

JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 62, 432-444 (1978)

Conditional Gradient Algorithms with Open Loop Step Size Rules

J. C. DUNN

*Department of Mathematics, North Carolina State University, Raleigh,
North Carolina 27607*

AND

S. HARSHBARGER

Center for Applied Mathematics, Cornell University, Ithaca, New York

Submitted by Peter D. Lax

1. INTRODUCTION

In Ref. [1], Demyanov and Rubinov analyze a general iterative process for minimizing a smooth convex functional f over a bounded convex set Ω in a Banach space. At each stage in this process, the original minimization problem is replaced by a (presumably) simpler problem in which the local linear approximation to f , specified by f 's derivative at the current iterate x_n in Ω , is minimized over Ω . Every solution y_n of the latter problem yields a descent direction vector $p_n = y_n - x_n$, and the next iterate x_{n+1} is then gotten by moving a certain distance from x_n in the direction p_n , after which the whole cycle is repeated. Demyanov and Rubinov refer to this procedure as the conditional gradient method and it does indeed resemble the classical gradient method for minimizing f over the entire space X (in the latter case, p_n is obtained by minimizing the local linear approximation to f at x_n over the unit ball with center x_n , as in Kantorovich [2]). Both methods are formally applicable to nonconvex functionals, however in such cases the resulting iterates may render f "stationary" in a certain sense [1] without achieving a global minimum over Ω .

A precursor of the general conditional gradient method appears in a paper by Frank and Wolfe [3] on quadratic programming problems in finite dimensional spaces. Variants of this basic procedure have since been rediscovered and analyzed by many workers, mainly in the context of optimal control theory (cf. Kelley [4], Gilbert [5], Gilbert and Barr [6], Barr [7], Barnes [8], Meyer [9], Pecsvaradi [10], Dunn [11], Dunn and Kumar [12], and Kumar [13]); the method is especially interesting in this setting because it is effective on problems

with singular as well as nonsingular solutions [12, 13], and because the iterated linear minimization subproblem is frequently trivial [1]. From a certain general perspective, the conditional gradient method, like the saddle point seeking process of Brown and Robinson [14] and the stochastic approximation process of Robbins and Monro [15], is a special type of Mann iteration [16] for computing the fixed points of multivalued operators; from another viewpoint, it is a limiting case of the projected gradient method treated in [1]; finally, it and the projected gradient method are both members of the general class of "feasible direction" methods considered at some length in the recent survey article by Polak [17].

Step lengths along descent directions p_n are traditionally determined by minimizing f over the half line specified by x_n and p_n , e.g., in the classical gradient method, one puts

$$x_{n+1} = x_n + \omega_n p_n, \quad (1)$$

with

$$f(x_n + \omega_n p_n) = \min_{\omega \geq 0} f(x_n + \omega p_n), \quad (2)$$

while in the case of the conditional gradient method (with Ω bounded), the vector $y_n = x_n + p_n$ is typically on the boundary of Ω and the line minimization in (2) is therefore replaced by

$$f(x_n + \omega_n p_n) = \min_{1 \geq \omega \geq 0} f(x_n + \omega p_n). \quad (3)$$

There are two things wrong with this procedure. First, exact solutions for ω_n in (2) or (3) are usually unattainable and one must then resort to approximate line search techniques. When f is known only in terms of complicated transcendental processes, this may prove to be computationally expensive, as in certain optimal control problems where each evaluation of f entails the numerical solution of an initial value problem for a system of ordinary differential equations [12]. The second objection is more subtle: while line minimization is clearly a "locally" optimal strategy in so far as it produces the greatest possible decrease in f at each iteration, *given* x_n and p_n , it may be quite far from optimal in securing the greatest decrease in f over *many* iterations; this is clearly seen in the behavior of the gradient method when the level sets of f resemble highly elongated ellipsoids (see Rosenbrock's "ravine" problem in [18, p. 1] and also Luenberger's analysis for quadratic functionals [19]). For these reasons, interest has shifted from line minimization schemes to other more easily implemented step size rules such as those proposed by Goldstein [20] and Armijo [21] for gradient methods, and by Demyanov and Rubinov [1] for conditional gradient methods (see Note 4, Section 2). Although considerably more tractable than line minimization, these rules are still "closed loop" (to use Polak's phrase [17]) in the sense that step lengths are computed from formulas involving local properties of f

at x_n (and possibly other less accessible global properties as well, e.g., Lipschitz constants). Simplest of all are the "open loop" step size rules which determine admissible sequences $\{\omega_n\}$ beforehand, i.e., without references to the course of the iteration.

Open loop step size rules of the threshold type have been proposed and analyzed for gradient methods in Hilbert space by Goldstein [22] Polyak [23] and Vainberg [24]. As Polak points out in [17], a serious difficulty with such rules is that they require information about f which is usually not available (e.g. Lipschitz constants). Dunn [25] investigates gradient processes in Hilbert space with open loop step size constraints

$$\lim_{n \rightarrow \infty} \omega_n = 0, \quad (4a)$$

$$\sum_{n=0}^{\infty} \omega_n = \infty. \quad (4b)$$

The functionals treated are convex and have continuous and uniformly bounded second derivatives, however strict convexity of f is not invoked and it is not necessary to have a *value* for an upper bound on the spectra of f'' . Equation (4a) insures that ω_n will eventually remain below the upper thresholds of Goldstein and Polyak, whatever their values may be, while Eq. (4b) replaces the lower threshold with a weaker implicit restriction on the *rate* at which $\{\omega_n\}$ may converge to 0; the price of abandoning the lower threshold is sublinear convergence. In an early paper, Ermol'ev [26] considers open loop conditional gradient methods with the step size restriction (4) and the descent property constraint (i.e., a step is taken if and only if f is reduced) however the convergence proof given in [26] for these processes is apparently defective. More recently, Bruck [27] investigates still more general iterative processes for a certain type of variational inequality in Hilbert space; his formulation includes open loop versions of gradient and projected gradient processes with step size restrictions similar to (4). The present article extends the analysis to open loop conditional gradient processes with step size constraints like (4) and functionals f which are smooth, convex, and bounded below on a bounded set Ω in a Banach space.

A principal aim here is to establish conditions on $\{\omega_n\}$ which insure that the corresponding conditional gradient iterate sequence $\{x_n\}$ is "minimizing," i.e., $\lim_{n \rightarrow \infty} f(x_n) = \inf_{\Omega} f$. It is shown in Section 2 that $\{x_n\}$ is a minimizing sequence if f' is Lipschitz continuous and $\{\omega_n\} \subset (0, 1]$ satisfies the condition:

$$\sum_{n=0}^{\infty} \omega_n^2 < \infty \quad (5a)$$

$$\sum_{n=0}^{\infty} \omega_n = \infty \quad (5b)$$

in place of (4). In particular, for the sequence $\{\omega_n\}$ generated recursively by

$$\omega_{n+1} = \omega_n - (\omega_n)^2/2; \quad \omega_0 = 1, \quad (6)$$

it turns out that $f(x_n) - \inf_{\Omega} f = O(1/n)$ as $n \rightarrow \infty$ (a remarkable conclusion, in view of the fact that $O(1/n)$ convergence is the best one can expect of the standard closed loop conditional gradient process for certain f 's and Ω 's included in the present formulation; see Note 4). While this result does seem to depend in an essential way on the assumption of Lipschitz continuity for f' , it is possible to establish at least the fact of convergence under the weaker assumption of uniform continuity provided $\{\omega_n\}$ is confined to a certain proper subclass of the sequences satisfying (6); this subclass contains the sequence generated by (6) and the prototypical sequence $\{1/(n+1)\}$.

The question of whether the open loop conditional gradient iterates $\{x_n\}$ of Section 2 converge in some sense to the set of minimizing vectors for f is taken up briefly in Section 3.

2. CONVERGENCE OF THE FUNCTIONAL VALUES $f(x_n)$

In this and the next section, X is a real Banach space with norm, $\|\cdot\|$, X^* is the dual of X with induced norm, $\|\cdot\|_*$, $\langle u, v \rangle$ signifies the action of the linear functional $u \in X^*$ on the vector $v \in X$, Ω is a nonempty bounded convex set in X , $f: X \rightarrow \mathbb{R}^1$ is bounded below, convex, and Frechet differentiable, and f' is at least uniformly continuous on Ω .

Let Ω_f denote the set of minimizers of f in Ω , i.e.,

$$\Omega_f = \{\xi \in \Omega \mid f(\xi) = \inf_{\Omega} f\}. \quad (7)$$

At the present level of generality Ω_f may be empty, however in any case, $\xi \in \Omega_f$ iff $-f'(\xi)$ falls in the normal cone for Ω at ξ , i.e., iff

$$\langle f'(\xi), \xi - u \rangle \leq 0 \quad (8)$$

for all $u \in \Omega$ (cf. Rockafellar [28], and also [1]). Alternatively, for each $x \in \Omega$ put

$$T(x) = \{\bar{x} \in \Omega \mid \langle f'(x), \bar{x} \rangle = \inf_{u \in \Omega} \langle f'(x), u \rangle\}. \quad (9a)$$

Then by construction, $\xi \in \Omega_f$ iff $\xi \in T(\xi)$, i.e., ξ minimizes f iff ξ is a fixed point of the set-valued mapping $T: \Omega \rightarrow 2^{\Omega}$ defined by (9a).

Consider the iterative scheme

$$x_{n+1} = (1 - \omega_n) x_n + \omega_n \bar{x}_n; \quad \bar{x}_n \in T(x_n) \quad (9b)$$

with $x_0 \in \Omega$ and $\{\omega_n\}$ a given sequence with range in $(0, 1]$. In one sense, (9b) is a weighted averaging process of the Mann type for constructing fixed points of multivalued operators T , however with T given by (9a), the scheme (9b) becomes an open loop conditional gradient process for f as described in Section 1, and the following questions arise: (a) given $x_0 \in \Omega$ and $\{\omega_n\} \subset (0, 1]$, does (9) have a corresponding solution $\{x_n\} \subset \Omega$ beginning at x_0 ?, and (b), if $\{x_n\}$ is a solution of (9), does $\{f(x_n)\}$ converge to $\inf f$?

For convex Ω and $\{\omega_n\} \subset (0, 1]$, (9) will have solutions $\{x_n\} \subset \Omega$ if the linear minimization problem for $f'(x)$ always has solutions in Ω , i.e., if $T(x)$ is never empty. This condition is certainly met if Ω is weakly compact since the functionals $f'(x)$ are weakly continuous (Ω weakly compact also insures that $\Omega_f \neq \emptyset$ since f is weakly lower semicontinuous). On the other hand, once the existence of $\{x_n\}$ is granted, the convergence of $\{f(x_n)\}$ has nothing to do with weak compactness. Moreover, all of the subsequent convergence proofs for $\{f(x_n)\}$ are readily altered to suit a modified version of (9) in which \bar{x}_n is determined by the condition,

$$\langle f'(x_n), \bar{x}_n \rangle \leq \inf_{u \in \Omega} \langle f'(x_n), u \rangle + \delta_n, \quad (10)$$

with $\delta_n > 0$ and $\lim_{n \rightarrow \infty} \delta_n = 0$; this process *always* has solutions $\{x_n\}$ for bounded Ω (see Note 6). For these reasons, the weak compactness assumption is not invoked until Section 3, where it enters into the analysis at a somewhat deeper level.

With regard to the convergence question for $\{f(x_n)\}$, two possible lines of approach are apparent. First, one might attempt to establish convergence theorems for $\{x_n\}$ under conditions on T which derive naturally from its relationship to f through (9a), (e.g., conditions of the monotonicity type), and then deduce $f(x_n) \rightarrow \inf_{\Omega} f$ from the continuity properties of f . Alternatively, one might establish convergence theorems for $\{f(x_n)\}$ directly and then draw conclusions about the convergence of $\{x_n\}$ under various compactness and uniform convexity conditions on Ω , as in [1] for closed loop conditional gradient processes. The latter approach turns out to be the more fruitful for open loop conditional gradient processes as well, since the general theory of convergence for (9b) with T multivalued is presently limited to a class of bounded operators satisfying certain conditions of the monotonicity type in a Hilbert space (see Bruck [29], Rhoades [30] and Dunn [31]). Even in Hilbert space, how such conditions might follow from (9a) is evident only for the narrow class of positive definite quadratic functionals [31]. Nevertheless certain analytical tools developed for the treatment of (9b) also play a fundamental part in the analysis which follows.

LEMMA 1. *Let $\{\beta_n\} \subset [0, \infty)$ satisfy*

$$\beta_{n+1} \leq (1 - \omega_n) \beta_n + \omega_n \epsilon_n \quad (11)$$

with $\{\omega_n\} \subset (0, 1]$ and $\{\epsilon_n\} \subset [0, \infty)$, for $n \geq 0$. Suppose that $\{\omega_n\}$ and $\{\epsilon_n\}$ also satisfy the conditions:

$$n\omega_n \leq C \quad (12a)$$

$$(1 - \omega_{n+1}) = \frac{\omega_{n+1}}{\omega_n} \quad (12b)$$

for some $C > 0$ and for all n sufficiently large, and

$$\lim_{n \rightarrow \infty} \epsilon_n = 0. \quad (13)$$

Then $\lim_{n \rightarrow \infty} \beta_n = 0$.

Proof. Conditions (11) and (12), and a straightforward induction give

$$0 \leq \beta_{n+1} \leq \frac{C}{n} \left[\frac{\beta_1}{\omega_0} + \sum_{i=1}^n \epsilon_i \right]. \quad (14)$$

With reference to (13), choose M so large that

$$i \geq M \Rightarrow 0 \leq \epsilon_i \leq \epsilon. \quad (15)$$

Then (13) and (15) give

$$0 \leq \beta_{n+1} \leq \frac{C}{n} \left(\frac{\beta_1}{\omega_0} \right) + \frac{C}{n} \sum_{i=1}^{M-1} \epsilon_i + \frac{C(n-M)\epsilon}{n} \quad (16)$$

and therefore

$$0 \leq \lim_{n \rightarrow \infty} \beta_{n+1} \leq \overline{\lim}_{n \rightarrow \infty} \beta_{n+1} \leq C\epsilon. \quad (17)$$

Since ϵ can be arbitrarily small, this means that $\lim \beta_n = 0$.

Q.E.D.

Note 1. The class of sequences $\{\omega_n\} \subset (0, 1]$ defined by (12) has $\{1/(n+1)\}$ as its prototype. To see this, write (12b) as follows:

$$\omega_{n+1} \geq \frac{\omega_n}{1 + \omega_n}$$

for $n \geq M$; a simple induction then yields

$$\omega_n \geq \frac{\omega_M}{1 + (n-M)\omega_M}$$

for $n \geq M$. Thus, every sequence satisfying (12) is bounded above by $\{C/n\}$ for some constant C , and is bounded below by a sequence which is asymptotically like $1/n$ as $n \rightarrow \infty$. Bruck [29] analyzes the convergence of (9b) for bounded set valued monotone Hilbert space operators and for $\omega_n = 1/(n+1)$. Rhoades [30] uses (12b) in his generalization of Bruck's result.

THEOREM 1. *Let Ω be a nonempty closed bounded convex subset of a real Banach space X , and let $f: X \rightarrow R^1$ be bounded below and convex, with a uniformly continuous Frechet derivative f' . Furthermore, let $\{x_n\} \subset \Omega$ and $\{\omega_n\} \subset (0, 1]$ satisfy (9) and (12). Then $\lim_{n \rightarrow \infty} f(x_n) = \inf_{\Omega} f > -\infty$.*

Proof. Since f is continuously differentiable and Ω is convex, the mean value theorem gives

$$\begin{aligned} f(x_{n+1}) &= f(x_n) + \langle f'(y_n), x_{n+1} - x_n \rangle \\ &= f(x_n) + \langle f'(x_n), \bar{x}_n - x_n \rangle \omega_n + e(x_n, \omega_n) \omega_n, \end{aligned} \quad (18a)$$

where

$$e(x_n, \omega_n) = \langle f'(y_n) - f'(x_n), \bar{x}_n - x_n \rangle \quad (18b)$$

and $y_n \in \Omega$, with

$$\begin{aligned} y_n &= (1 - \sigma) x_n + \sigma x_{n+1} \\ &= x_n + \sigma \omega_n (\bar{x}_n - x_n), \end{aligned} \quad (18c)$$

for some $\sigma \in [0, 1]$. Put $r_n = f(x_n) - \inf_{\Omega} f$; then (18) gives

$$0 \leq r_{n+1} \leq r_n + \langle f'(x_n), \bar{x}_n - x_n \rangle \omega_n + |e(x_n, \omega_n)| \omega_n. \quad (19)$$

Given $\epsilon > 0$, choose $\eta \in \Omega$ so that $f(\eta) \leq \inf_{\Omega} f + \epsilon$. From (9a) and the convexity of f , one then obtains,

$$\begin{aligned} f(\eta) &\geq f(x_n) + \langle f'(x_n), \eta - x_n \rangle \\ &\geq f(x_n) + \langle f'(x_n), \bar{x}_n - x_n \rangle \end{aligned}$$

and therefore,

$$\langle f'(x_n), \bar{x}_n - x_n \rangle \leq -r_n + \epsilon. \quad (20)$$

Since $\epsilon > 0$ is arbitrarily small, (19) and (20) give

$$0 \leq r_{n+1} \leq (1 - \omega_n) r_n + |e(x_n, \omega_n)| \omega_n. \quad (21a)$$

Furthermore, (18b) and (18c) give

$$\begin{aligned} 0 \leq |e(x_n, \omega_n)| &\leq \|f'(y_n) - f'(x_n)\|_* \cdot \|\bar{x}_n - x_n\| \\ &\leq \|f'(y_n) - f'(x_n)\|_* \cdot d \end{aligned} \quad (21b)$$

and

$$\|y_n - x_n\| \leq \omega_n d, \quad (21c)$$

with $d = \text{diam } \Omega < \infty$, therefore since f' is uniformly continuous on Ω and since $\omega_n \rightarrow 0$, it follows that

$$\lim_{n \rightarrow \infty} |e(x_n, \omega_n)| = 0. \quad (22)$$

Finally (21), (22), and Lemma 1 given $\lim_{n \rightarrow \infty} f(x_n) = \inf_{\Omega} f$. Q.E.D.

Note 2. The process (9) can be *formally* modified to deal with f 's which are merely "differentiable with respect to directions", as in [1] (every convex functional has this property). Moreover, in some cases at least, the resulting algorithm will actually produce minimizing sequences. For instance, let $\Omega = [-1, 1] \subset \mathbb{R}^1$ and $f(x) = |x|$; then f has one sided derivatives at 0 but f' does not exist there. For this f and Ω , the appropriate operator T for (9b) is given by $T(x) = \text{sgn}(x)$ for $x \neq 0$, and $T(x) = \{-1, 1\}$ for $x = 0$. It can be shown that $x_n \rightarrow 0$ for every sequence $\{x_n\}$ generated by (9b) with this T and any $\{\omega_n\} \subset (0, 1]$ satisfying (4), however this can *not* be established by a general argument of the kind used in the proof of Theorem 1.

THEOREM 2. *Let f and Ω satisfy the hypotheses of Theorem 1, and in addition, let f' be Lipschitz continuous, ie., $\|f'(x) - f'(y)\|_* \leq L \|x - y\|$ for some $L > 0$ and all $x, y \in \Omega$. Furthermore, let $\{x_n\} \subset \Omega$ and $\{\omega_n\} \subset (0, 1]$ satisfy (9). Put $r_n = f(x_n) - \inf_{\Omega} f$ and $d = \text{diam } \Omega < \infty$. Then*

$$0 \leq r_n \leq B\beta_n, \quad (23a)$$

for $n \geq 1$, with

$$B = \max\{r_1, \frac{1}{2}Ld^2\} \quad (23b)$$

and with $\{\beta_n\}$ generated by

$$\beta_{n+1} = (1 - \omega_n)\beta_n + \omega_n^2; \quad \beta_1 = 1. \quad (23c)$$

If $\{\omega_n\}$ also satisfies (5), then $\lim_{n \rightarrow \infty} \beta_n = 0$ and therefore $\lim_{n \rightarrow \infty} f(x_n) = \inf_{\Omega} f$.

Proof. From Lemma 1.2, p. 117 in [1], one obtains:

$$f(x_{n+1}) \leq f(x_n) + \langle f'(x_n), \bar{x}_n - x_n \rangle \omega_n + \frac{1}{2}L \|\bar{x}_n - x_n\|^2 \omega_n^2 \quad (24)$$

(a somewhat coarser inequality can be gotten from (18a) and (21b)). Therefore,

$$0 \leq r_{n+1} \leq r_n + \langle f'(x_n), \bar{x}_n - x_n \rangle \omega_n + \frac{1}{2}L d^2 \omega_n^2. \quad (25)$$

Given $\epsilon > 0$, choose $\eta \in \Omega$ so that $f(\eta) \leq \inf_{\Omega} f + \epsilon$. Then (10a) and the convexity of f give (20), which together with (25) yields

$$0 \leq r_{n+1} \leq (1 - \omega_n)r_n + \frac{1}{2}L d^2 \omega_n^2. \quad (26)$$

This result and a simple induction produces (23). Finally if $\{\omega_n\} \subset (0, 1]$ satisfies (5), then $\lim_{n \rightarrow \infty} \beta_n = 0$ is immediate from Lemma 1 in [31]. Q.E.D.

Note 3. Condition (12) implies (5) (see Note 1) however (5) does not imply (12), e.g., for $1 > \nu > \frac{1}{2}$, $\{n^{-\nu}\}$ satisfies (5) but not (12).

THEOREM 3. *Let f and Ω satisfy the hypotheses of Theorem 2. Furthermore, let $\{\omega_n\}$ denote the sequence generated by (6) and let $\{x_n\} \subset \Omega$ and $\{\beta_n\}$ satisfy (9) and*

(23c), respectively, with $\{\omega_n\}$. Put $r_n = f(x_n) - \inf_{\Omega} f$ and $d = \text{diam } \Omega$. Then $\{\omega_n\}$ and $\{\beta_n\}$ have range in $(0, 1]$ and decrease monotonically to 0, with

$$0 < \omega_n \leq \frac{2}{n}, \quad (27)$$

$$0 < \beta_n \leq \frac{4}{n}, \quad (28)$$

$$0 \leq r_n \leq \frac{1}{2}L d^2 \beta_n, \quad (29)$$

for $n \geq 1$, and with $\omega_n \sim 2/n$ and $\beta_n \sim 4/n$ as $n \rightarrow \infty$. Moreover, if $\{\hat{\omega}_n\}$ is any other sequence with range in $(0, 1]$ and if $\{\hat{\beta}_n\}$ is the corresponding sequence in (23c), then

$$0 < \beta_n \leq \hat{\beta}_n. \quad (30)$$

Proof. The quadratic function $g(\omega) = (1 - \omega)\beta_n + \omega_n^2$ attains its minimum at $\omega = \beta_n/2$. For each $n \geq 1$, put $\omega_n = \beta_n/2$ in (23c) to obtain

$$\beta_{n+1} = \beta_n - (\beta_n)^2/4; \quad \beta_1 = 1 \quad (31)$$

for $n \geq 1$, and therefore

$$\omega_{n+1} = \omega_n - \omega_n^2/2; \quad \omega_0 = 1$$

for $n \geq 0$, as in (6). By construction, the solution $\{\beta_n\}$ of (31) satisfies (23c) with the sequence $\{\omega_n\}$ in (6). Both sequences are clearly monotone nonincreasing with range in $(0, 1]$. The inequalities (27) and (28), and the asymptotic properties of ω_n and β_n are immediate from (6), (31), and Lemma 2 of [31]. Since $\omega_0 = 1$, the inequality (26) gives $0 \leq r_1 \leq \frac{1}{2}L d^2$, consequently $B = \frac{1}{2}L d^2$ in (23b), and therefore (23a) reduces to (29). Finally, since ω_n minimizes $g(\omega)$, one has

$$\begin{aligned} \beta_{n+1} &= (1 - \omega_n)\beta_n + \omega_n^2 \\ &\leq (1 - \hat{\omega}_n)\hat{\beta}_n + \hat{\omega}_n^2 \end{aligned}$$

for any $\hat{\omega}_n$. In particular, if $\hat{\omega}_n \in [0, 1]$ and if $0 \leq \beta_n \leq \hat{\beta}_n$, this can be carried further to

$$\beta_{n+1} \leq (1 - \hat{\omega}_n)\hat{\beta}_n + \hat{\omega}_n^2 = \hat{\beta}_{n+1}.$$

Since $\beta_1 = \hat{\beta}_1$, the inequality (30) now follows for all $n \geq 1$, by induction.

Q.E.D.

Note 4. The closed loop step size rule of Demyanov and Rubinov, i.e.,

$$\omega_n = \min \left\{ 1, \frac{\langle f'(x_n), x_n - \bar{x}_n \rangle}{L \|x_n - \bar{x}_n\|^2} \right\} \quad (32)$$

may be derived by minimizing the bound on r_{n+1} given by the right side of (25); compare this with the way ω_n is obtained in the proof of Theorem 3, viz., by minimizing the parameter β_{n+1} appearing in the sequence of bounds on r_n generated by (23). It is shown in [1] that $f(x_n) - \inf_{\Omega} f = O(1/n)$ is always obtained with (32) when f is convex, and that the convergence is actually linear for certain uniformly convex Ω 's, provided $\|f'(x)\|_*$ is bounded away from 0 on Ω . On the other hand, since (32) is equivalent to line minimization for certain quadratic f 's, a result of Cannon and Cullum [32] shows that $O(1/n)$ convergence is sometimes the *best* one can obtain with (32) when Ω is not uniformly convex. In these cases, one does just as well with the simple open loop rule (6). Moreover (6) can be implemented regardless of whether a value for L is known.

Note 5. The sequence $\{\omega_n\} \subset (0, 1]$ generated by (6) satisfies the condition (12) involved in Theorem 1, e.g., (12a) is immediate from (27), and (12b) follows from the fact that $(1 + \omega_n) \omega_{n+1} = (1 + \omega_n) (1 - \frac{1}{2}\omega_n) \omega_n = [1 + \frac{1}{2}\omega_n(1 - \omega_n)] \omega_n \geq \omega_n$. Thus $f(x_n) \rightarrow \inf_{\Omega} f$ is assured for (6) if f' is uniformly continuous (and possibly under more general circumstances as well; see Note 2).

Note 6. If the vector \bar{x}_n in (9b) is determined by (10) instead of (9a), then inequality (21a) in the proof of Theorem 1 is replaced by

$$0 \leq r_{n+1} \leq (1 - \omega_n) r_n + (|e(x_n, \omega_n)| + \delta_n) \omega_n$$

however since $\delta_n > 0$ and $\lim_{n \rightarrow \infty} \delta_n = 0$, the remainder of the proof goes through as before. Similarly, if \bar{x}_n is obtained from (10) with $\delta_n = \delta \omega_n$, then inequality (26) is replaced by

$$0 \leq r_{n+1} \leq (1 - \omega_n) r_n + (\frac{1}{2}L d^2 + \delta) \omega_n^2$$

and Theorems 2 and 3 survive with $\frac{1}{2}L d^2$ changed to $\frac{1}{2}L d^2 + \delta$ in (23b) and (29). This shows how one may approximate the infimum of the *convex* functional f with arbitrarily small error if one knows how to approximate the infima of the *linear* functionals $f'(x)$ with arbitrarily small error (regardless of whether any of the infima in question are actually attained on Ω).

3. CONVERGENCE OF THE ITERATES x_n

If the set Ω_f of minimizers for f is not empty, it seems reasonable that the minimizing sequence property, $f(x_n) \rightarrow \inf_{\Omega} f$, should force $\{x_n\}$ to converge in some sense to Ω_f , irrespective of the method used to generate $\{x_n\}$. Actually,

something like this does happen if f is lower semicontinuous with respect to a given topology on $\Omega \subset X$, since the x_n 's are then contained in a nested sequence of neighborhoods, N_n , of Ω_f converging downward on Ω_f (e.g., let $\epsilon_k \rightarrow 0$ monotonically from above, let $L(\epsilon_k) = \{x \in \Omega \mid f(x) \leq \inf f + \epsilon_k\}$, let n_k be a strictly increasing integer sequence for which $n \geq n_k \Rightarrow x_n \in L(\epsilon_k)$, and put $N_n = L(\epsilon_k)$ for $n_k \leq n < n_{k+1}$; then $x_n \in N_n$ for all $n \geq n_1$, and $N_n \downarrow \Omega_f$ monotonically). Under these circumstances, every cluster point of a minimizing sequence must lie in Ω_f . Moreover, if Ω is compact with respect to the topology in question, then Ω_f is not empty and the cluster point inclusion property implies that every neighborhood of Ω_f must contain all but finitely many members of a minimizing sequence $\{x_n\}$, i.e., $\{x_n\}$ must converge to Ω_f in the given topology. This result has several simple but interesting consequences for the present development, since the convex functionals f in Section 2 are continuous in the norm topology and lower semicontinuous in the weak topology. Thus, when Ω is strongly compact, Ω_f is nonempty and the minimizing sequences $\{x_n\}$ of Section 2 must converge to Ω_f in the sense that $\lim_{n \rightarrow \infty} \{\inf_{y \in \Omega_f} \|x_n - y\|\} = 0$. When Ω is weakly compact, Ω_f is again nonempty and $\{x_n\}$ converges to Ω_f in the weak topology. In particular, if Ω_f has a single element, ξ , then $\{x_n\}$ converges to ξ strongly in the first case and weakly in the second case. Under certain additional conditions of the uniform convexity type on f or Ω , a minimizing sequence $\{x_n\}$ may actually converge strongly to a unique minimizer ξ even though Ω is only weakly compact [1].

The foregoing observations do not rest on the particular structure of the conditional gradient method (9) (or (9b)-(10)), i.e., they apply to *any* minimizing sequence for f . What more can be said about sequences $\{x_n\}$ produced by (9)? In the case of certain smooth Hilbert space functionals, the iterate sequences generated by either the classical or projected gradient methods with step size constraints of the type (4) or (5), are known to converge at least weakly to some $\xi \in \Omega_f$ as long as $\Omega_f \neq \emptyset$ [25] [27], however simple counter-examples in R^2 show that this need not happen for conditional gradient iterates when Ω_f has more than one element, because of the multivalued character of the operator T in (9a). For certain positive semidefinite quadratic functionals with multiple minima, it can be shown that $\lim_{n \rightarrow \infty} \{\inf_{y \in \Omega_f} \|x_n - y\|\} = 0$ when $\{x_n\}$ is generated by (9) with the step size constraints (5), even though Ω_f is only weakly compact; how far this conclusion extends into the class of nonquadratic f 's is not known at present. Finally, when ξ is a unique minimizer of f , there is reason to believe that the asymptotic behavior of iterate sequences generated by both open and closed loop conditional gradient methods varies significantly according to whether the operator T in (9a) is single-valued or multivalued at ξ (with the former circumstance being the more favorable). On an abstract level, there are certain interesting connections here with the notions of nonsingular and singular extremals proposed by Dunn in [33] for optimal control problems. These and other related points will be treated at length elsewhere.

REFERENCES

1. V. F. DEMYANOV AND A. M. RUBINOV, "Approximate Methods in Optimization Problems," American Elsevier, New York, 1970.
2. L. V. KANTOROVICH, Functional analysis and applied mathematics, *Uspehi Mat. Nauk* **3** (1948), 89–185.
3. M. FRANK AND P. WOLFE, An algorithm for quadratic programming, *Nav. Res. Logist. Quart.* **3** (1956), 95–110.
4. H. J. KELLEY, Method of gradients, in "Optimization Techniques," Chap. 6, Academic Press, New York, 1962.
5. E. G. GILBERT, An iterative procedure for computing the minimum of a quadratic form on a convex set, *SIAM J. Control*, **A 4** (1966), 61–80.
6. R. O. BARR AND E. G. GILBERT, Some iterative procedures for computing optimal controls, in "Proceedings of the 3rd IFAC Congress, Butterworth, London, 1966.
7. R. O. BARR, An efficient computational procedure for a generalized quadratic programming problem, *SIAM J. Control* **7** (1969), 415–429.
8. E. R. BARNES, A geometrically convergent algorithm for solving optimal control problems, *SIAM J. Control* **10** (1972), 434–443.
9. G. G. L. MEYER, Accelerated Frank-Wolfe algorithms, *SIAM J. Control* **12**, 4 (1974), 655–663.
10. T. PECSVARADI AND K. S. NARENDRA, A new iterative procedure for the minimization of a quadratic form on a convex set, *SIAM J. Control* **8** (1970), 396–402.
11. J. C. DUNN, A simple averaging process for approximating the solutions of certain optimal control problems, *J. Math. Anal. Appl.* **48** (1974).
12. J. C. DUNN AND V. KUMAR, An averaging technique for solving extremal boundary value problems, in "Recent Advances in Engineering Science," Vol. 8, Scientific Publishers, Boston, 1975.
13. V. KUMAR, A control averaging technique for solving a class of singular optimal control problems, *Int. J. Control* **23**, 3 (1976), 361–380.
14. J. ROBINSON, An iterative method of solving a game, *Annal. Math.* **54** (1956), 296–301.
15. H. ROBBINS AND S. MONRO, A stochastic approximation method, *Ann. Math. Stat.* **22** (1951), 400–407.
16. W. R. MANN, Mean value methods in iteration, *Proc. Amer. Math. Soc.* **4** (1953), 506–510.
17. E. POLAK, An historical survey of computational methods in optimal control, *SIAM Rev.* **15** (1973), 553–584.
18. E. BELTRAMI, "An Algorithmic Approach to Nonlinear Analysis and Optimization," Academic Press, New York, 1970.
19. D. G. LUENBERGER, "Optimization by Vector Space Methods," Academic Press, New York, 1962.
20. A. A. GOLDSTEIN, On steepest descent, *SIAM J. Control* **3** (1965), 147–151.
21. L. ARMIJO, Minimization of functions having continuous partial derivatives, *Pacific J. Math.* **16** (1966), 1–3.
22. A. A. GOLDSTEIN, Minimizing functionals on Hilbert space, in "Computer Methods in Optimization Problems," pp. 159–165, Academic Press, New York, 1964.
23. B. T. POLYAK, Gradient methods for the minimization of functionals, *Zh. Vychisl. Mat. i Mat. Fiz.* **3** (1963), 643–653.
24. M. M. VAINBERG, On the convergence of the method of steepest descents for nonlinear equations, *Soviet Math. Dokl.* **1** (1960), 1–4.
25. J. C. DUNN, Convexity, monotonicity, and gradient processes in Hilbert space, *J. Math. Anal. Appl.* **53** (1976).

26. YU. M. ERMOL'EV, Methods of solution of nonlinear extremal problems, *Cybernetics* 2 (1966), 1-14.
27. R. E. BRUCK, JR., An iterative solution of a variational inequality for certain monotone operators in Hilbert space, *Bull. Amer. Math. Soc.* 81 (1975), 890-892.
28. R. T. ROCKAFELLAR, "Convex Analysis," Princeton University Press, Princeton, New Jersey, 1970.
29. R. E. BRUCK, JR., The iterative solution of the equation $y \in x + Tx$ for a monotone operator T in Hilbert space, *Bull. Amer. Math. Soc.* 79 (1973), 1258-1261.
30. B. E. RHOADES, Fixed point iterations using infinite matrices: part III, in "Proc. Conference on Computing Fixed Points, With Applications," Clemson University, June 1974, Academic Press, New York.
31. J. C. DUNN, Iterative construction of fixed points for multivalued operators of the monotone type, *J. Functional Anal.*, in press.
32. M. D. CANNON AND C. D. CULLUM, A tight upper bound on the rate of convergence of the Frank-Wolfe algorithm, *SIAM J. Control* 6 (1968), 509-516.
33. J. C. DUNN, On the classification of singular and nonsingular extremals for the Pontryagin maximum principle, *J. Math. Anal. Appl.* 17 (1967), 1-36.