58

COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN
1210 West Dayton Street
Madison, Wisconsin 53706

CONVERGENCE OF SOME GRADIENT-LIKE METHODS FOR CONSTRAINED MINIMIZATION

by

James W. Daniel

Computer Sciences Technical Report #65

July 1969

Prepared under Contract Number N00014-67-A-0128-0004 at the University of Wisconsin. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. Introduction

For the problem of iteratively minimizing a nonlinear functional f over a real Hilbert space $\mathcal{H}$ by gradient-type methods a rather thorough analysis of the direction problem and the step-length problem has been given is some generality by a number of authors [1, 2, 5, 7, 15, 16, 17, 20]; to a somewhat lesser extent this has been accomplished for constrained problems [1, 2, 5, 6, 15, 20, 21, 25]. We shall take some steps in that direction by analyzing two basic step-length algorithms for general (feasible) directions and indicating the applicability of these results to the conditional gradient and variable metric projected gradient methods.

We wish to minimize a real valued Frechet differentiable functional f over a (often convex) subset C of a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot , \cdot \rangle$ by an iterative method generating a sequence $\{x_n\}$, taken to lie in C although more generality is possible [5, 20, 21, 22]. We shall generally think of the sequence being generated by moving certain step-lengths along directions p_n which point into C and along which f is non-increasing.

Definition 1.1 A direction sequence $p_n \equiv p_n(x_n)$ will be called feasible (for the points x_n in C) if and only if $p_n = x'_n - x_n$ where $\lambda x'_n + (1-\lambda) x_n$ is in C for λ in $[0, 1]$ and $\langle p_n, \nabla f(x_n) \rangle \leq 0$.

We shall prove that, under certain methods of choosing the step-length "along" p_n, we have $\langle p_n, \nabla f(x_n) \rangle$ converging to zero; we shall then show how this condition can be made useful.

2. Step-length using Lipschitz continuity.

Theorem 2.1: Let f be bounded below on C, ∇f satisfy $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$ for x, y in C, and $p_n = p_n(x_n)$ be feasible directions. Pick $\delta_1, \delta_2, \delta_3$ all greater than zero and let γ_n lie in $[\min(\delta_1, \frac{\delta_2\|p_n\|^2}{-<\nabla f(x_n), p_n>}), \frac{2}{L} - \delta_3]$ for all n. For each n let $x''_n = x_n + t_n p_n$ where t_n is defined via

$$t_n = \min\left(1, \frac{\gamma_n <-\nabla f(x_n), p_n>}{\|p_n\|^2}\right)$$

and let x_{n+1} in C satisfy $f(x_{n+1}) \leq \beta f(x''_n) + (1-\beta)f(x_n)$ for a fixed β in $(0, 1]$. Then $f(x_n)$ decreases to a limit. If $\|p_n\|$ is uniformly bounded, for example if C is bounded, then $\lim_{n\to\infty} < \nabla f(x_n), p_n> = 0$. If $\|p_n\| \to 0$ implies $< \nabla f(x_n), \frac{p_n}{\|p_n\|} > \to 0$, then $\lim_{n\to\infty} < \nabla f(x_n), \frac{p_n}{\|p_n\|} > = 0$.

Proof:

$$\begin{aligned} \frac{1}{\beta}[f(x_{n+k}) - f(x_n)] &\leq f(x''_n) - f(x_n) \leq \\ &\leq < \nabla f(x_n), x''_n - x_n> + \int_0^1 < \nabla f(x_n + \lambda t_n p_n) - \nabla f(x_n), t_n p_n> d\lambda \\ &\leq < \nabla f(x_n), x''_n - x_n> + \frac{L}{2} t_n^2 \|p_n\|^2 \\ &\leq - t_n < - \nabla f(x_n), p_n> + \frac{L}{2} t_n^2 \|p_n\|^2 . \end{aligned}$$

If $1 \leq \gamma_n \frac{< - \nabla f(x_n), p_n>}{\|p_n\|^2}$ then $t_n = 1$, x''_n is in C, and

$$\frac{1}{\beta}[f(x_{n+1})-f(x_n)] \le \langle -\nabla f(x_n), p_n\rangle \left[-1+\frac{L}{2}\ \frac{\|p_n\|^2}{\langle -\nabla f(x_n), p_n\rangle}\right]$$

$$\le \langle -\nabla f(x_n), p_n\rangle \left[-1+\frac{L\gamma_n}{2}\right] \le \frac{\delta_3 L}{2}\ \langle -\nabla f(x_n), p_n\rangle \le 0\ .$$

If however $1 > t_n = \gamma_n \dfrac{\langle -\nabla f(x_n), p_n\rangle}{\|p_n\|^2}$, then x''_n is in C and

$$\frac{1}{\beta}[f(x_{n+1})-f(x_n)] \le \gamma_n \frac{\langle -\nabla f(x_n), p_n\rangle^2}{\|p_n\|^2} + \frac{L}{2}\|p_n\|^2\ \gamma_n^2\ \frac{\langle -\nabla f(x_n), p_n\rangle^2}{\|p_n\|^4}$$

$$\le \frac{\langle -\nabla f(x_n), p_n\rangle^2}{\|p_n\|^2}\left[\frac{\gamma_n^2 L}{2} - \gamma_n\right] \le \text{ either}$$

$$-\frac{\delta_1\delta_3 L}{2}\ \frac{\langle -\nabla f(x_n), p_n\rangle^2}{\|p_n\|^2} \quad \text{or} \quad \frac{-\delta_2\delta_3 L}{2}\ \langle -\nabla f(x_n), p_n\rangle\ .$$

In either case $\frac{1}{\beta}[f(x_{n+1}) - f(x_n)] \le 0$ and $f(x_n)$ decreases to a limit.

If $\|p_n\| = \|x'_n - x_n\|$ is bounded, then from the three inequalities bounding the decrease in f we obtain a $\delta > 0$ such that

$$f(x_n) - f(x_{n+1}) \ge \delta \langle -\nabla f(x_n), p_n\rangle^r\ ,$$

for $r = 1$ or $r = 2$, which implies $\lim_{n\to\infty} \langle -\nabla f(x_n), p_n\rangle = 0$. Since

$\left\langle \nabla f(x_n) \dfrac{p_n}{\|p_n\|}\right\rangle = \dfrac{\langle \nabla f(x_n), p_n\rangle}{\|p_n\|}$, the final conclusion also follows.

Q.E.D.

<u>Remark</u> If in particular one chooses x_{n+1} so as to minimize $f(x_n + t\,p_n)$ for t with $x_n + t\,p_n$ in C, then certainly $f(x_{n+1}) \le f(x''_n)$ and the theorem applies without <u>explicit</u> use of the Lipschitz constant;

this is also true of course if x_{n+1} minimizes f over some simplex such as that generated by $x_{n-k}, x_{n-k+1}, \ldots, x_n+p_n$. More generally one need only reduce f to nearly the value $f(x_n'')$; thus if x_n' is itself computed by some method it need not be computed exactly. Note that the step-size choice in this theorem has been well analyzed for unconstrained problems and partially so for constrained problems [1, 2, 5, 7, 20, 21] .

3. Step-length using a range function.

The burden of the proof of Theorem 2.1 is to show that f "decreases enough," that is, in such a way as to force $\langle \nabla f(x_n), p_n \rangle$ to zero. As our second step-length algorithm we discuss a method developed for unconstrained problems which attains this sufficient decrease more directly [7, 15, 16, 17] .

Definition 3.1 A real valued function d defined on $[0, \infty)$ is called a forcing function if and only if $d(t) \geq 0$ whenever $t \geq 0$ and $\lim_{n\to\infty} d(t_n) = 0$ if and only if $\lim_{n\to\infty} t_n = 0$.

We shall determine admissible values of t_n in terms of a so-called range function

$$g(x, t, p) \equiv \frac{f(x) - f(x+tp)}{-t \langle \nabla f(x), p \rangle}$$

which is continuous at $t = 0$ if we define $g(x, 0, p) \equiv 1$. Given a feasible sequence $p_n \equiv p_n(x_n)$ satisfying, for the moment, $\|p_n\| = 1$, a real number δ in $(0, \frac{1}{2}]$ and a forcing function d with $d(t) \leq \delta t$, we move from x_n to

x_{n+1} as follows. If, for $t_n = 1$ and $x_n' = x_n + p_n$ we find

$$g(x_n, t_n, p_n) \geq \frac{d(<-\nabla f(x_n), p_n>)}{<-\nabla f(x_n), p_n>} \tag{3.1}$$

we set $x_n'' = x_n'$; otherwise find t_n in $(0, 1)$ satisfying Equation 3.1 and also

$$|g(x_n, t_n, p_n) - 1| \geq \frac{d(<-\nabla f(x_n), p_n>)}{<-\nabla f(x_n), p_n>} \tag{3.2}$$

Finally x_{n+1} is any point in C with $f(x_{n+1}) \leq \beta f(x_n'') + (1-\beta) f(x_n)$ for a fixed β in $(0, 1]$. We observe that the algorithm is well defined. Since $g(x_n, 0, p_n) = 1$ and $1 - \frac{d(t)}{t} \geq \frac{d(t)}{t}$ for all t, if we have $g(x_n, 1, p_n) < \frac{d(z)}{z}$ where $z = <-\nabla f(x_n), p_n>$, then by the continuity of $g(x_n, t, p_n)$ in t and the fact that $x_n + tp_n$ is in C for t in $[0, 1]$ since p_n is a feasible direction there exists t_n in $(0, 1)$ with $\frac{d(z)}{z} \leq g(x_n, t_n, p_n) \leq 1 - \frac{d(z)}{z}$ which certainly satisfied Equations 3.1 and 3.2. We now prove the convergence of this method, following the proof for the unconstrained case [7].

Theorem 3.1: Let f be bounded below on C, ∇f be uniformly continuous on C, and $p_n \equiv p_n(x_n)$ be a feasible direction sequence with $\|p_n\| = 1$. Let d be a forcing function with $d(t) \leq \delta t$ for δ in $(0, \frac{1}{2}]$. Let the algorithm described above be applied. Then $\lim_{n\to\infty} <\nabla f(x_n), p_n> = 0$.

Proof: Define the reverse modulus of continuity [7]

$$s(t) = \inf \{\|x-y\|;\ \|\nabla f(x) - \nabla f(y)\| \geq t,\ x, y \text{ in } C\}.$$

By the uniform continuity of ∇f on C, s is a monotonic decreasing forcing function. By Equation 3.1, $\{f(x_n)\}$ is decreasing and

$$\frac{1}{\beta}[f(x_n) - f(x_{n+1})] \geq f(x_n) - f(x''_n) \geq t_n d(\langle -\nabla f(x_n), p_n \rangle) . \qquad (3.3)$$

If infinitely often $\langle -\nabla f(x_n), p_n \rangle \geq \varepsilon > 0$, we cannot have $t_n = 1$ infinitely often for then, by Equation 3.3, $f(x_n)$ is not bounded below. Thus it must be that $t_n = 1$ does not satisfy Equation 3.1 and hence t_n is in $(0, 1)$. For these n, we write

$$f(x''_n) - f(x_n) = \langle \nabla f(x_n + \lambda_n t_n p_n), t_n p_n \rangle \text{ for some } \lambda_n \text{ in } (0,1).$$

Thus, from Equation 3.2,

$$\frac{d(\langle -\nabla f(x_n), p_n \rangle)}{\langle -\nabla f(x_n), p_n \rangle} \leq |g(x_n, t_n, p_n) - 1|$$

$$\leq \left| \frac{\langle \nabla f(x_n + \lambda_n t_n p_n) - \nabla f(x_n), p_n \rangle}{\langle \nabla f(x_n), p_n \rangle} \right|$$

$$\leq \frac{\| \nabla f(x_n + \lambda_n t_n p_n) - \nabla f(x_n) \|}{\langle -\nabla f(x_n), p_n \rangle} .$$

Therefore $\| \nabla f(x_n + \lambda_n t_n p_n) - \nabla f(x_n) \| \geq d(\langle -\nabla f(x_n), p_n \rangle)$ and hence

$$t_n = \| x_{n+1} - x_n \| \geq \| \lambda_n t_n p_n \| \geq s(\| \nabla f(x_n + \lambda_n t_n p_n) - \nabla f(x_n) \|)$$

$$\geq s(d(\langle -\nabla f(x_n), p_n \rangle)) \qquad (3.4)$$

Hence, using Equation 3.3, we conclude that

$$\frac{1}{\beta}[f(x_n) - f(x_{n+1})] \geq f(x_n)-f(x''_n) \geq d(<-\nabla f(x_n), p_n>)s(d(<-\nabla f(x_n), p_n>))$$

which implies that $\lim_{n\to\infty} <\nabla f(x_n), p_n> = 0$.

Q.E.D.

For problems in which C is not the whole space $\mathcal{H}$, that is, in which there are constraints, the restriction $\|p_n\| = 1$ is unrealistic; the following corollary shows that is is not needed so long as p_n cannot be "too small" compared to how "near" one is to a solution.

<u>Corollary 3.1</u> Under the hypotheses of Theorem 3.1 above with the assumption $\|p_n\| = 1$ replaced by

1) $\|p_n\| \geq d_1(<-\nabla f(x_n), \frac{p_n}{\|p_n\|}>)$ for a forcing function d_1,

2) $<-\nabla f(x_n), \frac{p_n}{\|p_n\|}> \to 0$ whenever $\frac{d(<-\nabla f(x_n), p_n>)}{\|p_n\|} \to 0$,

it follows that $\lim_{n\to\infty} <\nabla f(x_n), \frac{p_n}{\|p_n\|}> = 0$.

<u>Proof:</u> Under these hypotheses Equation 3.3 for $t_n = 1$ becomes

$$\frac{1}{\beta}[f(x_n)-f(x_{n+1})] \geq f(x_n)-f(x''_n) \geq d(<-\nabla f(x_n), p_n>) = d(<-\nabla f(x_n), \frac{p_n}{\|p_n\|}> \|p_n\|)$$

so that either $<-\nabla f(x_n), \frac{p_n}{\|p_n\|}>$ or $\|p_n\| \geq d_1(<-\nabla f(x_n), \frac{p_n}{\|p_n\|}>)$ tends to zero. On the other hand, for t_n in $(0, 1)$, Equation 3.4 becomes

$$t_n \| p_n \| \geq s \left(\frac{d(\langle -\nabla f(x_n), p_n \rangle}{\| p_n \|} \right)$$

and thereby

$$\frac{1}{\beta} [f(x_n) - f(x_{n+1})] \geq s \left(\frac{d(\langle -\nabla f(x_n), p_n)}{\| p_n \|} \right) \frac{d(\langle -\nabla f(x_n), p_n \rangle)}{\| p_n \|}$$

which implies that $\dfrac{d(\langle -\nabla f(x_n), p_n \rangle)}{\| p_n \|}$ tends to zero; the conclusion follows by 2) above.

Q.E.D.

Corollary 3.2 Under the assumptions on f, p_n, and C in Theorem 3.1 and Corollary 3.1, if x_{n+1} is chosen such that $f(x_{n+1}) = \min\limits_{0 \leq t \leq 1} f(x_n + t p_n)$, the conclusions of that theorem and corollary are valid.

Remark. i) The assumption 2) in Corollary 3.1 is valid if for instance $\| p_n \|$ is bounded above or $d(t) = qt$ for some $q \neq 0$. ii) If $\| p_n \|$ is bounded then also $\lim\limits_{n \to \infty} \langle \nabla f(x_n), p_n \rangle = 0$.

The algorithm above is not computational in the sense that it may well be very difficult to locate a t_n in $(0, 1)$ satisfying Equations 3.1 and 3.2 when $t_n = 1$ does not satisfy Equation 3.1. A known algorithm [3, 5, 7] for handling this problem for unconstrained minimization fortunately carries over easily and yields a much more valuable computational scheme.

Theorem 3.2: Under the hypotheses of Theorem 3.1 and Corollary 3.1, t_n may be chosen as the first of the numbers $\alpha^0, \alpha^1, \alpha^2, \ldots$ satisfying Equation 3.1 for a fixed α in $(0, 1)$ and then

$$\lim_{n \to \infty} \langle \nabla f(x_n), \frac{p_n}{\|p_n\|} \rangle = 0 .$$

Proof: We note first that since α^j tends to zero a first such value α^j exists. As in the proofs of the previous theorem and corollary, the case $t_n = 1$ is easily handled to show $\lim_{n \to \infty} \langle \nabla f(x_n), \frac{p_n}{\|p_n\|} \rangle = 0$ for those values of n; we consider the case in which $t_n = \alpha^j$, $j \geq 1$. Thus we have $x''_n = x_n + \alpha^j p_n$; let $x'''_n \equiv x_n + \alpha^{j-1} p_n$. Since α^j is the first value satisfying Equation 3.1, we have

$$f(x_n) - f(x'''_n) < \alpha^{j-1} d(\langle -\nabla f(x_n), p_n \rangle)$$

$$f(x_n) - f(x''_n) \geq \alpha^j d(\langle -\nabla f(x_n), p_n \rangle) .$$

Therefore

$$f(x''_n) - f(x'''_n) < (1 - \alpha) \alpha^{j-1} d(\langle -\nabla f(x_n), p_n \rangle) .$$

We can write

$$f(x''_n) - f(x'''_n) = \langle \nabla f(\lambda_n x'''_n + (1 - \lambda_n) x''_n), x''_n - x'''_n \rangle$$

for some λ_n in (0, 1). This leads to

$$\langle -\nabla f(\lambda_n x'''_n + (1 - \lambda_n) x''_n), p_n \rangle < d(\langle -\nabla f(x_n), p_n \rangle).$$

Recalling that $d(t) \le \delta t$ for some δ in $(0, \frac{1}{2}]$, we write

$$\|p_n\| \ \|\nabla f(\lambda_n x_n''' + (1-\lambda_n)x_n'') - \nabla f(x_n)\| \ge \langle \nabla f(\lambda_n x_n''' + (1-\lambda_n)x_n'') - \nabla f(x_n), p_n \rangle$$

$$\ge \langle -\nabla f(x_n), p_n \rangle - \langle -\nabla f(\lambda_n x_n''' + (1-\lambda_n)x_n''), p_n \rangle$$

$$\ge \langle -\nabla f(x_n), p_n \rangle - d(\langle -\nabla f(x_n), p_n \rangle)$$

$$\ge (1-\delta) \langle -\nabla f(x_n), p_n \rangle .$$

Defining s as the reverse modulus of continuity of ∇f as in the proof of Theorem 3.1, we then have

$$\|x_n'' - x_n\| = \alpha \|x_n''' - x_n\| \ge \alpha \|\lambda_n x_n''' + (1-\lambda_n)x_n'' - x_n\|$$

$$\ge \alpha\, s(\, (1-\delta) \langle -\nabla f(x_n), \frac{p_n}{\|p_n\|} \rangle) .$$

From this and Equation 3.1 we have

$$\frac{1}{\beta}[f(x_n) - f(x_{n+1})] \ge f(x_n) - f(x_n'') \ge \alpha^j d(\langle -\nabla f(x_n), p_n \rangle)$$

$$\ge \|x_n'' - x_n\| \ \frac{d(\langle -\nabla f(x_n), p_n \rangle)}{\|p_n\|}$$

from which it follows that $\lim_{n\to\infty} \langle \nabla f(x_n), \frac{p_n}{\|p_n\|} \rangle = 0$.

Q.E.D.

Using the step-length algorithms above, we conclude always that $\lim_{n\to\infty} \langle \nabla f(x_n), p_n \rangle = 0$ or $\lim_{n\to\infty} \langle \nabla f(x_n), \frac{p_n}{\|p_n\|} \rangle = 0$; for these to be useful results, the condition (for example) "$\langle \nabla f, p \rangle = 0$ in the limit" should somehow be related to a necessary or sufficient condition for a minimizing point. In unconstrained problems for example, one can take $p_n = -\nabla f(x_n)$ in which case the limiting condition $\nabla f(x) = 0$ is a necessary optimality condition. It appears in fact that any reasonable optimality condition of the form $\langle \nabla f, p \rangle = 0$ can be used to generate direction sequences for which the above step-length algorithms are useful. We consider two well known methods.

4. Conditional gradients.

Suppose that C is a convex set. Then a well known necessary condition for x^* to minimize f over C, one that is sufficient if f is convex, is that $\langle x-x^*, \nabla f(x^*) \rangle \geq 0$ for all x in C, that is, every direction into C is a direction of increase for f. If one has a point x_n which does not satisfy this condition, then it is reasonable to seek the x'_n which most violates this condition and then take $p_n = x'_n - x_n$; this method is well-known [5, 6, 9, 10, 20, 25]. Thus we seek x'_n such that

$\langle \nabla f(x_n), x'_n - x_n \rangle \leq \inf_{x \text{ in } C} \langle \nabla f(x_n), x-x_n \rangle + \varepsilon_n$ for some positive ε_n tending to zero. If C is bounded we can always find x'_n; if C is bounded and norm closed as well as convex then we can take $\varepsilon_n = 0$ if desired, although this causes unnecessary computation.

Proposition 4.1: Let f be convex, bounded below on the bounded convex set C, and attain its minimum at some point x^* in C. Let x_n be a sequence in C such that $\langle \nabla f(x_n), p_n \rangle$ tends to zero, where $p_n \equiv x'_n - x_n$

and x'_n satisfies $<\nabla f(x_n), x'_n - x_n> \leq \inf_{x \text{ in } C} <\nabla f(x_n), x-x_n> + \varepsilon_n$ for a sequence of positive ε_n tending to zero. Then x_n is a minimizing sequence, that is, $\lim_{n \to \infty} f(x_n) = f(x^*)$.

Proof: We have, using the convexity of f,

$$0 \leq f(x_n) - f(x^*) \leq <\nabla f(x_n), x_n - x^*>$$

$$\leq <\nabla f(x_n), x_n - x'_n> + <\nabla f(x_n), x'_n - x_n> - <\nabla f(x_n), x^* - x_n>$$

$$\leq <-\nabla f(x_n), p_n> + \varepsilon_n$$

which tends to zero.

Q.E.D

Remark. It is a simple matter to add hypotheses to f or C which guarantee that any minimizing sequence must in fact converge to x^* [5, 20, 21, 22]. For example, this is true if f is a weakly lower semi-continuous uniformly quasi-convex functional [5, 20, 22].

Thus we clearly may apply the step-size algorithm of Theorem 2.1.

Corollary 4.1: Let f be convex, bounded below on the bounded convex set C, and attain its minimum over C at x^*. Let f satisfy $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$ for x, y in C, and for each x_n such that for some x in C $<\nabla f(x_n), x-x_n> < 0$ let x'_n satisfy $<\nabla f(x_n), x'_n - x_n> \leq \inf_{x \in C} <\nabla f(x_n), x - x_n> + \varepsilon_n$ for a sequence of positive ε_n converging to zero; set $p_n \equiv x'_n - x_n$. If the minimization algorithm of theorem 2.1 is then applied, $\{x_n\}$ is a minimizing sequence.

To use the algorithms of Theorem 3.1 and 3.2 we note that if $\|\nabla f\|$ is bounded on C then $\lim_{n \to \infty} \|p_n\| = 0$ implies

$\lim_{n \to \infty} \langle \nabla f(x_n), p_n \rangle = 0$ and hence $\{x_n\}$ is a minimizing sequence. If however $\|p_n\|$ is bounded away from zero then for those n the forcing function d_1 in Corollary 3.1 exists and we can argue as in that corollary. Thus we conclude

Corollary 4.2: Let f be convex, bounded below on the bounded convex set C, and attain its minimum over C at x^*; let ∇f be uniformly continuous and $\|\nabla f(x)\|$ be uniformly bounded on C. Then the step-size algorithms of Theorems 3.1 and 3.2 applied to the direction algorithm of Corollary 4.1 yields a minimizing sequence $\{x_n\}$ for f over C.

5. Projected gradients.

The steepest descent method for unconstrained problems, in which $p_n = -\nabla f(x_n)$, has been a popular method for many years, for some applications undeservedly. For constrained problems that direction need not point into the constraint set C so it is not directly applicable. Perhaps the most successful way of handling this has been to "project" the direction onto C; more precisely one proceeds in the direction $p_n = x'_n - x_n$ where x'_n is the orthogonal projection onto C of $x_n - \alpha_n \nabla f(x_n)$ for some scalar $\alpha_n > 0$. This is the well known gradient projection method [23, 24]. In view of the numerical evidence that so-called variable metric methods are much better than steepest descent for unconstrained problems [8] and the growing interest in such methods for constrained problems [11, 12, 13, 14], we consider an analogous variable metric projected gradient method. We suppose that $\{A_n\}$ is uniformly bounded, uniformly positive definite family of self-adjoint linear operators on $\mathcal{H}$, that is, that there are $m > 0$, $M < \infty$ such that $m\langle x, x\rangle \le \langle A_n x, x\rangle \le M\langle x, x\rangle$ for all x in $\mathcal{H}$. For each n, let x'_n be the projection, with respect to the variable metric $\langle \cdot, A_n \cdot\rangle$, of $x_n - \alpha_n A_n^{-1} \nabla f(x_n)$ onto C; that is, x'_n minimizes $\langle x - (x_n - \alpha_n A_n^{-1} \nabla f(x_n)), A_n[x - (x_n - \alpha_n A_n^{-1} \nabla f(x_n))]\rangle$ over x in C. If C is norm closed and convex a unique such $x'_n - \alpha_n A_n^{-1} \nabla f(x_n)$, we know that for all x in C we must have

$$\langle x - x'_n, A_n(x'_n - w_n)\rangle \ge 0 . \qquad (5.1)$$

If we set $x \equiv x_n$ in this inequality, we obtain

$$0 \ge \langle x_n - x'_n, A_n(w_n - x'_n)\rangle = \langle x_n - x'_n, A_n(w_n - x_n)\rangle + \langle x_n - x'_n, A_n(x_n - x'_n)\rangle$$

and since $w_n - x_n = -\alpha_n A_n^{-1} \nabla f(x_n)$ we obtain

$$\langle x_n - x'_n, -\alpha_n \nabla f(x_n)\rangle \le -\langle x_n - x'_n, A_n(x_n - x'_n)\rangle$$

or

$$\alpha_n \langle -\nabla f(x_n), p_n\rangle \ge \langle p_n, A_n p_n\rangle . \qquad (5.2)$$

Therefore the direction sequence is feasible. We now show that the condition $\lim_{n \to \infty} \langle \nabla f(x_n), p_n\rangle = 0$ is useful.

Theorem 5.1: Let f be convex, bounded below on the norm closed, bounded, convex set C, and attain its minimum over C at x^*. Let

x_n be a sequence in C such that the projected gradient directions p_n satisfy $\lim_{n \to \infty} \langle \nabla f(x_n), p_n \rangle = 0$ and $\alpha_n \geq \varepsilon > 0$. Then $\{x_n\}$ is a minimizing sequence.

Proof: We write

$$
\begin{aligned}
0 \leq f(x_n) - f(x^*) &\leq \langle \nabla f(x_n), x_n - x^* \rangle \\
&\leq \langle \nabla f(x_n), x_n - x'_n \rangle + \langle \nabla f(x_n), x'_n - x^* \rangle \\
&\leq \langle -\nabla f(x_n), p_n \rangle + \frac{1}{\alpha_n} \langle x_n - \alpha_n A_n^{-1} \nabla f(x_n) - x'_n, A_n(x^* - x'_n) \rangle \\
&\quad + \frac{1}{\alpha_n} \langle x_n - x'_n, A_n(x'_n - x^*) \rangle \\
&\leq \langle -\nabla f(x_n), p_n \rangle + \frac{1}{\alpha_n} \langle x_n - x'_n, A_n(x'_n - x^*) \rangle \text{ by Equation 5.1}
\end{aligned}
$$

Therefore

$$
0 \leq f(x_n) - f(x^*) \leq \langle -\nabla f(x_n), p_n \rangle + \frac{M \| x^* - x'_n \|}{\varepsilon^{\frac{1}{2}} M^{\frac{1}{2}}} [\langle -\nabla f(x_n), p_n \rangle]^{\frac{1}{2}}
$$

which tends to zero.

Q.E.D.

Thus it is reasonable to consider the application of our step-length algorithms to such direction sequences. We consider first the method of Theorem 2.1 .

Corollary 5.1: Let f be bounded below on the norm closed, bounded, convex set C, and let $\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$ for x, y in C. Let $p_n = x'_n - x_n$ where x'_n minimizes $\langle x - w_n, A_n(x - w_n) \rangle$ for x in C with $w_n \equiv x_n - \alpha_n A_n^{-1} \nabla f(x_n)$ with $\{A_n\}$ as described above, i.e., $m\langle x, x\rangle \le \langle A_n x, x\rangle \le M\langle x, x\rangle$. Set $x''_n = x_n + t'_n p_n$ in C and let x_{n+1} be any point in C such that $f(x_{n+1}) \le \beta f(x''_n) + (1-\beta) f(x_n)$ for fixed β in $(0, 1]$. If there exist positive constants $\varepsilon_1, \varepsilon_2$ such that

$$0 < \varepsilon_1 \le t'_n \le \frac{m}{\alpha_n}\left[\frac{2}{L} - \varepsilon_2\right], \quad t'_n \le 1,$$

then $\langle -\nabla f(x_n), p_n\rangle$ tends to zero. If f is convex and $\alpha_n \ge \varepsilon_3 > 0$ then $\{x_n\}$ is a minimizing sequence.

Proof: We seek to use Theorem 2.1. We define

$$\gamma_n \equiv \frac{t'_n \|p_n\|^2}{\langle -\nabla f(x_n), p_n\rangle}$$

and immediately see, since $t'_n \le 1$, that our t'_n equals the t_n of Theorem 2.1; hence, if γ_n satisfies the hypotheses of that theorem, we can conclude that $\langle \nabla f(x_n), p_n\rangle$ tends to zero. For that question we have, by Equation 5.2,

$$\gamma_n = \frac{\|p_n\|^2}{\langle -\nabla f(x_n), p_n\rangle} t'_n \le \frac{\|p_n\|^2 \alpha_n}{\langle p_n, A_n p_n\rangle} t'_n \le \frac{\alpha_n t'_n}{m} \le \frac{2}{L} - \varepsilon_2$$

as required. For the lower bound,

$$\gamma_n = \frac{\|p_n\|^2}{\langle -\nabla f(x_n), p_n\rangle} t'_n \ge \varepsilon_1 \frac{\|p_n\|^2}{\langle -\nabla f(x_n), p_n\rangle}$$

as required. The final conclusions follow from Theorem 5.1.

Q.E.D.

Remark: If we take $0 < \varepsilon_1 \leq \alpha_n \leq \frac{2}{L} - \varepsilon_2$, $t'_n = 1$ for all n, $x_{n+1} = x''_n$, and $A_n = I$ we have the method presented in [20]. The problem with that version is computational; one must know the value of L so far as the first theorem derived for the method stated. Our corollary shows that any point along the gradient direction, so long as α_n is bounded above and away from zero, may be used if x_{n+1} is chosen well; in particular, if $f(x_{n+1}) = \min_{0 \leq t \leq 1} f(x_n + t p_n)$ the method works without knowledge of L.

Arguing as we did prior to Corollary 4.2 concerning the algorithms of Theorems 3.1 and 3.2, we easily prove the following.

Corollary 5.2: Let f be convex and bounded below on the norm closed, bounded, convex set C; let ∇f be uniformly continuous and $\|\nabla f(x)\|$ be uniformly bounded on C. Then the step-size algorithms of Theorems 3.1 and 3.2 applied to the direction algorithm of Corollary 5.1 yields a minimizing sequence $\{x_n\}$ for f over C.

We note that our projected gradient method for $A_n = I$, $H = \mathbb{R}^\ell$, and C a polyhedral set, is not quite the same as the gradient projection method originally described in [23, 24] since that requires that x'_n be the projection onto one of the faces to which x_n belongs or, in some implementations [4], onto a small neighborhood of x_n in C. The computational versions of gradient projection in use apply a special technique near edges of C which turns out to be essentially equivalent to bounding α_n away from zero but keeping it small enough so that the projection is always very near x_n. Thus it is clear that a simple convergence proof for Rosen's original computational gradient projection method can be fashioned in this way from our results above; this has been done [19]. If one however does not take α_n small, one needs a good, efficient method for projection, in an arbitrary quadratic metric, onto a full polyhedral set. Such an algorithm has been brought to our attention [18] and raises the possibility of using larger α_n which may well be more powerful than the original gradient projection approach, at least far away from the solution.

References

1. Altman, M., "Generalized gradient methods of minimizing a functional," Bull. Acad. Polon. Sci., vol. 14, 313-318 (1966).

2. Altman, M., "A generalized gradient method for the conditional minimum of a functional," Bull. Acad. Polon. Sci., vol. 14, 445-451 (1966).

3. Armijo, L., "Minimization of functions having Lipschitz continuous first partial derivatives," Pacific J. Math., vol. 16, 1-3 (1966).

4. Cross, K. E., "A gradient projection method for constrained optimization," Union Carbide Nuclear Division Report K-1746 (1968).

5. Daniel, J.W., "Theory and methods for the approximate minimization of functionals," lecture notes from Ecole déte analyse numerique (1969).

6. Demyanov, V.F., Rubinov, A.M., "The minimization of a smooth convex functional on a convex set," J. SIAM Control, vol. 5 (1967), 280-294.

7. Elkin, R. M., "Convergence theorems for Gauss-Seidel and other minimization algorithms," Computer Sci. Report #68-59, U. of Maryland, College Park (1968).

8. Fletcher, R., Powell, M., "A rapidly convergent descent method for minimization," Computer J., vol. 6, 163-168 (1963).

9. Frank, M., Wolfe, P., "An algorithm for quadratic programming," Nav. Res. Log. Quar., vol. 3, 95-110 (1956).

10. Gilbert, E. G., "An iterative procedure for computing the minimum of a quadratic form on a convex set," SIAM J. Control, vol. 4, 61-81 (1966).

11. Goldfarb, D., "A conjugate gradient method for nonlinear programming," Dissertation, Princeton Univ. (1966).

12. Goldfarb, D., "Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints," to appear in SIAM J. Appl. Math. (1969).

13. Goldfarb, D., "Sufficient conditions for the convergence of a variable metric algorithm," to appear in Proc. Joint Conf. on Opt., U. of Keele (1969).

14. Goldfarb, D., Lapidus, L., "Conjugate gradient method for nonlinear programming problems with linear constraints," I. and E. C. Fundamentals, vol. 7, 142-151 (1968).

15. Goldstein, A. A., "Convex programming in Hilbert space," Bull. Amer. Math. Soc., vol. 70, 709-710 (1964).

16. Goldstein, A. A., "Minimizing functionals on Hilbert space," 159-166 in Computing methods in optimization problems, ed. by Balakrishnan and Neustadt, Academic Press, New York (1964).

17. Goldstein, A. A., "Minimizing functionals on normed linear spaces," J. SIAM Control, vol. 4, 81-89 (1966).

18. Golub, G. H. (1969), private communication.

19. Kreuser, J., Mangasarian, O., private communication (1969).

20. Levitin, E. S., Poljak, B. T., "Constrained minimization methods," (Russian), Zh. vych. Mat. mat. Fiz., vol. 6, 787-823. Also translated in USSR Comput. Math. Math. Phys., vol. 6, 1-50 (1966).

21. Levitin, E. S., Poljak, B. T., "Convergence of minimizing sequences in conditional extremum problems," Soviet Math. Dokl., vol. 7, 764-767 (1966).

22. Poljak, B. T., "Existence theorems and convergence of minimizing sequences in extremum problems with restrictions," Soviet Math. Dokl., vol. 7, 72-75 (1966).

23. Rosen, J. B., "The gradient projection method for nonlinear programming. Part I: linear constraints," J. SIAM, vol. 8, 181-217 (1960).

24. "________. Part II: nonlinear constraints," J. SIAM, vol. 9, 514-532 (1961).

25. Topkis, D. M., Veinott, A. F., Jr., "On the convergence of some feasible direction algorithms for nonlinear programming," SIAM J. Control, vol. 5, 268-279 (1967).

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

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

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
University of Wisconsin	Unclassified
	2b. GROUP

3. REPORT TITLE

CONVERGENCE OF SOME GRADIENT-LIKE METHODS FOR CONSTRAINED MINIMIZATION

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

Computer Sciences Technical Report, July 1969

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

James W. Daniel

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
July, 1969	22	25

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-67-A-0128-0004	Computer Sciences Technical Report 65
b. PROJECT NO	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Releasable without limitations on dissemination

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Mathematics Branch Office of Naval Research Washington, D. C. 20360

13. ABSTRACT

Some step-length algorithms are analyzed and convergence proved for a general class of feasible direction algorithms for constrained minimization. Applications are given to the conditional gradient and variable metric projected gradient methods.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Constrained minimization Step-length algorithms Conditional gradients Variable metric Projected gradient Gradient methods						