than those presented here.

3. LAGRANGE MULTIPLIERS

Kuhn and Tucker's generalization [6] of the method of Lagrange multipliers to the maximization of a function f whose variables $x_1, \ldots, x_n$ lie in the constraint set given by the inequalities

$$f_i(x) \geq 0 \quad (i = 1, \ldots, m) \tag{3.1}$$

is based on the observation that f has a local extreme value at x if, and only if, the normal hyperplane to the gradient vector $\partial f(x)$ at x is, locally, a supporting hyperplane of the constraint set. (This formulation applies equally well to the classical case $f_i(x) \equiv 0$.) If x is to be a local maximum, then the constraint set and the gradient drawn from x must lie on opposite sides of the supporting hyperplane. If, furthermore, the constraint set is convex (as when the functions f_i are convex or, in particular, linear), and f is concave, then, as seen in Section 2, this condition is necessary and sufficient that f have a global maximum at x.

Noting that a hyperplane touching the convex constraint set at a point x satisfies the above condition if, and only if, x has the maximum projection (of all points of the convex) along the outward normal to the hyperplane, a necessary and sufficient condition that x_0 be a solution of PI is that

$$\partial f(x_0)x_0 = \text{Max}\,[\partial f(x_0)w \mid w \geq 0,\ Aw \leq b], \quad (w \text{ an } n \times 1 \text{ matrix}). \tag{3.2}$$

But by the Duality Theorem for linear programming [App. a] (x_0 is fixed), the right side of (3.2) is equal to

$$\text{Min}\,[ub \mid u \geq 0,\ uA \geq \partial f(x_0)] \quad (u \text{ a } 1 \times m \text{ matrix}). \tag{3.3}$$

Substituting this for the right side of (3.2) and transposing, we obtain that:

$$\begin{gathered} x_0 \text{ is a solution of PI if, and only if,} \\ \text{Max}\,[\partial f(x_0)x_0 - ub \mid u \geq 0,\ uA \geq \partial f(x_0)] = 0\,. \end{gathered} \tag{3.4}$$

Note that $\partial f(x) = p - 2x'C$. The function

$$g(x, u) \equiv \partial f(x)\,x - ub = px - ub - 2x'Cx \tag{3.5}$$

extracted from (3.4), which is similar to the generalized Lagrangian of [6], plays a dominant role in the sequel. It is evidently a concave function of the variables (x, u). We suppose henceforth that (x, u) is constrained by the linear relations used in PI, (3.2), and (3.4), namely,

$$\left\{\begin{array}{ll} x \geq 0, & u \geq 0, \\ Ax \leq b, & \partial f(x) \leq uA\,. \end{array}\right. \tag{3.6}$$

Now under these constraints

$$g(x,u) \leq 0, \tag{3.7}$$

since $g(x,u) = \partial f(x)\, x - ub \leq uAx - ub = u(Ax - b) \leq 0$. But then the Max of (3.4) is never positive, and we have proved the result on which our use of the generalized Lagrange multiplier method rests, the

SOLUTION CRITERION:[2] x is a solution of PI if, and only if, for some u such that (x, u) satisfies (3.6),

$$g(x,u) = 0\,. \tag{3.8}$$

More information connecting f and g is given by the following easy consequence of the concavity of f [App. e]:

For any x and y,

$$f(y) - f(x) \leq \partial f(x)\,(y - x)\,. \tag{3.9}$$

If now (x, u) satisfies (3.6), we have that for any feasible y,

$$\begin{aligned} f(y) - f(x) &\leq \partial f(x)\, y - \partial f(x)\, x \leq uAy - \partial f(x)\, x \\ &\leq ub - \partial f(x)\, x \qquad = -\, g(x,u)\,. \end{aligned} \tag{3.10}$$

Evidently if f is not bounded above for all feasible y, then such (x, u) cannot be found. On the other hand, [App. i] if f is so bounded then there exists a solution, x_0, and thus, by the Solution Criterion, (x_0, u) satisfying (3.6) exists. We have thus proved the

BOUNDEDNESS CRITERION: f is bounded above on the constraint set if, and only if, the joint constraints (3.6) are feasible.

Finally, supposing in (3.10) that y is a solution of PI—that is, that $f(y) = M$, we have the

APPROXIMATION CRITERION: If (x, u) satisfies (3.6), then

$$M - f(x) \leq -\, g(x,u)\,. \tag{3.11}$$

THE PROBLEM PII: From the above it appears that a solution x of PI, if there is any, may be read off from any (x, u) for which the maximum, zero, of the concave quadratic

---

[2] The Solution Criterion was originally obtained from Lemma 1 of [6]: our $g(x,u)$ is $\phi_x^{o'}\, x^o + \phi_u^{o'}\, u^o$ in Kuhn and Tucker's notation, so that (3.8) is equivalent to part of the conditions of their Lemma 1. (3.6) is equivalent to the remaining conditions, and Theorems 1 and 2 of [6] then establish the Solution Criterion. Barankin and Dorfman [1] obtain the Criterion and the transformed problem PII in this way.

function $g(x,u)$ is achieved under the constraints (3.6). It is, in fact, this latter problem, PII below, which is solved in Section 4. Also, the Approximation Criterion (3.11) yields an estimate $f(x) \leq M \leq f(x) - g(x,u)$ for M when any (x,u) satisfying (3.11) is at hand.

As is customary in solving linear programs [3, p. 339] the constraints (3.6) will be written as equations by adjoining one non-negative variable y_i $(i = 1, \ldots, m)$ to each of the m inequalities of $Ax \leq b$, and one v_j $(j = 1, \ldots, n)$ to each of the n inequalities of $\partial f(x) \leq uA$ (written for our purposes as $-\partial f(x) + uA \geq 0$, or, using matrix transposition, as $2Cx + A'u' \geq p'$), obtaining

$$\text{(3.12)} \qquad \begin{aligned} &x \geq 0, \quad u \geq 0, \quad y \geq 0, \quad v \geq 0, \\ &Ax + y = b, \\ &2Cx + A'u' - v' = p' \end{aligned}$$

(y an $m \times 1$ matrix, v an $n \times 1$ matrix).

We also have, after substitution from (3.12),

$$\text{(3.13)} \qquad g(x,u) = \partial f(x)\, x - ub = (uA - v)\,x - u(Ax + y) = -(vx + uy)\,.$$

In these terms, PII consists in finding vectors x, u, y, v satisfying (3.12) such that

$$\text{(3.14)} \qquad vx + uy = 0\,.$$

This last condition may be interpreted geometrically by noting that it is equivalent to

$$\text{(3.15)} \qquad v_j \neq 0 \text{ implies } x_j = 0, \quad u_i \neq 0 \text{ implies } y_i = 0 \quad (\text{all } i, j)\,.$$

Since $\partial f(x) = uA - v = \Sigma\, v_j\,(-e_j) + \Sigma\, u_i\, A_{i.}$ is an expression for $f(x)$ as a non-negative linear combination of the outward normals to those bounding hyperplanes $e_j x = 0$, $A_{i.}\, x = b_i$ on which x lies (i.e., for which x_j, y_i vanish), the solution criterion (3.14) is equivalent to:

(3.16) x solves PI if, and only if, $\partial f(x)$ belongs to the convex cone spanned by the outward normals to the bounding hyperplanes on which x lies.

This condition is a geometric dual of the condition stated at the beginning of this section, and has been obtained from it by the duality theorem for linear programming.

BASIC SOLUTIONS: The system (3.12) consists of $m+n$ equations in $2(m+n)$ non-negative variables. It is easily shown that any extreme point of the set of points x, u, y, v feasible for the constraints (3.12) has at most $m+n$ positive components. An extreme point of this convex will be called a basic feasible vector.[3] Evidently the solution of a linear program can always be found as a basic feasible vector and the simplex process does precisely so [App. b].

[3]After Dantzig. Under the non-degeneracy hypothesis of [3], a basic feasible vector has exactly $m+n$ positive components.

Now the conditions (3.15) entail that at most $m+n$ components of any solution of PII are positive, so that we may hope to find a solution of PII among the basic feasible vectors of the constraint set for PII. We show in Section 4 that this is indeed the case.[4]

This remarkable consequence of the use of Lagrange multipliers might suggest the direct use of the simplex method to maximize the objective function; however, as discussed in Section 5, this is not possible. Nevertheless, the simplex method machinery is usefully exploited in the "gradient-and-interpolation" method presented in the next section.

4. THE COMPUTATION

PRELIMINARIES: For present purposes the four interdependent vectors x, u, y, v may be combined into a single constrained vector z. Accordingly let

$$z = \begin{bmatrix} x \\ u' \\ y \\ v' \end{bmatrix}, \quad B = \begin{bmatrix} A & O & I & O \\ 2C & A' & O & -I \end{bmatrix}, \quad d = \begin{bmatrix} b \\ p' \end{bmatrix}, \tag{4.1}$$

the two instances of I being identity matrices of appropriate rank. z, B, and d are then $2(m+n)\times 1$, $(m+n)\times 2(m+n)$, and $(m+n)\times 1$ matrices, respectively; and the system of constraints (3.12) for PII assumes the form.

$$z \geq O, \quad Bz = d . \tag{II}$$

With each z as above associate the "adjoint"

$$\tilde{z} = [v, y', u, x'] . \tag{4.2}$$

The new objective function may be conveniently expressed using this linear operation, for by (3.13)

$$g(x,u) = -(vx+uy) = -\frac{1}{2}\tilde{z}z . \tag{4.3}$$

Further, for $z = [x', u, y', v]'$ and $Z = [X', U, Y', V]'$ feasible with respect to (II) above,

$$\begin{cases} \tilde{z}Z = \tilde{Z}z , \text{ and} \\ (\tilde{Z} - \tilde{z})(Z - z) \geq 0 . \end{cases} \tag{4.4}$$

The first relation is evident. The second relation is given by

$$\begin{aligned} \frac{1}{2}(\tilde{Z}-\tilde{z})(Z-z) &= (V-v)(X-x) + (U-u)(Y-y) \\ &= [2(X-x)'C + (U-u)A](X-x) + (U-u)A(x-X) \\ &= 2(X-x)'C(X-x) \geq O , \end{aligned}$$

which thus reflects the concavity of g.

[4] This has been shown independently by Barankin and Dorfman when the system (3.12) is non-degenerate [1].

In these terms, the final version of the Lagrange multiplier problem obtained in the last section may be stated as

PII: Find z for which the maximum, zero, of the inner product

$$-\tilde{z}\,z$$

is assumed under the constraints

$$z \geq 0\,, \quad Bz = d\,. \tag{II}$$

By the Boundedness Criterion, PI has a (finite) solution if, and only if, PII is feasible; by the Solution Criterion, x solves PI if, and only if, $z = [x' \ u \ y' \ v]'$ solves PII for some y, u, and v.

Although this has not been necessary before, for application of the simplex method we suppose that the constraint equations (II) have, where necessary, been multiplied by -1, so that the right side is non-negative.

THE ALGORITHM: The process below will yield a basic feasible vector solution of PII. Phase I initiates the process, while Phase II is iterated until it yields the required vector.

Phase I: The constraints (II) are tested for feasibility. (Most commonly employed is the artificial basis [App. c].) If they are feasible, a basic feasible vector z_1 is produced with which to begin Phase II. If not, the last n equations may be discarded and the remainder similarly tested for feasibility. If these are feasible, then the quadratic problem is feasible, but unbounded above; otherwise, the quadratic problem is infeasible.

Phase II: This phase is defined inductively: At the k^{th} instance of Phase II we have the feasible (II) vector w_k which does not solve PII, and also a basic feasible vector z_k with which to start the simplex machinery. (At the first instance, let $w_1 = z_1$.)

Employ the simplex technique in the maximization of the linear form

$$-\tilde{w}_k\,z\,,$$

obtaining the sequence of basic feasible vectors $z^1 = z_k$, z^2, z^3, .. such that

$$-w_k z^1 < -w_k z^2 < \ldots\,[5].$$

Stop at the first z^h such that either

$$\tilde{z}^h z^h = 0\,, \quad \text{or} \tag{4.5}$$

$$-\tilde{w}_k z^h \geq -\frac{1}{2}\tilde{w}_k w_k\,. \tag{4.6}$$

---

[5]In the likely event that the constraints (II) are degenerate, "$\leq$" may occur here for a while, but not for long. Dantzig's method for handling degeneracy [2, 4] is exceptionally easy to use here, owing to the presence (except for sign) of an identity matrix in the constraint matrix.

If (4.5) obtains, then z^h solves the problem. Otherwise let

$$(4.7) \qquad \begin{cases} z_{k+1} = z^h \\ \mu = \operatorname{Min}\left\{\dfrac{\tilde{w}_k(w_k - z_{k+1})}{(\tilde{z}_{k+1} - \tilde{w}_k)(z_{k+1} - w_k)}, 1\right\}, \text{ and} \\ w_{k+1} = w_k + \mu(z_{k+1} - w_k). \end{cases}$$

Repeat Phase II, using w_{k+1} and z_{k+1}.

PROOF OF CONVERGENCE: It will be shown that the new objective $-\tilde{w}_k w_k$ converges monotonically to zero, and that in some instance of Phase II a basic feasible vector solution satisfying (4.5) is found.

(a) In order to show that z^h satisfying (4.6) will be found in a given instance of Phase II (if (4.5) is not obtained), it is sufficient to observe that the linear programming problem Max $\{-\tilde{w}_k z \mid (\text{II})\}$ has a finite maximum, since $-\tilde{w}_k z \leq 0$. This maximum is not less than $-\frac{1}{2}\tilde{w}_k w_k$, since if Z solves PII, i.e., $\tilde{Z}Z = 0$, we have $\tilde{w}_k w_k - 2\tilde{w}_k Z = (\tilde{w}_k - \tilde{Z})(w_k - Z) \geq 0$, by (4.4).

For the time being, let $w_k = w$, $z_{k+1} = Z$, $w_{k+1} = W$. Then (4.6) yields $\tilde{w}(w - Z) \geq \frac{1}{2}\tilde{w}w > 0$, and (4.4) that $(\tilde{Z} - \tilde{w})(Z - w) \geq 0$, so that $0 < \mu \leq 1$ (no difficulty arises if the denominator in the expression for μ vanishes). Thus w_{k+1} is a convex combination of w_k and z_{k+1}; indeed, it has been chosen so as to <u>maximize the new objective on the segment</u> $\overline{w_k z_{k+1}}$.

(b) Convergence. Dropping subscripts as in (a), when $Z = z^h$ satisfying (4.6) is obtained in the k^{th} instance of Phase II, we have

$$\begin{aligned} \tilde{W}W &= \tilde{w}w + 2\mu\tilde{w}(Z - w) + \mu^2(\tilde{Z} - \tilde{w})(Z - w) \\ &= \tilde{w}w + \mu\tilde{w}(Z - w) + \mu[\mu(\tilde{Z} - \tilde{w})(Z - w) - \tilde{w}(w - Z)] \\ &\leq \tilde{w}w + \mu\tilde{w}(Z - w) \qquad (4.8) \\ &\leq \tilde{w}w - \mu \cdot \frac{1}{2}\tilde{w}w \\ &= (1 - \tfrac{1}{2}\mu)\tilde{w}w. \end{aligned}$$

Letting F be the (compact) convex hull of the set of all basic feasible vectors of (II), let

$$(4.9) \qquad L = \operatorname{Max}\left\{(\tilde{z}^1 - \tilde{z}^2)(z^1 - z^2) \mid z^1, z^2 \text{ in } F\right\}.$$

If $\mu < 1$, then

$$\frac{1}{2}\mu \geq \frac{\tilde{w}(w - Z)}{2L} \geq \frac{\tilde{w}w}{4L} \quad \text{by (4.6)},$$

so that, in any case,

$$1 - \frac{\mu}{2} \leq \operatorname{Max}\left\{1 - \frac{\tilde{w}w}{4L}, \frac{1}{2}\right\},$$

and finally by (4.8)

$$\frac{\tilde{W}W}{4L} \leq \frac{\tilde{w}w}{4L} \operatorname{Max}\left\{1 - \frac{\tilde{w}w}{4L}, \frac{1}{2}\right\}. \tag{4.10}$$

Resuming subscripts, let $a_k = \dfrac{\tilde{w}_k w_k}{4L}$. Then (4.10) becomes

$$a_{k+1} \leq \operatorname{Max}\left\{1 - a_k, \frac{1}{2}\right\} a_k \quad \text{for } k \geq 0. \tag{4.11}$$

Since if $a_k \geq \frac{1}{2}$ then $a_{k+1} \leq \frac{1}{2} a_k$, there exists K such that $a_k < \frac{1}{2}$ whenever $k \geq K$. Now for $a_k < \frac{1}{2}$ we have $a_{k+1} \leq a_k(1 - a_k)$, whence

$$\frac{1}{a_{k+1}} \geq \frac{1}{a_k} \frac{1}{1 - a_k} = \frac{1 + a_k + a_k^2 + \ldots}{a_k} \geq \frac{1}{a_k} + 1,$$

so that

$$\text{for } k \geq K, \quad \frac{1}{a_k} \geq \frac{1}{a_K} + (k - K).$$

Thus

$$a_k \leq \begin{cases} \left(\frac{1}{2}\right)^k a_o & \text{for } k \leq K \\ \dfrac{1}{k - K + \dfrac{1}{a_K}} & \text{for } k > K. \end{cases} \tag{4.12}$$

Hence $\tilde{w}_k w_k$ approaches zero at worst like $1/k$, and the convergence is proved.

(c) Finiteness. Now suppose that the linear program of Phase II never yields a basic feasible vector which solves PII; then in particular no z_k obtained in the iteration of Phase II solves PII. Since each w_k is a convex combination of $\{z_{k'} \mid k' < k\}$, there is a point of accumulation w_o of $\{w_k\}$, belonging to the convex hull of the finite set of basic feasible vectors $\{z_k \mid \tilde{z}_k z_k > 0\}$ for which $\tilde{w}_o w_o = 0$; yet $w_o = \Sigma t_k z_k$ with $t_k \geq 0$, $\Sigma t_k = 1$, so that $\tilde{w}_o w_o = \sum_{j,k} t_j t_k \tilde{z}_j z_k > 0$: a contradiction. The proof that Phase II terminates in a solution is complete.

(d) Approximation. Some iterations of Phase II can be avoided if only a solution of PI to a predetermined degree of approximation is desired; for the Approximation Criterion (3.11) yields that, where x_k is the x-part of w_k,

$$M - f(x_k) \leq \frac{1}{2} \tilde{w}_k w_k ; \tag{4.13}$$

Phase II is repeated until $\tilde{w}_k w_k$ is sufficiently small.

Computation Time: Experience with a few very small numerical examples (of the order of $n = 5$, $m = 3$) suggests that the total number of changes of basis (several in each instance of the application of the Simplex Technique to Phase II) required to solve the quadratic problem will, in practice, be of the order of $2(m + n)$. The instances of Phase II after the first commonly require only one or two changes of basis. Thus this algorithm may be able to solve quadratic problems as quickly as linear problems of comparable size can be solved; but we have no theoretical estimate for this.

5. OBITER DICTA

We briefly consider here some alternative computational proposals which aim directly to exploit the fact that some basic feasible vector of (II) is a solution of the problem PII by using the simplex change of basis to pass from a basic feasible vector to adjacent basic feasible vectors.

Perhaps the most attractive possibility is that of the existence of a function defined for basic feasible vectors which assumes its minimum at a solution and which can always be decreased with a single change of basis if the solution has not been reached. In view of the Solution Criterion (3.8), the function $-g(x, u) = \frac{1}{2} \tilde{z} z$ naturally merits consideration from this angle, as well as the integer-valued function which counts the number of positive components of $\tilde{z}$ that coincide in position with the positive components of z; these functions have been studied by Barankin and Dorfman [1]. Quadratic programming problems may be constructed, however, in which certain non-solving basic feasible vectors occur as "local minima" for these functions, which cannot be decreased with a single change of basis. Whether or not a function of the desired type exists is not known.

One could also systematically explore all the basic feasible vectors of (II) for a solution, changing basis to pass from vertex to adjacent vertex in the graph consisting of all the basic feasible vectors and their connecting edges in (II). Such a process has been proposed by Charnes [2] for obtaining all solutions of a linear programming problem, given one. The amount of information that must be recorded if a graph of unknown design is to be traced out has been studied as the "labyrinth problem" [5]. The tracing process becomes less efficient (in terms of edges retraced) the less the data recorded. A minimum record is a list of all vertices already traversed. The best available upper bound [7] for the number of vertices of (II) being of the order of $\binom{2m+2n}{m+n-2}$, it seems that such an approach to PII would be infeasible for large-scale problems because of the excessive demands made on the memory of a high-speed computer.

6. CONCAVE PROGRAMMING

The gradient-and-interpolation method of Section 4 applied to the quadratic concave function $g(x, u)$ may be equally well applied to the problem of maximizing a more general concave function f, satisfying the hypothesis (A) below, whose variables are constrained by linear relations. Generalized Lagrange multipliers do not seem to have the same utility here that they do in the quadratic case, since the constraint $\partial f(x) \leq uA$ entailed by their use is no longer linear, and thus the features which ensured the termination of the quadratic programming in a finite number of steps are not found here.

Let f be a concave function of n variables possessing continuous second derivatives. The concave programming problem is:

<u>PC</u>: Maximize $f(x)$ under the constraints

$$x \geq 0 , \quad Ax \leq b . \tag{6.1}$$

Here A, b are $m\times n$, $m\times 1$ matrices, $\partial f(x)$ is the $1\times n$ matrix $\left(\frac{\partial f(x)}{\partial x_i}\right)$, $\partial^2 f(x)$ is the $n\times n$ matrix $\left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)$.

The problem will be supposed to satisfy the hypothesis

(A): For each feasible x, $\{\partial f(x)\, y \mid y \text{ feasible}\}$ is bounded above.

In particular, (A) holds if the constraint set is bounded, and in any case implies that f is bounded above on the constraint set, since for x fixed we have [App. e] the consequence of concavity

$$f(y) \leq f(x) + \partial f(x)\,(y - x) . \tag{6.2}$$

(A) is also easily shown for the problem PII.

Supposing now that the problem is feasible, let F be the (compact) convex hull of the set of basic feasible vectors of the constraint set. Such basic feasible vectors exist, since by (A) the linear program Max $\{\delta f(x)\, y \mid y \geq 0,\ Ay \leq b\}$ has solutions. There exists, then, a constant $L > 0$ such that for each x, y, and w in F,

$$-(y - x)'\, \partial^2 f(w)\, (y - x) \leq L . \tag{6.3}$$

The computational method follows.

<u>Phase One</u>: As in the quadratic case, determine the feasibility of the constraints (6.1), obtaining the initial basic feasible vector z_1.

<u>Phase Two</u>: At the k^{th} instance of Phase Two, when we suppose given the basic feasible vector z_k and the feasible vector x_k ($x_1 = z_1$), find a basic feasible vector z_{k+1} which solves the linear program

$$\text{Max } \{\partial f(x_k)\, w \mid w \geq 0,\ Aw \leq b\} . \tag{6.4}$$

Let s be such that

$$-(z_{k+1} - x_k)' \, \partial^2 f(w) \, (z_{k+1} - x_k) \le s \ge 0 \tag{6.5}$$

for all w on the segment $\overline{x_k z_{k+1}}$.[6] Let

$$\begin{cases} \mu = \operatorname{Min}\left\{ \dfrac{\partial f(x_k)(z_{k+1} - x_k)}{s}, \, 1 \right\}, \\ \\ x_{k+1} = x_k + \mu \, (z_{k+1} - x_k) , \end{cases} \tag{6.6}$$

and repeat Phase Two with z_{k+1}, x_{k+1}.

PROOFS: We shall prove that $f(x_k)$ converges to M, the supremum of f on the constraint set. Let for the moment $x_k = x$, $x_{k+1} = X$, and $z_{k+1} - x_k = y$.

From (6.2), for any feasible w we have $f(w) \le f(x) + \partial f(x)(w - x) \le f(x) + \partial f(x)(z_{k+1} - x) = f(x) + \partial f(x) \, y$, since z_{k+1} maximizes $\partial f(x_k)$; and thus

$$M - f(x) \le \partial f(x) \, y . \tag{6.7}$$

Now by Taylor's theorem, since $0 \le \mu \le 1$,

$$f(X) = f(x + \mu y) = f(x) + \mu \, \partial f(x) \, y + \frac{1}{2}\mu^2 \, y' \, \partial^2 f(w) \, y ,$$

where w lies on $\overline{xX}$, so that, from (6.7), (6.6), and (6.5),

$$\begin{aligned} M - f(X) &= M - f(x) - \frac{1}{2}\mu \, \partial f(x) \, y - \frac{1}{2}\mu \, [\partial f(x) \, y + \mu \, y' \, \partial^2 f(w) \, y] \\ &\le M - f(x) - \frac{1}{2}\mu \, [M - f(x)] - \frac{1}{2}[s\mu + \mu \, y' \, \partial^2 f(w) \, y] \\ &\le (1 - \frac{1}{2}\mu) \, [M - f(x)] . \end{aligned} \tag{6.8}$$

Since, moreover,

$$1 - \frac{1}{2}\mu = \operatorname{Max}\left\{ 1 - \frac{1}{2}, \, \frac{\partial f(x) \, y}{-y' \, \partial^2 f(w) \, y} \right\} \le \operatorname{Max}\left\{ \frac{1}{2}, \, \frac{M - f(x)}{L} \right\},$$

(6.8) yields, dividing by L, resuming subscripts,

[6] Of course $s \le L$. For this method, s should be chosen as small as possible, but we have no suggestions for finding it.

$$\frac{M - f(x_{k+1})}{L} \leq \frac{M - f(x_k)}{L} \operatorname{Max}\left\{1 - \frac{M - f(x_k)}{L}, \frac{1}{2}\right\}, \tag{6.9}$$

which is the same as (4.9) with $\frac{\widetilde{W}W}{4L}$ replaced by $\frac{M-f}{L}$, so that $\frac{M - f(x_k)}{L}$ converges to zero like $\frac{1}{k}$, and $f(x_k) \to M$.

Finally, we have in analogy to the Approximation Criterion of Section 3 the estimate for M given by (6.7):

$$f(x_k) \leq M \leq f(x_k) + \partial f(x_k)(z_{k+1} - x_k). \tag{6.10}$$

This estimate improves throughout the process; letting $z_{k+1} - x_k = y_k$, we show $\partial f(x_k) y_k \to 0$.

Let $\delta_k = \operatorname{Min}\{\partial f(x_k) y_k, L\}$. Then for some w between x and $x + \frac{\delta_k}{L} y$,

$$M \geq f(x + \frac{\delta_k}{L} y_k) = f(x_k) + \frac{\delta_k}{L} \partial f(x_k) y_k + \frac{1}{2} \frac{\delta_k^2}{L^2} y_k' \partial^2 f(w) y_k$$

$$\geq f(x_k) + \frac{\delta_k}{L} \partial f(x_k) y_k - \frac{\delta_k^2}{2L} \geq f(x_k) + \frac{\delta_k^2}{2L}.$$

Since $f(x_k) \to M$, we have $\delta_k \to 0$.

In the absence of the property (A), a modification of the gradient-and-interpolation procedure which takes account of possible infinite solutions of the linear program (6.4) can be made, for which convergence of the function to the maximum can be shown; however, estimates of the rate of convergence (6.9) and error (6.10) cannot be constructed.

7. APPENDIX

(a) The duality theorem for linear programming asserts, where A, b, c are, respectively, $m \times n$, $m \times 1$, $1 \times n$ matrices:

$$\sup\{cx \mid x \geq 0, Ax \leq b\} = \inf\{ub \mid u \geq 0, uA \geq c\},$$

the supremum (infimum) over an empty set being $-\infty$ (∞); the extrema are assumed if finite.

(b) The simplex method for solving the linear program $\operatorname{Max}\{cx \mid x \geq 0, Ax = b\}$ employs at each step the basis consisting of m columns of A such that their totality B is nonsingular and $B^{-1} b \geq 0$. The vector having components $B^{-1} \cdot b$ with appropriate zeroes adjoined is a basic feasible vector. A change of basis replaces one column of B by another column from A in such a way as to increase the objective cx on the basic feasible vector. This is continued until the maximum is attained. The constraints are non-degenerate if always $B^{-1} b > 0$. The non-degeneracy assumption of [3] has been removed [4].

(c) An initial basic feasible vector for the linear program is found by applying the simplex method to the expanded problem $\operatorname{Max}\{-y \mid x \geq 0, y \geq 0, Ax + Iy = b\}$, $x = 0$, $y = b \geq 0$

being an initial basic feasible vector for this problem. If Max < 0, then the original problem is not feasible; otherwise, for the Max, $y = 0$, so x is a basic feasible vector. This formulation is used with a further modification to eliminate degeneracy [2, 4].

(d) The function f (having continuous second derivative) on the convex set S is concave if, and only if, for all x, y in S and $0 \leq \mu \leq 1$, using $h(\mu) = f(x + \mu(y - x))$, we have $h''(\mu) \leq 0$, whence:

$$(y - x)' \, \partial^2 f(w) \, (y - x) \leq 0 , \tag{7.1}$$

for all w between x and y, is a necessary and sufficient condition for concavity. If S has interior points, (7.1) is equivalent to the condition that the matrix $\partial^2 f(w)$ be negative semi-definite.

If $f(x) = px - x'Cx$ then $\partial^2 f = -C$, so that f is concave for all x if, and only if, C is positive definite.

(e) For f as in (d), if $x, y \in S$, by Taylor's theorem

$$f(y) - f(x) = \partial f(x)(y - x) + \frac{1}{2}(y - x)' \, \partial^2 f(w)(y - x)$$

for some w between x and y, so that

$$f(y) - f(x) \leq \partial f(x)(y - x) \tag{7.2}$$

is a necessary condition for concavity of f [6, Lemma 3]. If the left of (7.1) is positive for some x, y, w, then $\bar{x}$, $\bar{y}$ can be chosen sufficiently close to w that $(\bar{y} - \bar{x})' \, \partial^2 f(\bar{w})(\bar{y} - \bar{x}) > 0$ for all $\bar{w}$ between $\bar{x}$ and $\bar{y}$, so that (7.2) fails, whereby it is also sufficient.

(f) Since, in general, if f is concave then $\{x \mid f(x) \geq c\}$ is convex for all c, the solution set $\{x \in S \mid f(x) \geq \text{Max}\}$ of any concave program on a convex set S is convex.

(g) If C is positive semi-definite and symmetric and $x'Cx = 0$ then for all μ and z, $0 \leq (x + \mu z)' C (x + \mu z) = 2\mu x'Cz + \mu^2 z'Cz$, so that $x'Cz = 0$ for all z, whence $x'C = Cx = 0$.

(h) If x and y belong to the solution set of the quadratic problem, then for all $0 \leq \mu \leq 1$,

$$M = f(x + \mu(y - x)) = f(x) + \mu \, \partial f(x)(y - x) - \frac{1}{2}\mu^2 (y - x)' C (y - x) ,$$

whence $(y - x)' C (x - y) = 0$, so $(y - x)' C = 0$, hence $f(x) = p - 2x'C = p - 2y'C = f(y)$. Thus, ∂f is constant on the solution set. Furthermore, $f(x + \mu(y - x)) = f(x)$ for all μ, so that the solution set is the intersection of the constraint set with some linear manifold.

(i) We prove by induction on k that if the quadratic function $f(x) = px + x'Cx$ (not necessarily concave) is bounded on the polyhedral convex set R of dimension k, then it assumes its supremum on R.

We may write any $k+1$-dimensional polyhedron R as $\{s + \mu t \mid s \in S, \, t \in T, \, \mu \geq 0\}$, where S is a certain bounded convex polyhedron and T is the intersection of a certain convex polyhedral cone with the unit sphere. Since for $r \in R$, $t \in T$, $f(r + \mu t) = f(x) + \mu(p + 2r'C)t + \mu^2 t'Ct$ is bounded for $\mu \geq 0$, $t'Ct \leq 0$ for all t in T.

If, on the one hand, $t'Ct < 0$ for all t in T, then there exist $\delta > 0$ and D such that $t'Ct < -\delta$ and $(p + 2s'C)t < D$ for all t in T and s in S, so that the maximum of f is assumed on the compact set $S + \frac{D}{\delta} T$.

Suppose, on the other hand, some $t_0' C t_0 = 0$. If for all $r \epsilon R$ we have $r + \mu t_0 \epsilon R$ for all μ, then the boundedness of $f(r + \mu t)$ implies $(p + 2r'C)t_0 = 0$ for all r, so that the values of f on R are unchanged by projection into the k-dimensional subspace normal to t_0, to which the induction hypothesis may be applied.

Otherwise, for each r, $r + \mu t_0 \notin R$ for some μ, so that for each r, $b_r = r + \mathrm{Min}\,\{\mu \mid r + \mu t_0 \epsilon R\} t_0$ lies on the boundary of R, and $f(b_r) \geq f(r)$, since $(p + 2r'C)t_0 \leq 0$. Since the supremum of f on each k-dimensional bounding hyperplane of R is assumed, so is it on R. (Irving Kaplansky has pointed out that the above result does not obtain for polynomials of degree greater than two. It is clearly not true for polynomials of odd degree; and the function $x^2 + (1 - xy)^2$ does not assume its greatest lower bound, zero, on the plane.)

REFERENCES

[1] Barankin, E. W., and Dorfman, R., Toward Quadratic Programming, O. N. R. Logistics Projects at Columbia Univ. and U. C. L. A., Univ. of California at Berkeley.

[2] Charnes, A., Cooper, W. W., and Henderson, A., An Introduction to Linear Programming, Wiley, 1953.

[3] Dantzig, G. B., "Maximization of a linear function of variables subject to linear inequalities," Activity Analysis of Production and Allocation (T. C. Koopmans ed.), Wiley, 1951.

[4] Dantzig, G. B., Orden, A., and Wolfe, P., "The generalized simplex method for minimizing a linear form under linear inequality restraints," Pacific J. Math. 5, 183-195 (June 1955).

[5] König, D., Theorie der Endlichen und Unendlichen Graphen, Leipzig, 1936.

[6] Kuhn, H. W., and Tucker, A. W., "Nonlinear programming," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 481-492 (1951).

[7] Saaty, T. L., "The number of vertices of a polyhedron," Amer. Math. Monthly 62, 326-331 (May 1955).

ADDITIONAL BIBLIOGRAPHY

(Items 4-7 discuss computational proposals.)

1. Debreu, G., "Definite and semidefinite quadratic forms," Econometrica 20, 295-300 (1952).
2. Mann, H. B., Quadratic forms with linear constraints," Amer. Math. Monthly, 58, 430-433 (1943).
3. Phipps, C. G., Maxima and minima under restraint," Amer. Math. Monthly 59, 230-235 (1952).
4. Arrow, K. J., and Hurwicz, L., "A gradient method for approximating saddle points and constrained maxima," RAND P-223 (June 1951). To appear in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.
5. Charnes, A., and Lemke, C. E., "Minimization of non-linear separable convex functionals (Computational theory of linear programming IV)," O.N.R. Research Memorandum No. 16, Carnegie Institute of Technology (May, 1954).

391368 O - 56 - 8

6. Manne, A. S., "Concave programming for gasoline blends," RAND P-383 (April 1953).
7. Markowitz, H., "The optimization of quadratic functions subject to linear constraints," Naval Research Logistics Quarterly, Vol. 3, Nos. 1, 2, 1956.

* * *