J. SIAM Control
Vol. 4, No. 1, 1966
Printed in U.S.A.

# AN ITERATIVE PROCEDURE FOR COMPUTING THE MINIMUM OF A QUADRATIC FORM ON A CONVEX SET*

ELMER G. GILBERT†

**1. Introduction.** This paper presents an iterative procedure for computing the minimum of a quadratic form on a compact convex set $C$. The sole characterization required of $C$ is the availability of a method for solving linear programs on $C$. This characterization differs from the usual set of functional inequalities given in quadratic programming problems [6], and is particularly appropriate to the solution of problems in optimal control. In fact, some of the results presented here arose from an attempt to provide a convergence proof for the extension by Fancher [5] of a procedure due to Ho [8]. Section 8 and [1] give several direct applications of the iterative procedure to problems in optimal control. By using the algorithm of this paper as a means of projecting points into convex sets it is possible to develop additional algorithms for solving other problems in programming and control [1], [7].

It should be noted that the iterative procedure of this paper is very similar to that given in the latter part of the paper by Frank and Wolfe [6]. However the emphasis and setting of the two papers are quite different, and the overlap is small.

The paper is organized as follows: in §2 notation, definitions, and a basic problem (BP) are considered; in §§3, 4, and 5 the algorithm for BP is described, error bounds are derived, and convergence is proved and investigated in detail; in §6 the algorithm is related to a gradient method for solving BP; in §7 the previous results are extended to a general quadratic programming problem GP; and in §8 the connection with problems in optimal control is made.

**2. Preliminaries, the basic problem.** The following notation is employed: $z = (z^1, \cdots, z^n)$, a vector in Euclidean $n$-space $E^n$; $y \cdot z = \sum_{i=1}^{n} y^i z^i$; $|z| = (z \cdot z)^{1/2}$; $N(x; \omega) = \{y \mid |y - x| < \omega\}$, $\omega > 0$, the open sphere at $x$ with radius $\omega$; $\bar{N}(x; \omega) = \{y \mid |y - x| \leqq \omega\}$, the corresponding closed sphere; $L(x; y) = \{z \mid z = x + \omega(y - x), -\infty < \omega < \infty\}$, $x \neq y$, the line passing through $x$ and $y$; $Q(x; y) = \{z \mid z \cdot y = x \cdot y\}$, $y \neq 0$,

* Received by the editors June 30, 1965. Presented at the First International Conference on Programming and Control, held at the United States Air Force Academy, Colorado, April 15, 1965.

† Information and Control Engineering, University of Michigan, Ann Arbor, Michigan. This research was supported by the United States Air Force under Grant No. AF-AFOSR-814-65.

the hyperplane (dimension $n - 1$) through $x$ with normal $y$; $\partial X$, the boundary of the set $X$.

Now consider some notation and results applicable to a set $K \subset E^n$, which is compact and convex. Let $\eta(y) = \max_{z \in K} z \cdot y$ denote the *support function* of $K$. Since $K$ is compact, $\eta(y)$ is defined for all $y$. Furthermore, it can be shown that $\eta(y)$ is a convex function on $E^n$, a result which implies that $\eta(y)$ is continuous on $E^n$ [3]. Let $P(y)$, $y \neq 0$, be the hyperplane $\{x \mid x \cdot y = \eta(y)\}$. Since $z \cdot y \leqq \eta(y)$ for all $z \in K$ and $P(y) \cap K$ is not empty, $P(y)$ is the (unique) *support hyperplane* of $K$ with outward normal $y$. For each $y \neq 0$ the set $S(y) = P(y) \cap K$ is called the *contact set* of $K$. It follows that $S(y)$ is not empty, $S(y) \subset \partial K$, $S(\omega y) = S(y)$ for $\omega > 0$. If for every $y \neq 0$, $S(y)$ contains only a single point, then $K$ is *strictly convex*.

DEFINITION. A function, $s(y)$, defined on $E^n$ is a *contact function* of $K$ if $s(y) \in S(y)$, $y \neq 0$, and $s(0) \in K$.

From the preceding it may be concluded that $s(\cdot)$ is bounded; $s(y) = s(\omega y)$, $\omega > 0$; and $\eta(y) = s(y) \cdot y$. Furthermore, on the set $\{y \mid \mid y \mid > 0\}$ each of the following is true if and only if $K$ is strictly convex: $s(\cdot)$ is uniquely determined, $s(\cdot)$ is continuous. The continuity result is proved in [10].

If for every $y$ there is a method for determining a point $x(y) \in K$ such that $x(y) \cdot y = \max_{z \in K} z \cdot y = \eta(y)$, then this method may be used to evaluate a contact function of $K$. Such an evaluation, which corresponds to the solution of a linear programming problem on $K$ (see §1), is essential to the computing procedure which follows. Consider now the basic problem:

BP. *Given $K$, a compact convex set in $E^n$, find a point $z^* \in K$ such that $\mid z^* \mid = \min_{z \in K} \mid z \mid$.*

Since $K$ is compact and $\mid z \mid$ is a continuous function of $z$, a solution $z^*$ exists. The following additional results hold:

*Solution properties.* (i) *$z^*$ is unique,* (ii) *$\mid z^* \mid = 0$ if and only if $0 \in K$,* (iii) *for $\mid z^* \mid > 0$, $z^* \in \partial K$,* (iv) *for $\mid z^* \mid > 0$, $z = z^*$ if and only if $z \in P(-z) \cap K = S(-z)$.*

Properties (ii) and (iii) are obvious. Property (i) is proved by contradiction. Suppose $z_1^*$ and $z_2^*$ are distinct solutions. Then by convexity $\tilde{z} = \frac{1}{2}z_1^* + \frac{1}{2}z_2^* \in K$, which means $\mid \tilde{z} \mid \geqq \mid z_1^* \mid = \mid z_2^* \mid$. But this implies

$$\mid \tfrac{1}{2}z_1^* + \tfrac{1}{2}z_2^* \mid^2 \geqq \tfrac{1}{2} \mid z_1^* \mid^2 + \tfrac{1}{2} \mid z_2^* \mid^2,$$

which can be written $\mid z_1^* - z_2^* \mid^2 \leqq 0$, an inequality which is only true for $z_1^* = z_2^*$. Consider (iv). The condition $z \in P(-z) \cap K$ implies $z \in P(-z) = Q(z; z)$. But $Q(z; z)$ is the support hyperplane for the closed sphere $\bar{N}(0; \mid z \mid)$ whose outward normal is $z$ and whose contact point is $z$. Therefore $Q(z; z)$ is a (separating) support hyperplane for $K$

and $\bar{N}(0; |z|)$. Thus $K \cap N(0; |z|)$ is empty. Since $z \in K \cap \bar{N}(0; |z|)$, this implies $z = z^*$. The steps of this argument may be reversed to obtain the converse result.

**3. The iterative procedure for the basic problem.** In this section the iterative procedure for computing the solution to BP is described.

As a first step, let $s(\cdot)$ be a specific contact function of $K$ and consider

$$(3.1)\quad \beta(z) = \begin{cases} |z - s(-z)|^{-2} z\cdot(z - s(-z)) & \text{if } z - s(-z) \neq 0, \\ 0 & \text{if } z - s(-z) = 0, \end{cases}$$

and

$$(3.2)\quad \gamma(z) = \begin{cases} |z|^{-2} z\cdot s(-z) & \text{if } |z| > 0,\ z\cdot s(-z) > 0, \\ 0 \quad \text{if } z = 0 & \text{or } |z| > 0,\ z\cdot s(-z) \leqq 0. \end{cases}$$

Thus $\beta(\cdot)$ and $\gamma(\cdot)$ are functions which are defined on $K$. Their geometric significance is as follows: $x = z + \beta(z)\ (s(-z) - z)$ is the point on the line $L(z; s(-z))$ with minimum Euclidean length; $\gamma(z)z$ is either the point $L(0; z) \cap P(-z)$ or the origin, depending on whether or not $L(0; z) \cap P(-z)$ is on the line segment connecting $z$ and the origin. The functions $\beta(\cdot)$ and $\gamma(\cdot)$ have the following properties.

THEOREM 1. *Let $K$ be the set described in BP and restrict $z$ to $K$. Then*

(i) $\beta(z) \geqq 0$,

(ii) $\beta(z) = 0$ *if and only if* $z = z^*$,

(iii) $0 \leqq \gamma(z) \leqq 1$,

(iv) *if* $0 \in K$, $\gamma(z) \equiv 0$,

(v) *if* $0 \notin K$, $\gamma(z) = 1$ *if and only if* $z = z^*$,

(vi) $\gamma(z)$ *is continuous.*

*Proof.* In this paragraph $z$ always denotes a point in $K$. In §4 (inequality (4.5)) it is shown that $0 \leqq z\cdot(z - s(-z))$. Hence, (i) and (iii) follow from (3.1) and (3.2). For the time being assume $|z^*| > 0$. The conditions $\beta(z) = 0$ and $\gamma(z) = 1$ both imply $z\cdot(-z) = s(-z)\cdot(-z) = \eta(-z)$ which requires $z \in P(-z)$. Since $z \in K$, solution property (iv) yields $z = z^*$. Reversing these arguments completes the proof of (ii) for $|z^*| > 0$ and of (v). Now take $|z^*| = 0$. Inequality (4.4) then implies $s(-z)\cdot z \leqq 0$ which by (3.2) yields (iv). If $\beta(z) = 0$, then it must follow from (3.1) that $s(-z)\cdot z = |z|^2$. Because of $s(-z)\cdot z \leqq 0$ this implies $z = 0 = z^*$. Since $z = z^* = 0$ also yields $\beta(z) = 0$, the proof of (ii) is complete. For $|z| \geqq |z^*| > 0$, the continuity of $\gamma(z)$ follows from (3.2) and the conti-

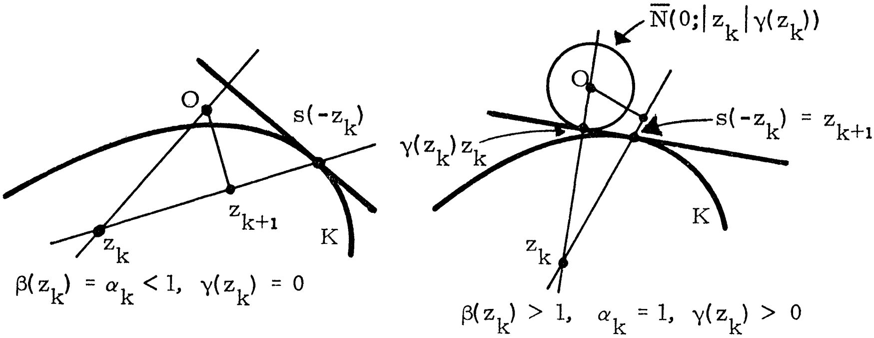


FIG. 1. *Geometric interpretation of the iterative procedure* (*O origin*)

nuity of the support function $\eta(y) = s(y) \cdot y$. For $|z^*| = 0$, it is trivially true from (iv).

It is of interest to note that $\beta(\cdot)$ may be discontinuous on $K$, even though $s(\cdot)$ is continuous on $K$. See Example 3, §5.

The iterative procedure defines a sequence of vectors $\{z_k\}$ by

$$z_{k+1} = z_k + \alpha_k(s(-z_k) - z_k), \qquad k = 0, 1, 2, \cdots, \tag{3.3}$$

where $z_0$ is an arbitrary point in $K$ and the scalars $\alpha_k$ are selected arbitrarily from the closed interval $I(z_k)$,

$$I(z) = [\min\{\delta\beta(z), 1\}, \min\{(2 - \delta)\beta(z), 1\}], \qquad 0 < \delta = \text{fixed number} \leqq 1. \tag{3.4}$$

Fig. 1 gives the geometric interpretation of the iterative procedure for the case where $\delta = 1$ and $\alpha_k \in I(z_k)$ reduces to $\alpha_k = \operatorname{sat} \beta(z_k)$ ($\operatorname{sat} \omega = \omega$, $0 \leqq \omega \leqq 1$; $\operatorname{sat} \omega = 1$, $\omega > 1$). If $\beta(z_k) > 0$ an improvement is obtained on the $k$th step, i.e., $|z_{k+1}| < |z_k|$; if $\beta(z_k) = 0$, $z_k = z^*$ and the iterative process is finite, i.e., the solution has been obtained in $k$ steps. From Fig. 1 it is also clear that $|z_k|\gamma(z_k) \leqq |z^*| \leqq |z_k|$. Thus on each step upper and lower bounds on $|z^*|$ may be computed. Notice that in applying the iterative procedure it is not necessary to know beforehand whether or not $0 \in K$. A more precise and complete statement of results is contained in the following theorem.

THEOREM 2. *Let* $s(\cdot)$ *be an arbitrary contact function of the set* $K$ *specified in BP. Take* $z_0 \in K$ *and, by means of* (3.3) *with* $\alpha_k \in I(z_k)$, *generate the sequence* $\{z_k\}$. *Then for* $k \geqq 0$ *and* $k \to \infty$:

(i) $z_k \in K$,

(ii) *the sequence* $\{|z_k|\}$ *is decreasing and* $|z_k| \to |z^*|$,

(iii) $z_k \to z^*$,

(iv) $|z_k|\gamma(z_k) \leqq |z^*|$ *and* $|z_k|\gamma(z_k) \to |z^*|$,

(v) $|z_k - z^*| \leqq \sqrt{1-\gamma(z_k)}\,|z_k|$ *and* $\sqrt{1-\gamma(z_k)}\,|z_k| \to 0$,

(vi) $|s(-z_k) - z^*| \leqq |s(-z_k) - \gamma(z_k)z_k|$.

Since the bounds given in parts (iv), (v), and (vi) are computable as the iterative process proceeds, they may be used to generate stopping criteria for the termination of the iterative process. Example problems show $\{|z_k|\gamma(z_k)\}$ is not necessarily increasing. Thus $|z_k| - \max_{i\leqq k}|z_i|\gamma(z_i)$ is more satisfactory as an upper bound for $|z_k| - |z^*|$ than $|z_k| - |z_k|\gamma(z_k)$. Since examples also show that $\{|z_k - z^*|\}$ and $\{|s(-z_k) - z^*|\}$ are not necessarily decreasing, it is not possible to improve similarly the bounds given in (v) and (vi).

Suppose $|z^*| > 0$ and $s(\cdot)$ is continuous in a neighborhood of $-z^*$ (the latter is certainly implied if $K$ is strictly convex). Then it follows from the continuity of $\gamma(\cdot)$ and (iii) that the upper bound in (vi) converges to zero. Thus $\{s(-z_k)\}$ may be used as an approximating sequence, an approach which may be advantageous in some situations (see §8). In addition it is clear from (iv) that

$$|s(-z_k)| - |z^*| \leqq |s(-z_k)| - \max_{i\leqq k}|z_i|\gamma(z_i),$$

where the right side converges to zero. Therefore meaningful stopping criteria are available.

**4. Proof of Theorem 2.** First, some basic inequalities are stated. From $z^* \in P(-z^*)$, $0 \notin K$, and $s(-y) \in P(-y)$, $y \neq 0$, it follows by the definition of $P(\cdot)$ that

$$z^*\cdot z^* \leqq z\cdot z^*, \quad z \in K, 0 \notin K; \tag{4.1}$$

$$s(-y)\cdot y \leqq z\cdot y, \quad z \in K, y \in E^n. \tag{4.2}$$

These inequalities lead to

$$|z^*|^2 \leqq s(-y)\cdot z^*, \quad 0 \notin K, y \in E^n; \tag{4.3}$$

$$s(-y)\cdot y \leqq z^*\cdot y, \quad y \in E^n; \tag{4.4}$$

$$s(-z)\cdot z \leqq |z|^2, \quad z \in K; \tag{4.5}$$

$$|y - z^*|^2 + z^*\cdot(y - z^*) \leqq y\cdot(y - s(-y)), \quad y \in E^n; \tag{4.6}$$

$$|z - z^*|^2 \leqq z\cdot(z - s(-z)), \quad z \in K, 0 \notin K. \tag{4.7}$$

Inequalities (4.3), (4.4), and (4.5) are deduced from (4.1) and (4.2) by obvious substitutions. From the identity

$$| y - z^* |^2 + z^* \cdot (y - z^*) + y \cdot (z^* - s(-y)) = y \cdot (y - s(-y)),$$

(4.6) follows by (4.4). Inequality (4.7) follows from (4.6) by use of (4.1).

Part (i) of the theorem depends on $\alpha_k \in I(z_k)$ which insures $0 \leqq \alpha_k \leqq 1$. Thus from (3.3), $s(-z_k) \in K$, and the convexity of $K$, $z_k \in K$ implies $z_{k+1} \in K$.

Consider now the inequalities in (iv), (v), and (vi). From (4.4) and the Schwarz inequality, $s(-y) \cdot y \leqq | y | \cdot | z^* |$. Thus (iv) follows from (3.2). The proof of the inequalities in (v) and (vi) makes use of $z = z_k \in K$. For $s(-z) \cdot z > 0$, $z \cdot (z - s(-z)) = | z |^2 (1 - \gamma(z))$ and from (4.7) the inequality in (v) is true. Now consider $s(-z) \cdot z \leqq 0$, which corresponds to $\gamma(z) = 0$. For $z^* = 0$, $\gamma(z) = 0$ (Theorem 1) and (v) holds as an equality; for $z^* \neq 0$, the inequality in (v) follows from (4.1) which insures

$$| z - z^* |^2 = | z |^2 - 2z \cdot z^* + 2 | z^* |^2 - | z^* |^2 \leqq | z |^2.$$

If $z^* = 0$ the inequality in (vi) is trivially true. Consider now $z^* \neq 0$. If $s(-z) \cdot z \leqq 0$, (vi) reduces to $-2s(-z) \cdot z^* + | z^* |^2 \leqq 0$ which is true by (4.3). The following identity is easily verified:

$$| s(-z) - z^* |^2 = | s(-z) - \gamma(z) z |^2 + | z |^{-2} (s(-z) \cdot z)^2 + | z^* |^2 - 2s(-z) \cdot z^*.$$

Assuming $s(-z) \cdot z > 0$ and using $s(-z) \cdot z \leqq | z | \cdot | z^* |$ yields $| z |^{-2} (s(-z) \cdot z)^2 \leqq | z^* |^2$. Thus

$$| s(-z) - z^* |^2 \leqq | s(-z) - \gamma(z) z |^2 + 2(| z^* |^2 - s(-z) \cdot z^*)$$

and by (4.3) the inequality in (vi) follows.

In order to complete the proof of the theorem, the function

$$\Gamma(z) = | z |^2 - | z^* |^2 = | z - z^* |^2 + 2z^* \cdot (z - z^*) \tag{4.8}$$

is introduced. For $0 \notin K$ inequality (4.1) gives

$$0 \leqq | z - z^* |^2 \leqq \Gamma(z), \quad z \in K, \tag{4.9}$$

a result which is obviously true for $0 \in K$. In the following paragraphs it will be shown that $\{\Gamma(z_k)\}$ is decreasing and $\Gamma(z_k) \to 0$. By (4.8) and (4.9) this proves (ii) and (iii). The remaining results in (iv) and (v) follow from the known value of $\gamma(z^*)$, the continuity of $\gamma(\cdot)$, and (iii).

For simplicity let

$$\Delta(z; \alpha) = \Gamma(z) - \Gamma(z + \alpha(s(-z) - z)), \tag{4.10}$$

and assume tacitly in what follows that $z \in K$. Then from (4.8),

$$\Delta(z; \alpha) = 2\alpha(| z |^2 - s(-z) \cdot z) - \alpha^2 | z - s(-z) |^2. \tag{4.11}$$

Because the coefficient of $\alpha^2$ is not positive, $\min_{\alpha \in I(z)} \Delta(z; \alpha)$ is attained at one of the end points of $I(z)$. It is readily shown that

$$\Delta(z; \delta\beta(z)) = \Delta(z; (2 - \delta)\beta(z)).$$

Thus from the definition of $I(z)$,

$$\min_{\alpha \in I(z)} \Delta(z; \alpha) = \begin{cases} \Delta(z; \delta\beta(z)) & \text{if} \quad \beta(z) \leqq \delta^{-1}, \\ \Delta(z; 1) & \text{if} \quad \beta(z) \geqq \delta^{-1}. \end{cases} \tag{4.12}$$

Equation (4.12) is now used to obtain a lower bound on $\Delta(z; \alpha)$, $\alpha \in I(z)$. From (4.11) and (3.1) it follows that

$$\Delta(z; \delta\beta(z)) = |z - s(-z)|^{-2}[z \cdot (z - s(-z))]^2(2\delta - \delta^2). \tag{4.13}$$

Let

$$\mu = \max_{z_1, z_2 \in K} |z_1 - z_2| \tag{4.14}$$

denote the diameter of $K$ and recall that $0 < \delta \leqq 1$. Then

$$\Delta(z; \delta\beta(z)) \geqq \mu^{-2}\delta[z \cdot (z - s(-z))]^2. \tag{4.15}$$

From (4.8) and (4.6),

$$\Gamma(z) \leqq 2\,|z - z^*|^2 + 2z^* \cdot (z - z^*) \leqq 2z \cdot (z - s(-z)) \tag{4.16}$$

(for $z^* = 0$ this may be sharpened to $\Gamma(z) \leqq z \cdot (z - s(-z))$). Thus

$$\Delta(z; \delta\beta(z)) \geqq \tfrac{1}{4}\mu^{-2}\delta\Gamma^2(z). \tag{4.17}$$

For $\beta(z) \geqq 1$, $z \cdot (z - s(-z)) \geqq |z - s(-z)|^2$ and consequently

$$\Delta(z; 1) = 2z \cdot (z - s(-z)) - |z - s(-z)|^2 \geqq z \cdot (z - s(-z)).$$

Therefore (4.16) yields

$$\Delta(z; 1) \geqq \tfrac{1}{2}\Gamma(z), \quad \beta(z) \geqq 1. \tag{4.18}$$

Finally, utilizing (4.17) and (4.18) in (4.12) yields

$$\Delta(z; \alpha)|_{\alpha \in i(z)} \geqq \min\{\tfrac{1}{4}\mu^{-2}\delta\Gamma^2(z), \tfrac{1}{2}\Gamma(z)\}. \tag{4.19}$$

Letting $z = z_k$ in (4.19), using (3.3), and returning to (4.10), it is seen that

$$\Gamma(z_k) - \Gamma(z_{k+1}) \geqq \min\{\tfrac{1}{4}\mu^{-2}\delta\Gamma^2(z_k), \tfrac{1}{2}\Gamma(z_k)\} \geqq 0. \tag{4.20}$$

Therefore the sequence $\{\Gamma(z_k)\}$ is decreasing and, since it is bounded from below by zero, has a limit point. Thus passing to the limit on the left side of (4.20) gives zero and therefore from the right side $\Gamma(z_k) \to 0$.

**5. Nature of convergence.** This section gives further results on the convergence of the iterative procedure. Theorem 3 establishes upper bounds on the elements of the sequences $\{|z_k|\}$ and $\{|z_k - z^*|\}$. Several example problems are analyzed to demonstrate still more fully the nature of convergence. Finally, a few numerical results are given. Emphasis is on the case $0 \notin K$, since it appears that it is most important in applications.

THEOREM 3. *Let*

$$\theta_k = \theta_0(1 + \tfrac{1}{4}\mu^{-2}\delta\theta_0 k)^{-1}, \quad \theta_0 = |z_0|^2 - |z^*|^2, \tag{5.1}$$

*and assume that* $|z_0|^2 \leqq |z^*|^2 + 2\mu^2\delta^{-1}$. *Then if* $\{z_k\}$ *is generated by the iterative procedure, the following inequalities hold for* $k \geqq 0$:

$$|z_k| \leqq \sqrt{\theta_k + |z^*|^2}, \tag{5.2}$$

$$|z_k - z^*| \leqq \sqrt{\theta_k}\,. \tag{5.3}$$

The assumption on $|z_0|$ is often met in practice. For example, it is easily shown that it must be satisfied if $|z^*| \leqq \frac{1}{2}(2\delta^{-1} - 1)\mu$. In any case, $z_0$ may be interpreted as a suitable intermediate point in the iterative process, and inequalities (5.2) and (5.3) may be used to estimate the subsequent rate of convergence.

For $|z^*| > 0$ and $k \geqq 1$ inequalities (5.2) and (5.3) imply

$$|z_k| - |z^*| < 2\mu^2|z^*|^{-1}\delta^{-1}k^{-1}, \tag{5.4}$$

$$|z_k - z^*| < 2\mu\delta^{-1/2}k^{-1/2}, \tag{5.5}$$

results which conform closely to (5.2) and (5.3) for $k$ sufficiently large. In Examples 1 and 2, which appear later in this section, it is demonstrated that within a constant multiplicative factor it is impossible to obtain bounds on $|z_k| - |z^*|$ and $|z_k - z^*|$ which approach zero more rapidly than those given in (5.4) and (5.5).

*Proof of Theorem* 3. Since $|z_0|^2 \leqq |z^*|^2 + 2\mu^2\delta^{-1}$, it follows from the previous section that $\Gamma(z_k) \leqq \Gamma(z_0) \leqq 2\mu^2\delta^{-1}$, $k \geqq 0$. From (4.20) this implies

$$\Gamma(z_{k+1}) \leqq \Gamma(z_k) - \tfrac{1}{4}\mu^{-2}\delta\Gamma^2(z_k), \quad k \geqq 0.$$

Since

$$1 - \tfrac{1}{4}\mu^{-2}\delta\Gamma \leqq (1 + \tfrac{1}{4}\mu^{-2}\delta\Gamma)^{-1}$$

for all $\Gamma \geqq 0$, it is possible to write

$$\Gamma(z_{k+1}) \leqq \Gamma(z_k)(1 + \tfrac{1}{4}\mu^{-2}\delta\Gamma(z_k))^{-1}, \quad k \geqq 0. \tag{5.6}$$

But substitution shows that $\theta_k$ is the solution of

$$\theta_{k+1} = \theta_k(1 + \tfrac{1}{4}\mu^{-2}\delta\theta_k)^{-1}, \tag{5.7}$$

with $\theta_0 = |z_0|^2 - |z^*|^2 = \Gamma(z_0)$. Thus comparison of (5.6) and (5.7) yields $\Gamma(z_k) \leqq \theta_k$, $k \geqq 0$. Finally, (5.2) and (5.3) follow from (4.8) and (4.9).

The complexity of the difference equation (3.3) makes it difficult to obtain more specific analytic results than those obtained in Theorem 3. Thus the remainder of this section is limited to the presentation and discussion of three, somewhat specialized, example problems and a few numerical results.

*Example* 1. Take $\delta = 1$ and let $K$ be the convex hull of three points in 2-space, $(1, \nu)$, $(-1, \nu)$, $(0, 1 + \nu)$, where $\nu > 0$. Clearly $z^* = (0, \nu)$ and $|z^*| = \nu$. Simple inspection shows that the iterative process is finite ($z_1 = z^*$) if and only if $z_0$ is on the line segment connecting $(1, \nu)$ and $(-1, \nu)$. Moreover when the process is not finite, $z_k$, $k \geqq 1$, is determined by the scalar $\psi_k = |z_k^1|(z_k^2)^{-1}$. Thus the second order nonlinear difference equation (3.3) may be replaced by a first order difference equation in $\psi_k$. It is not difficult to show that

$$\psi_{k+1} = \psi_k(1 - \nu\psi_k)(1 + \nu\psi_k + 2\psi_k^2)^{-1}, \quad k \geqq 1. \tag{5.8}$$

For $\psi_k \ll 1$ this equation is approximated by $\tilde{\psi}_{k+1} = \tilde{\psi}_k(1 + 2\nu\tilde{\psi}_k)^{-1}$, an equation of the same form as (5.7). These observations and some tedious, but straightforward, computations lead to (the notation $o(\omega)$ means $\lim_{\omega\to 0} \omega^{-1} o(\omega) = 0$)

$$|z_k| - |z^*| = (2\nu k)^{-1} + o(k^{-1}), \tag{5.9}$$

$$|z_k - z^*| = (2\nu k)^{-1}\sqrt{1 + \nu^2} + o(k^{-1}). \tag{5.10}$$

Equation (5.9) demonstrates that it is impossible to obtain an upper bound on $|z_k| - |z^*|$ which approaches zero more rapidly than (const.)$k^{-1}$. For large $k$ the upper bound in (5.4) is conservative by a factor of sixteen. This factor can be traced to two sources each of which contributes a factor of four: in (4.15), $\mu$ is an unsatisfactory estimate of $|z_k - s(-z_k)|$, in the derivation of (4.6) the term $y \cdot (z^* - s(y))$ has been omitted. For this example the upper bound in (5.5) is a poor estimate because it is order $k^{-1/2}$ rather than order $k^{-1}$.

It is also possible to show that

$$|z^*| - \gamma(z_k)|z_k| = (2\nu k)^{-1} + o(k^{-1}), \tag{5.11}$$

$$\sqrt{1 - \gamma(z_k)}\,|z_k| = k^{-1/2} + o(k^{-1/2}). \tag{5.12}$$

By comparing (5.11) with (5.9) and (5.12) with (5.10) it is seen that in Theorem 2, part (iv) provides a reasonably good stopping criterion while (v) does not.

*Example* 2. Take $\delta = 1$ and let $K$ be the convex hull of three points in

3-space, $(1, 0, \nu)$, $(-1, 0, \nu)$, $(0, 1, \nu)$, where $\nu > 0$. Thus $z^* = (0, 0, \nu)$ and $|z^*| = \nu$. The iterative process is much the same as in Example 1, the points $z_k \in K$, $k \geqq 1$, being determined by the scalar $\psi_k = |z_k^1|(z_k^2)^{-1}$. The first order difference equation for $\psi_k$ is (5.8) with $\nu = 0$. By using the fact that $\tilde{\psi}_k = \tilde{\psi}_0(1 + 4\tilde{\psi}_0^2 k)^{-1/2}$ is the solution of $\tilde{\psi}_{k+1} = \tilde{\psi}_k(1 + 4\tilde{\psi}_k^2)^{-1/2}$ and that $(1 + 4\tilde{\psi}_k^2)^{1/2} \cong 1 + 2\psi_k^2$ for $\psi_k \ll 1$, the following results can be derived:

$$|z_k| - |z^*| = (8\nu k)^{-1} + o(k^{-1}), \tag{5.13}$$

$$|z_k - z^*| = (2k^{1/2})^{-1} + o(k^{-1/2}), \tag{5.14}$$

$$|z^*| - \gamma(z_k)|z_k| = 3(8\nu k)^{-1} + o(k^{-1}), \tag{5.15}$$

$$\sqrt{1 - \gamma(z_k)}\,|z_k| = (2k)^{-1/2} + o(k^{-1/2}). \tag{5.16}$$

Equation (5.14) shows that the asymptotic behavior of $|z_k - z^*|$ matches the bound given in (5.5), except for a multiplicative factor of eight. The bound given in (5.4) is conservative by a multiplicative factor of 64. Comparison of (5.15) with (5.13) and (5.16) with (5.14) shows that (iv) and (v) of Theorem 2 both provide reasonable stopping criteria.

*Example* 3. Take $\delta = 1$ and in $n$-space let

$$K = \{z \mid z^1 \geqq \nu + \frac{1}{2}\sum_{i=2}^{n} (z^i)^2\lambda_i^{-1},\ z^1 \leqq 2\nu\}, \qquad \nu, \lambda_2, \cdots, \lambda_n > 0. \tag{5.17}$$

In the neighborhood of $z^* = (\nu, 0, 0, \cdots, 0)$, $\partial K$ is the elliptic hyperparaboloid

$$z^1 = \nu + \frac{1}{2}\sum_{i=2}^{n} (z^i)^2\lambda_i^{-1},$$

where $\lambda_2, \cdots, \lambda_n$ are the principal radii of curvature at the vertex $z^*$. For many convex sets $K$, $\partial K$ in the neighborhood of $z^*$ may be closely approximated by such an elliptic hyperparaboloid. Thus this example is of more general interest than the previous examples.

For $y^1 < 0$ and

$$\frac{1}{2}\sum_{i=2}^{n} (y^1)^{-2}\lambda_i(y^i)^2 < \nu,$$

it is easy to show that

$$\begin{aligned} s^1(y) &= \nu + \frac{1}{2}\sum_{i=2}^{n} (y^1)^{-2}\lambda_i(y^i)^2, \\ s^i(y) &= -(y^1)^{-1}\lambda_i y^i, \qquad i = 2, \cdots, n. \end{aligned} \tag{5.18}$$

Let $\bar{\lambda} = \max_{i=2,\cdots,n}\{\lambda_i\}$ and assume the conditions

$$\zeta = \nu\sqrt{1 + 2\nu\bar{\lambda}^{-1}} > |y|, \quad -\nu \geqq y^1 \tag{5.19}$$

are satisfied, which in turn imply

$$\frac{1}{2}\sum_{i=2}^{n} (y^1)^{-2}\lambda_i(y^i)^2 < \nu.$$

Thus (5.19) defines a set on which (5.18) is valid. Using this fact, $z^1 \geqq \nu$ for $z \in K$, and (3.1) gives

$$\beta(z) = \left((z^1 - \nu)^2 + \nu(z^1 - \nu) + \sum_{i=2}^{n}\left(1 + \frac{1}{2}(z^1)^{-1}\lambda_i\right)(z^i)^2\right) \Big/$$

$$\left((z^1 - \nu)^2 + \sum_{i=2}^{n} [1 + (z^1)^{-1}\lambda_i + (z^1)^{-2}\nu\lambda_i + (z^1)^{-2}\lambda_i^2] \right. \tag{5.20}$$

$$\left. \cdot (z^i)^2 + \left[\frac{1}{2}\sum_{i=2}^{n} (z^1)^{-2}\lambda_i(z^i)^2\right]^2\right), \quad z \neq z^*, z \in K, |z| < \zeta.$$

Because $z \in K$, $|z| < \zeta$ imply $z^1 \geqq \nu$ and

$$\frac{1}{2}\sum_{i=2}^{n} (z^1)^{-2}\lambda_i(z^i)^2 < \nu,$$

it follows that

$$\beta(z) \geqq \frac{(z^1 - \nu)^2 + \sum_{i=2} (z^i)^2}{(z^1 - \nu)^2 + (1 + \frac{5}{2}\bar{\lambda}\nu^{-1} + \bar{\lambda}^{-2}\nu^{-2})\sum_{i=2}^{n} (z^i)^2}$$

$$\geqq \frac{1}{1 + \frac{5}{2}\bar{\lambda}\nu^{-1} + \bar{\lambda}^{-2}\nu^{-2}} = \underline{\beta}, \qquad z \neq z^*, z \in K, |z| < \zeta. \tag{5.21}$$

Because $\beta(z^*) = 0$ this inequality implies that $\beta(z)$ is discontinuous on $K$ at $z^*$.

By starting with (4.13) and repeating the derivation of §4 with

$$[z \cdot (z - s(-z))] \, |z - s(-z)|^{-2} = \beta(z) \geqq \underline{\beta},$$

it can be shown that

$$\Gamma(z_{k+1}) \leqq \Gamma(z_k)(1 - \tfrac{1}{2}\underline{\beta}\delta), \quad z_k \neq z^*. \tag{5.22}$$

For $z_k = z^*$, $\Gamma(z_{k+1}) = 0$ and (5.22) is trivially true. Thus

$$\Gamma(z_k) \leqq \Gamma(z_0)(1 - \tfrac{1}{2}\underline{\beta}\delta)^k, \quad k \geqq 0, \quad z_0 \in K, \quad |z_0| < \zeta.$$

Using (4.8) and (4.9) this leads to

$$|z_k| - |z^*| \leqq \tfrac{1}{2}\nu^{-1}\theta_0(1 - \tfrac{1}{2}\underline{\beta}\delta)^k, \tag{5.23}$$

$$|z_k - z^*| \leqq \sqrt{\theta_0}(1 - \tfrac{1}{2}\underline{\beta}\delta)^{k/2}, \tag{5.24}$$

TABLE 1

*Number of iterations to satisfy error criteria*

*First column: first $k$ for which $|z_k| - |z^*| \leqq \epsilon$*
*Second column: first $k$ for which $|z^*| - \gamma(z_k)|z_k| \leqq \epsilon$*

| Case | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda_2$ | 1 | | 10 | | 100 | | 1000 | | 100 | | 1000 | |
| $\lambda_3$ | 1 | | 10 | | 100 | | 1000 | | 10 | | 10 | |
| $\epsilon$ | | | | | | | | | | | | |
| $10^{-3}$ | 3 | 2 | 28 | 20 | 59 | 18 | 216 | 83 | 27 | 14 | 229 | 82 |
| $10^{-4}$ | 5 | 4 | 31 | 27 | 59 | 35 | 250 | 88 | 52 | 26 | 267 | 125 |
| $10^{-5}$ | 5 | 4 | 38 | 30 | 74 | 58 | 290 | 162 | 73 | 51 | 298 | 167 |
| $10^{-6}$ | 5 | 4 | 41 | 37 | 111 | 58 | 340 | 215 | 81 | 51 | 359 | 218 |

where $\theta_0$ is given as before in (5.1). Since $\beta > 0$ inequalities (5.23) and (5.24) guarantee that the convergence of $\{|z_k|\}$ and $\{|z_k - z^*|\}$ is geometric. However, the guaranteed rate of convergence is not rapid if $\beta \ll 1$, i.e., $\nu^2 \ll \bar{\lambda}^2$.

Table 1 presents some numerical results for Example 3, $\nu = 1$, $n = 3$, and $z_0 = (6, 2, 2)$. Similar results are obtained for different $z_0$. The extent of $K$ has been increased beyond $z^1 = 2\nu$ so that (5.18) is valid even though (5.19) is violated. Note that convergence is slow when $\nu \ll \bar{\lambda}$. Although the bounds derived in the preceding paragraph follow the same pattern it may be concluded from Table 1 that they are not sharp estimates of actual convergence rate. Better estimates than (5.23) and (5.24) have been obtained but their derivation is too lengthy to present here. It is interesting to note that Cases 3 and 4 exhibit rates of convergence which are respectively similar to Cases 5 and 6. Thus $\nu/\bar{\lambda}$ seems to be the key parameter while $\lambda_3/\lambda_2$ has little effect. This is not true when the gradient method of the next section is used ($\lambda_3 \gg \lambda_2$ corresponds to a "ridge" of $f(y)$).

Fig. 2 shows the details of Case 5. The irregularity of the sequences shown is typical. Various methods for accelerating convergence (based on different rules for selecting $\alpha_k \in I(z_k)$, the results of the next section, etc.) are being investigated and will be reported in a later paper.

**6. Relation to a gradient method.** The iterative procedure described in the preceding sections is related to a gradient method, which is similar in approach to certain gradient based methods which have been proposed for the solution of a variety of problems in optimal control [2], [4], [9], [10], [11]. The purpose of this section is to illustrate both the differences

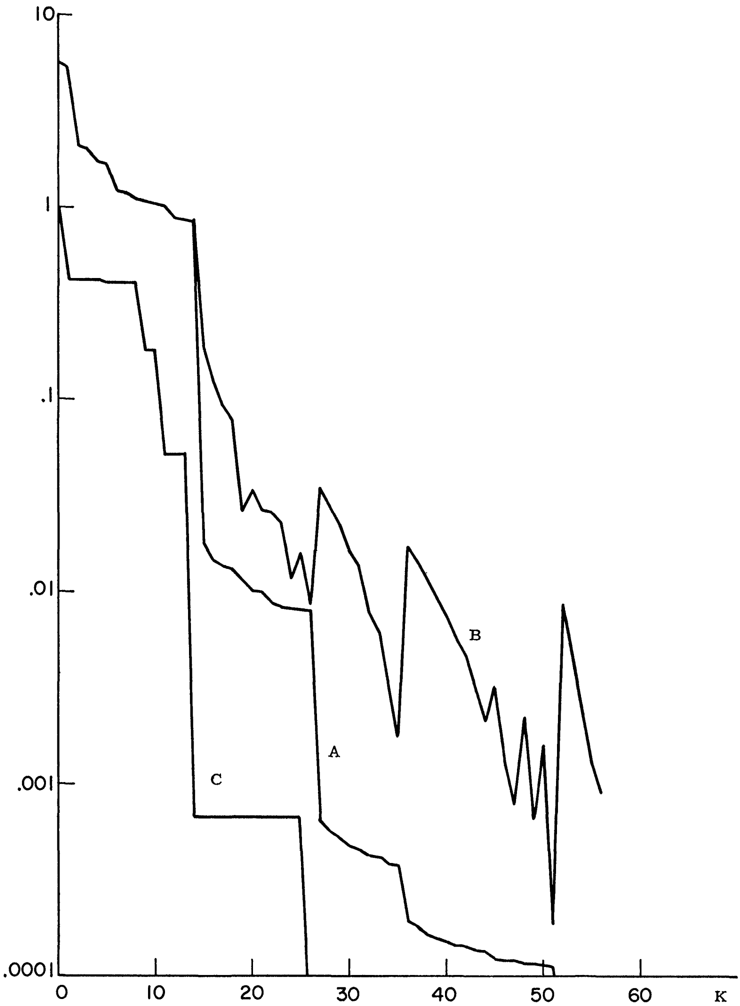


FIG. 2. *Numerical results for Case 5 of §6:* (A) $|z_k| - |z^*|$, (B) $|z_k - z^*|$, (C) $|z^*| - \max_{i \leq k} |z_i| \gamma(z_i)$. *For* $k \leqq 14$, $|z_k - z^*| \cong |z_k| - |z^*|$.

and strong connections between the two approaches. For brevity the developments which follow are presented somewhat superficially and without proof.

THEOREM 4. *Assume* $0 \notin K$ *and let* $J = \{y \mid y \cdot s(-y) > 0\}$. *Then for* $y \in J$ *the scalar function*

$$f(y) = |y|^{-1}(y \cdot s(-y)) = \gamma(y)|y| \tag{6.1}$$

*is defined and has the following properties*:

(i) $0 < f(y) \leqq |z^*|$,

(ii) $f(y) = |z^*|$ *if and only if* $y = \rho z^*$, $\rho > 0$.

*Further assume that* $K$ *is strictly convex. Then*:

(iii) $s(-\rho z^*) = z^*$, $\rho > 0$,

(iv) *the gradient of* $f(y)$ *exists and is given by*

$$\nabla f(y) = |y|^{-1}s(-y) - |y|^{-3}(y \cdot s(-y))y,$$

(v) $\nabla f(y) = 0$ *if and only if* $y = \rho z^*$, $\rho > 0$.

Theorem 4 forms the basis for the gradient method. A sequence of vectors $\{y_k\}$ is generated by

$$y_{k+1} = y_k + \sigma_k \nabla f(y_k), \quad y_0 \in J. \tag{6.2}$$

If $K$ is strictly convex, $0 \notin K$, and the positive numbers $\sigma_k$ are appropriately chosen, it can be shown that $y_k \in J$, $k \geqq 0$, $\{f(y_k)\}$ is increasing, and $y_k \to \rho z^*$, $\rho > 0$, for $k \to \infty$. Strict convexity of $K$ also assures that $s(y)$ is continuous on $J$. This, $s(\omega y) = s(y)$ for $\omega > 0$, and solution property (iv) (§2) guarantee that $\{s(-y_k)\}$ is an approximating sequence for $z^*$, i.e., $s(-y_k) \to z^*$. Disadvantages of the gradient method, relative to the procedure of §3, are: $K$ must be strictly convex, methods for choosing the values of $\sigma_k$ may be cumbersome and time consuming, the selection of a $y_0$ in $J$ may be difficult. On the other hand it is conceivable that the gradient method may yield more rapid convergence, particularly when variations of (6.2) are employed.

Consider now a modified version of the gradient method. Since from (iv) of Theorem 4, $\nabla f(\rho^{-1}y) = \rho \nabla f(y)$, $\rho > 0$, the difference equation

$$z_{k+1} = \rho_{k+1}\rho_k^{-1}(z_k + \sigma_k \rho_k^2 \nabla f(z_k)), \quad z_0 = y_0 \in J, \tag{6.3}$$

with $\rho_0 = 1$ and $\rho_k > 0$, $k > 0$, yields a sequence $\{z_k\}$ such that $z_k = \rho_k y_k$, $k \geqq 0$. Thus $s(-y_k) = s(-z_k)$, $k \geqq 0$. By letting

$$\rho_{k+1} = \rho_k[1 + \sigma_k \rho_k^2 |z_k|^{-1}(1 - \gamma(z_k))]$$

and

$$\alpha_k = \sigma_k \rho_k \rho_{k+1} |z_k|^{-1}, \ k \geqq 0, \tag{6.4}$$

it is easy to show that (6.3) becomes (3.3). Thus if $z_0 \in K \cap J$, (3.3) realizes the modified version of the gradient method, where the selection rule for $\alpha_k$ is (6.4) rather than (3.4). If the $\alpha_k$ as obtained from (6.4) happen to be in $I(z_k)$, $k \geqq 0$, then all the results of Theorem 2 follow; in particular $\{z_k\}$, whose elements are in $K$, is also an approximating sequence.

In any case the iterative procedure described in §3 takes "steps" in the same direction as those indicated by the modified gradient method. The "step size" prescribed by (3.4) may be much larger than that prescribed by (6.4). Thus with (3.4) the sequence $\{f(z_k)\}$ is not necessarily increasing.

**7. Extension to more general quadratic forms.** The iterative procedure for the Basic Problem can be extended without great difficulty to the general problem:

GP. *Given $C$, a compact convex set in $E^m$, and the quadratic form*

$$q(x) = |x|_G^2 + g\cdot x, \tag{7.1}$$

*where $|x|_G^2 = x\cdot Gx$, $G$ is a symmetric nonnegative definite $m \times m$ matrix, and $g$ is an $m$-vector in the range of $G$, find a point $x^* \in C$ such that*

$$q(x^*) = q^* = \min_{x\in C} q(x).$$

Clearly a solution $x^*$ exists. In order to obtain its essential properties and derive the iterative procedure it is convenient to write $q(x)$ as

$$q(x) = |Hx - a|^2 + q_0, \tag{7.2}$$

where $H$ is an $n \times m$ matrix, $n = \text{rank } G$, $G = H'H$ (the $'$ denotes matrix transpose), $a = \frac{1}{2}(HH')^{-1}Hg$ or equivalently $g = 2H'a$, and

$$q_0 = -|a|^2 = \min_{x\in E^m} q(x).$$

The existence of such a representation is a consequence of the hypotheses in the statement of GP. Introducing the set $K = \{z \mid z = Hx + a,\ x \in C\}$ it is clear that

$$q^* = \min_{z\in K} |z|^2 + q_0 = |z^*|^2 + q_0, \tag{7.3}$$

where $z^*$ is defined as before. Furthermore since $z^* \in K$ is unique it follows that $F = \{x \mid Hx + a = z^*,\ x \in C\}$ is the set of all solutions of GP. Since $F$ may sometimes contain more than a single point, $x^*$ is not necessarily unique.

The iterative procedure for GP is developed from the results of §3 by noting that for every point $x \in C$ there is, by means of

$$z = Hx + a, \tag{7.4}$$

a corresponding point $z \in K$. Thus, for example,

$$\max_{z\in K} y\cdot z = \max_{x\in C} y\cdot(Hx + a) = s_C(H'y)\cdot H'y + y\cdot a = y\cdot(Hs_C(H'y) + a),$$

where $s_C(\cdot)$ is a contact function of $C$. Therefore a contact function of $K$ is

$$s(y) = Hs_C(H'y) + a. \tag{7.5}$$

Using this result and $H'(Hx + a) = Gx + \frac{1}{2}g$, it is further seen that the equation

$$x_{k+1} = x_k + \alpha_k(s_C(-Gx_k - \tfrac{1}{2}g) - x_k) \tag{7.6}$$

when transformed by (7.4) yields the same sequence as (3.3). Hence if $\alpha_k \in I(Hx_k + a)$ and $x_0 \in C$, (7.6) yields a sequence $\{x_k\}$ with elements in $C$ such that $q(x_k)$ converges downward to $q^*$. This and other results are summarized in the following.

THEOREM 5. *Let $s_C(\cdot)$ be a contact function of the set $C$ specified in GP. Define*

$$v_k = Gx_k + \tfrac{1}{2}g; \tag{7.7}$$

$$\beta_k = \begin{cases} |x_k - s_C(-v_k)|_G^{-2}[v_k\cdot(x_k - s_C(-v_k))] \text{ if } G(x_k - s_C(-v_k)) \neq 0, \\ 0 \text{ if } G(x_k - s_C(-v_k)) = 0; \end{cases} \tag{7.8}$$

$$\gamma_k = \begin{cases} \max\{(1 - (q(x_k) - q_0)^{-1}[v_k\cdot(x_k - s_C(-v_k))]), 0\} \\ \qquad\qquad\qquad\qquad\qquad \text{if } q(x_k) \neq q_0, \\ 0 \text{ if } q(x_k) = q_0; \end{cases} \tag{7.9}$$

$$I_k = [\min\{\delta\beta_k, 1\}, \min\{(2 - \delta)\beta_k, 1\}], \quad 0 < \delta \leqq 1; \tag{7.10}$$

$$x_{k+1} = x_k + \alpha_k(s_C(-v_k) - x_k), \quad x_0 \in C, \quad \alpha_k \in I_k. \tag{7.11}$$

*By means of* (7.11) *generate* $\{x_k\}$. *Then* $\beta_k \geqq 0$, $k \geqq 0$; *and* $\beta_k = 0$ *implies* $x_k \in F$. *Furthermore, for* $k \geqq 0$ *and* $k \to \infty$:

(i) $x_k \in C$,

(ii) $\{q(x_k)\}$ *is decreasing and* $q(x_k) \to q^*$,

(iii) *there is a convergent subsequence of* $\{x_k\}$ *and every convergent subsequence of* $\{x_k\}$ *has its limit point in* $F$,

(iv) $\gamma_k^2 q(x_k) + (1 - \gamma_k^2)q_0 \leqq q^*$ *and* $\gamma_k^2 q(x_k) + (1 - \gamma_k^2)q_0 \to q^*$,

(v) $|x_k - \tilde{x}|_G^2 \leqq (1 - \gamma_k)(q(x_k) - q_0)$ *for all* $\tilde{x} \in F$ *and* $(1 - \gamma_k)(q(x_k) - q_0) \to 0$,

(vi) $|s_C(-v_k) - \tilde{x}|_G^2 \leqq |s_C(-v_k) - \gamma_k x_k|_G^2 + (1 - \gamma_k)g\cdot(s_C(-v_k) - \gamma_k x_k) - (1 - \gamma_k)^2 q_0$ *for all* $\tilde{x} \in F$.

*Proof.* Part (iii) follows from (i), (v), the definition of $F$, and the compactness of $C$. The remaining parts follow from Theorems 1 and 2 by straightforward substitutions.

If $C$ is strictly convex and $q^* > q_0$, the upper bound in (vi) converges to zero and $\{s_C(-v_k)\}$ serves as an approximating sequence (see remarks after Theorem 2). Also the results of §§5, 6 may be extended in an obvious way. For most applications the hypothesis that $g$ is in the range of $G$ holds. When it does not hold, by a different line of attack it is still possible to derive a theorem similar to Theorem 5.

The approach taken by Frank and Wolfe [6] to the concave programming problem can be extended to give a direct proof of (i), (ii), and (iii) of Theorem 5. A lower bound for $q(x_k)$ is also obtainable but it is not as sharp as (iv).

**8. Application to problem in optimal control.** Consider the dynamical system

$$\dot{x} = A(t)x + f(u(t); t), \quad x(0), \tag{8.1}$$

where $x$ is the $m$-dimensional state vector, $\dot{x}$ is its time derivative, $x(0)$ is the initial state; $u(t)$ is an $r$-dimensional vector control function, admissible if measurable on the control interval $[0, T]$, $0 < T < \infty$, with range in a compact set $U$; $A(t)$ is an $m \times m$ matrix function continuous on $[0, T]$; $f(\cdot\,; \cdot)$ is an $m$-dimensional vector function defined and continuous on $U \times [0, T]$. For every admissible control $u(t)$ there is an absolutely continuous solution function, $x_{u(t)}(t) = x_u(t)$, which satisfies (8.1) almost everywhere in $[0, T]$. It is desired to find an admissible control $u^*(t)$ such that $q(x_{u^*}(T)) = q^* \leqq q(x_u(T))$ for all admissible controls $u(t)$, where $q(\cdot)$ is prescribed in GP, §7. This optimal control problem has a number of practical applications [1].

Under the conditions just stated, Neustadt [12] has shown that the set

$$C = \{x \mid x = x_u(T), u(t) \text{ admissible}\}$$

is compact and convex. Thus if a method for evaluating a contact function of $C$ exists, the iterative procedure of §7 can be used to obtain approximations for $x_{u^*}(T) = x^*$ and $q^*$.

To obtain a contact function of $C$, $s_C(\cdot)$, note that

$$w \cdot x_u(T) = \psi(0; w) \cdot x(0) + \int_0^T \psi(\sigma; w) f(u(\sigma); \sigma)\, d\sigma, \tag{8.2}$$

where $\psi(t; w)$, defined on $[0, T] \times E^n$, is the solution of the adjoint differential equation

$$\dot{\psi} = -A'(t)\psi, \quad \psi(T) = w. \tag{8.3}$$

Equation (8.2) follows from (8.1) by integrating $\frac{d}{dt}(\psi(t; w) \cdot x_u(t))$. Suppose there exists an admissible control $u(t; w)$ such that almost everywhere in $[0, T]$,

$$\psi(t; w)\cdot f(u(t; w), t) = \max_{\tilde{u} \in U} \psi(t; w)\cdot f(\tilde{u}, t). \tag{8.4}$$

Then from (8.2) it is clear that $w \cdot x_{u(t;\, w)}(T) \geqq w \cdot x_{u(t)}(T)$ for every admissible control $u(t)$. Thus from the definition of $C$ a contact function of $C$ is

$$s_C(w) = x_{u(t;\, w)}(T). \tag{8.5}$$

This result agrees with the well-known fact that boundary points of the reachable set $C$ must "satisfy" the Pontryagin maximum principle. For all but the most elementary systems (8.1), $s_C(\cdot)$ is the only reasonable means for numerically characterizing the set $C$.

In most practical problems it is not difficult to obtain a function $u(t; w)$ which satisfies (8.4). Consider, for example, the case where $f(u; t) = B(t)u$, $B(t)$ is an $m \times r$ matrix function continuous on $[0, T]$, and $U$ is the unit hypercube $\{u \mid |u^i| \leqq 1, i = 1, \cdots, r\}$. Notice that (8.4) may not uniquely define $u(t; w)$ almost everywhere in $[0, T]$; suppose for instance that in the example of the preceding sentence $B'(t)\psi(t; w)$ has at least one component which is identically zero on $[0, T]$. This is of no concern, since different choices for $u(t; w)$ will at most lead only to different contact functions of $C$. Previous computational procedures [2], [4], [9], [10], [11] have required assumptions which correspond to a unique determination of $u(t; w)$ by (8.4). Such "unique maximum" assumptions imply strict convexity of $C$.

Computer evaluation of $s_C(\cdot)$ entails three steps: evaluation of $\psi(t; w)$ by solving (8.3) backwards from $t = T$ to $t = 0$, determination of $u(t; w)$ from $\psi(t; w)$ by (8.4), solution of (8.1) with $u(t) = u(t; w)$ from $t = 0$ to $t = T$. Thus when the iterative procedure is applied to the optimal control problem each iteration involves the sequential solution of two differential equations. This situation is handled efficiently by a hybrid computer which includes both digital and analog elements.

The details of applying the iterative procedure should be clear. There is no difficulty in choosing $x_0 \in C$, it is only necessary to set $x_0 = x_{u^0}(T)$ where $u^0(t)$ is an arbitrary admissible control. In the sense of Theorem 5, $\{x_k\}$ and $\{q(x_k)\}$ (and if $C$ is strictly convex and $q^* > q_0$, $\{s_C(-v_k)\}$ and $\{q(s_C(-v_k))\}$) are approximating sequences and error bounds may be computed.

The issue of finding admissible control functions corresponding to $x_k$ or $s_C(-v_k)$ remains. The control corresponding to $s_C(-v_k)$ is $u(t; -v_k)$, i.e.,

$$s_C(-v_k) = x_{u(t;\, -v_k)}(T).$$

Finding an admissible control which produces the terminal state $x_k$ is more difficult. From (7.11) it follows that

$$x_k = \sum_{i=1}^{k-1} \lambda_i s_C(-v_i) + \lambda_0 x_0 ,$$

where $\lambda_i \geqq 0$, $0 \leqq i < k$, and $\sum_{i=0}^{k-1} \lambda_i = 1$. Suppose $u_k(t)$ is an admissible control such that almost everywhere in $[0, T]$,

$$f(u_k(t), t) = \sum_{i=1}^{k-1} \lambda_i f(u(t; -v_i); t) + \lambda_0 f(u^0(t); t).$$

Then from the form of (8.1) it may be deduced that $x_{u_k(t)}(T) = x_k$. If for all $t \in [0, T]$ the sets $f(U; t)$ are convex such a choice is possible. If this is not the case an additional approximation process, the construction of a chattering control, is necessary [1]. For $f(u, t) = B(t)u$ and $U$ convex it follows that

$$u_k(t) = \sum_{i=1}^{k-1} \lambda_i u(t; -v_i) + \lambda_0 u_0(t)$$

or equivalently

$$u_{i+1}(t) = u_i(t) + \alpha_i[u(t, -v_i) - u_i(t)], \qquad i = 0, \cdots, k-1.$$

For additional details on application of the iterative procedure to a variety of problems in optimal control, see [1].

**9. Acknowledgments.** The author wishes to thank L. W. Neustadt and Robert O. Barr for helpful comments during the development of the material reported above. The computational results in §5 are due to Robert O. Barr.

REFERENCES

[1] R. O. Barr and E. G. Gilbert, *Some iterative procedures for computing optimal controls*, to appear.

[2] J. H. Eaton, *An iterative solution to time-optimal control*, J. Math. Anal. Appl., 5 (1962), pp. 329–344.

[3] H. G. Eggleston, *Convexity*, Cambridge University Press, Cambridge, 1958.

[4] E. J. Fadden and E. G. Gilbert, *Computational aspects of the time optimal problem*, Computing Methods in Optimization Problems, Academic Press, New York, 1964, pp. 167–192.

[5] P. S. Fancher, *Iterative computation procedures for an optimum control problem*, IEEE Trans. Automatic Control, AC-10 (1965), pp. 346–348.

[6] M. Frank and P. Wolfe, *An algorithm for quadratic programming*, Naval Res. Logist. Quart., 3 (1956), pp. 95–110.

[7] A. A. Goldstein, *Convex programming in Hilbert space*, Bull. Amer. Math. Soc., 70 (1964), pp. 709–710.

[8] Y. C. Ho, *A successive approximation technique for optimal control systems subject*

*to input saturation*, Trans. ASME Ser. D. J. Basic Engrg., 84 (1962), pp. 33–40.

[9] L. W. NEUSTADT, *Synthesizing time-optimal control systems*, J. Math. Anal. Appl., 1 (1960), pp. 484–492.

[10] ———, *On synthesizing optimal controls*, Proceedings of Second Congress of IFAC, Butterworth, London, 1964.

[11] ———, *Minimum effort control systems*, this Journal, 1 (1962), pp. 16–31.

[12] ———, *The existence of optimal controls in the absence of convexity conditions*, J. Math. Anal. Appl., 7 (1963), pp. 110–117.