CONVERGENCE OF A GRADIENT PROJECTION METHOD*

by

Eli M. Gafni and Dimitri P. Bertsekas

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

*This research was conducted at the M.I.T. Laboratory for Information and Decision Systems with partial support provided by the National Science Foundation under Grant NSF/ECS 79-20834.
Abstract

We consider the gradient projection method $x_{k+1} = P[x_k - \alpha_k \nabla f(x_k)]$ for minimizing a continuously differentiable function $f: H \to \mathbb{R}$ over a closed convex subset $X$ of a Hilbert space $H$, where $P(*)$ denotes projection on $X$. The stepsize $\alpha_k$ is chosen by a rule of the Goldstein-Armijo type along the arc $\{P[x_k - \alpha \nabla f(x_k)] | \alpha > 0\}$. A convergence result for this iteration has been given by Bertsekas [1] and Goldstein [2] for particular types of convex sets $X$. We show the validity of this result regardless of the nature of $X$. 
1. Introduction

We consider the problem

\[
\text{minimize } f(x) \\
\text{subject to } x \in X
\]  

where \( f : H \to \mathbb{R} \) is a continuously Frechet differentiable real-valued function on a Hilbert space \( H \), and \( X \) is a closed convex subset of \( H \). The inner product and norm on \( H \) are denoted \( \langle \cdot, \cdot \rangle \) and \( ||\cdot|| \) respectively. For any \( x \in H \) we denote by \( \nabla f(x) \) the gradient of \( f \) at \( x \), and by \( P(x) \) the unique projection of \( x \) on \( X \), i.e.

\[
P(x) = \arg \min_{z \in X} \{||z-x|| : z \in X\}, \quad \forall x \in H.
\]  

We say that \( x^* \in X \) is a stationary point for problem (1) if \( x^* = P[x^* - \nabla f(x^*)] \).

For any \( x \in X \) we consider the arc of points \( x(\alpha) \), \( \alpha > 0 \) defined by

\[
x(\alpha) = P[x - \alpha \nabla f(x)], \quad \forall \alpha > 0,
\]  

and the class of methods

\[
x_{k+1} = x_k(\alpha_k) = P[x_k - \alpha_k \nabla f(x_k)], \quad x_0 \in X.
\]  

The positive stepsize \( \alpha_k \) in (4) is chosen according to the rule

\[
\alpha_k = \beta^m_k
\]  

where \( m_k \) is the first nonnegative integer \( m \) for which

\[
f(x_k) - f[x_k(\beta^m)] \geq \sigma \langle \nabla f(x_k), x_k - x_k(\beta^m) \rangle,
\]
and where \( s > 0, \beta \in (0,1), \) and \( \sigma \in (0,1) \) are given scalars.

The stepsize rule (5), (6) was first proposed in Bertsekas [1], and reduces to the well known Armijo rule for steepest descent when \( X = H. \)

It provides a simple and effective implementation of the projection method originally proposed by Goldstein [3] and Levitin and Poljak [4] where the stepsize \( \alpha_k \) must be chosen from an interval that depends on a (generally unknown) Lipschitz constant for \( \nabla f \). One of the advantages of the rule (5), (6) is that, for linearly constrained problems, it tends to identify the active constraints at a solution more rapidly than other Armijo-like stepsize rules which search for an acceptable stepsize along the line segment connecting \( x_k \) and \( x_k(s) \) (see e.g., Daniel [5], Polak [6]).

The algorithm is quite useful for large-scale problems with relatively simple constraints, despite its limitation of a typically linear rate of convergence (see Dunn [7]). On the other hand we note that in order for the algorithm to be effective it is essential that the constraint set \( X \) has a structure which simplifies the projection operation.

It was shown in [1] that every limit point of a sequence \( \{x_k\} \) generated by the algorithm (4)-(6) is stationary if the gradient \( \nabla f \) is Lipschitz continuous on \( X \). The same result was also shown for the case where \( H = \mathbb{R}^n \) and \( X \) is the positive orthant but \( f \) is not necessarily Lipschitz continuous on \( X \). Goldstein [2] improved on this result by showing that it is valid if \( H \) is an arbitrary Hilbert space, \( \nabla f \) is continuous (but not necessarily Lipschitz continuous), and \( X \) has the property that

\[ \sigma \| x_k - x_k(\beta^m s) \|^2 \]

replaced by \( \frac{\| x_k - x_k(\beta^m s) \|^2}{\beta^m s} \). Every result subsequently shown for the rule (5), (6) applies to this variation as well.
While it appears that nearly all convex sets of practical interest (including polyhedral sets) have this property, there are examples (Kruskal [8]) showing that (7) does not hold in general. Goldstein [2] actually showed his result for the case where the stepsize $\alpha_k$ in iteration (4) is chosen to be $s$ if

$$f(x_k) - f[x_k(s)] \geq \sigma \langle \nabla f(x_k), x_k - x_k(s) \rangle,$$

and $\alpha_k$ is chosen to be any scalar $\alpha$ satisfying

$$(1-\alpha)\langle \nabla f(x_k), x_k - x_k(\alpha) \rangle \geq f(x_k) - f[x_k(\alpha)] \geq \sigma \langle \nabla f(x_k), x_k - x_k(\alpha) \rangle$$

if (8) is not satisfied. This rule is patterned after the well known Goldstein rule for steepest descent [9]. In what follows we focus attention on the Armijo-like rule (5),(6) but our proofs can be easily modified to cover the case where the algorithm uses a stepsize obtained by the Goldstein rule based on (8) and (9). We also note that Goldstein [2] assumes in addition that $\nabla f$ is uniformly continuous over $X$, but his proof can be easily modified to eliminate the uniformity assumption. By contrast the assumption (7) on the set $X$ is essential for his proof.

The purpose of this paper is to show that the convergence results described above hold without imposing a Lipschitz continuity assumption on $f$, or a condition such as (7) on the convex set $X$. This is the subject of Proposition 2 below. The following proposition establishes that the
algorithm (4)-(6) is well defined.

Proposition 1: For every $x \in X$ there exists $\alpha(x) > 0$ such that

$$f(x) - f[x(\alpha)] \geq \sigma \langle \nabla f(x), x - x(\alpha) \rangle, \quad \forall \alpha \in (0, \alpha(x)]$$

(10)

Proposition 2: If $\{x_k\}$ is a sequence generated by algorithm (4)-(6), then every limit point of $\{x_k\}$ is stationary.

The proofs of Propositions 1 and 2 are given in the next section.

The following lemma plays a key role.

Lemma 3: For every $x \in X$ and $z \in H$, the function $g: (0, \infty) \to \mathbb{R}$ defined by

$$g(\alpha) = \frac{||P(x+\alpha z) - x||}{\alpha}, \quad \forall \alpha > 0$$

is monotonically nonincreasing.

Proof: Fix $x \in X$, $z \in H$ and $\gamma > 1$. Denote

$$a = x + z, \quad b = x + \gamma z$$

(12)

Let $\overline{a}$ and $\overline{b}$ be the projections on $X$ of $a$ and $b$ respectively. It will suffice to show that

$$||\overline{b} - x|| \leq \gamma ||\overline{a} - x||.$$

(13)

If $\overline{a} = x$ then clearly $\overline{b} = x$ so (13) holds. Also if $a \in X$ then $\overline{a} = a = x + z$ so (13) becomes $||\overline{b} - x|| \leq \gamma ||z|| = ||b - x||$ which again holds by an elementary argument using the fact $\langle b - \overline{b}, x - \overline{b} \rangle \leq 0$. Finally if $\overline{a} = \overline{b}$ then
(13) also holds. Therefore it will suffice to show (13) in the case where $a \neq b$, $a \neq x$, $b \neq x$, $a \notin X$, $b \notin X$ shown in Figure 1.
Let $H_a$ and $H_b$ be the two hyperplanes that are orthogonal to $(\overline{b-a})$ and pass through $\overline{a}$ and $\overline{b}$ respectively. Since $<\overline{b-a}, b-b> > 0$ and $<\overline{b-a}, a-a> < 0$ we have that neither $a$ nor $b$ lie strictly between the two hyperplanes $H_a$ and $H_b$. Furthermore $x$ lies on the same side of $H_a$ as $a$, and $x \notin H_a$. Denote the intersections of the line $\{x+\alpha(b-x) | \alpha \in \mathbb{R}\}$ with $H_a$ and $H_b$ by $s_a$ and $s_b$ respectively. Denote the intersection of the line $\{x+\alpha(\overline{a-x}) | \alpha \in \mathbb{R}\}$ with $H_b$ by $w$. We have

$$
\gamma = \frac{||b-x||}{||a-x||} > \frac{||s_b-x||}{||s_a-x||} = \frac{||w-x||}{||a-x||} = \frac{||w-a|| + ||\overline{a-x}||}{||a-x||}$$

$$> \frac{||b-a|| + ||\overline{a-x}||}{||a-x||} \geq \frac{||b-x||}{||a-x||}$$

(14)

where the third equality is by similarity of triangles, the next to last inequality follows from the orthogonality relation $<w-b, b-a> = 0$, and the last inequality is obtained from the triangle inequality. From (14) we obtain (13) which was to be proved. Q.E.D.

2. Proofs of Propositions 1 and 2

From a well known property of projections we have

$$<x-x(\alpha), x - \alpha \nabla f(x) - x(\alpha)> \leq 0, \quad \forall x \in X, \alpha > 0.$$ 

Hence

$$<\nabla f(x), x - x(\alpha)> \geq \frac{||x-x(\alpha)||^2}{\alpha}, \quad \forall x \in X, \alpha > 0.$$ 

(15)
Proof of Proposition 1: If $x$ is stationary the conclusion holds with $\alpha(x)$ any positive scalar so assume that $x$ is nonstationary and therefore $||x-x(\alpha)|| \neq 0$ for all $\alpha > 0$. By the mean value theorem we have for all $x \in X$ and $\alpha \geq 0$

$$f(x) - f[x(\alpha)] = \langle \nabla f(x), x-x(\alpha) \rangle + \langle \nabla f(x(\alpha)) - \nabla f(x), x-x(\alpha) \rangle$$

where $\xi_\alpha$ lies on the line segment joining $x$ and $x(\alpha)$. Therefore (10) can be written as

$$\langle 1-\alpha \rangle \langle \nabla f(x), x-x(\alpha) \rangle \geq \langle \nabla f(x) - \nabla f(\xi_\alpha), x-x(\alpha) \rangle . \quad \text{(16)}$$

From (15) and Lemma 3 we have for all $\alpha \in (0,1]$

$$\langle \nabla f(x), x-x(\alpha) \rangle \geq \frac{||x-x(\alpha)||^2}{\alpha} \geq ||x-x(1)|| ||x-x(\alpha)|| .$$

Therefore (16) is satisfied for all $\alpha \in (0,1]$ such that

$$(1-\alpha)||x-x(1)|| \geq \langle \nabla f(x) - \nabla f(\xi_\alpha), \frac{x-x(\alpha)}{||x-x(\alpha)||} \rangle .$$

Clearly there exists $\alpha(x) > 0$ such that the above relation, and therefore also (16) and (10), are satisfied for $\alpha \in (0,\alpha(x)]$. Q.E.D.

Proof of Proposition 2: Proposition 1 together with (15) and the definition (5),(6) of the stepsize rule show that $\alpha_k$ is well defined as a positive number for all $k$, and that $\{f(x_k)\}$ is monotonically nonincreasing. Let $\bar{x}$ be a limit point of $\{x_k\}$ and let $\{x_k\}_k$ be the subsequence converging to $\bar{x}$. Since $\{f(x_k)\}$ is nontonically nonincreasing we have $f(x_k) \rightarrow f(\bar{x})$. Consider two cases:
Case 1: \( \liminf_{k \to \infty} \alpha_k \geq \bar{\alpha} > 0 \) for some \( \bar{\alpha} > 0 \).

Then from (15) and Lemma 3 we have for all \( k \in K \) that are sufficiently large

\[
f(x_k) - f(x_{k+1}) \geq \frac{\sigma}{\alpha_k^2} \left\| x_k - x_{k+1} \right\|^2 \geq \frac{\sigma}{\bar{\alpha}^2} \left\| x_k - x_{k}(s) \right\|^2
\]

Taking limit as \( k \to \infty, k \in K \) we obtain

\[
0 \geq \frac{\sigma \bar{\alpha} \left\| x - x(s) \right\|^2}{2s^2}
\]

Hence \( \bar{x} = \bar{x}(s) \) and \( \bar{x} \) is stationary.

Case 2: \( \liminf_{k \to \infty} \alpha_k = 0 \).

Then there exists a subsequence \( \{\alpha_k\}_K, \bar{K} \subset K \) converging to zero. It follows that for all \( k \in \bar{K} \) which are sufficiently large the test (6) will be failed at least once (i.e. \( m_k \geq 1 \)) and therefore

\[
f(x_k) - f[x_k(\beta^{-1} \alpha_k)] < \sigma <\!\!\!\!\!\left\langle \nabla f(x_k), x_k - x_k(\beta^{-1} \alpha_k) \right\rangle.
\]

Furthermore for all such \( k \in \bar{K}, x_k \) cannot be stationary since if \( x_k \) is stationary then \( \alpha_k = s \). Therefore

\[
\left\| x_k - x_k(\beta^{-1} \alpha_k) \right\| > 0.
\]
By the mean value theorem we have

\[ f(x_k) - f[x_k(\beta^{-1} \alpha_k)] = \langle \nabla f(x_k), x_k - x_k(\beta^{-1} \alpha_k) \rangle \\
+ \langle \nabla f(\xi_k) - \nabla f(x_k), x_k - x_k(\beta^{-1} \alpha_k) \rangle. \quad (18) \]

where \( \xi_k \) lies in the line segment joining \( x_k \) and \( x_k(\beta^{-1} \alpha_k) \). Combining (16) and (18) we obtain for all \( k \in K \) that are sufficiently large

\[ (1-\sigma) \langle \nabla f(x_k), x_k - x_k(\beta^{-1} \alpha_k) \rangle < \langle \nabla f(\xi_k) - \nabla f(x_k), x_k - x_k(\beta^{-1} \alpha_k) \rangle. \quad (19) \]

Using (15) and Lemma 3 we obtain

\[ \langle \nabla f(x_k), x_k - x_k(\beta^{-1} \alpha_k) \rangle \geq \frac{||x_k - x_k(\beta^{-1} \alpha_k)||^2}{\beta^{-1} \alpha_k} \]

\[ \geq \frac{1}{s} ||x_k - x_k(s)|| ||x_k - x_k(\beta^{-1} \alpha_k)|| \quad (20) \]

Combining (19) and (20), and using the Cauchy-Schwartz inequality we obtain

for all \( k \in K \) that are sufficiently large

\[ \frac{1-\sigma}{s} ||x_k - x_k(s)|| ||x_k - x_k(\beta^{-1} \alpha_k)|| < \langle \nabla f(\xi_k) - \nabla f(x_k), x_k - x_k(\beta^{-1} \alpha_k) \rangle \]

\[ \leq ||\nabla f(\xi_k) - \nabla f(x_k)|| ||x_k - x_k(\beta^{-1} \alpha_k)||. \quad (21) \]

Using (17) we obtain from (21)

\[ \frac{1-\sigma}{s} ||x_k - x_k(s)|| < ||\nabla f(\xi_k) - \nabla f(x_k)||. \quad (22) \]
Since $\alpha_k \to 0$ and $x_k \to \bar{x}$ as $k \to \infty$, $k \in \mathcal{K}$ it follows that $\xi_k \to \bar{x}$, as $k \to \infty$, $k \in \mathcal{K}$.

Taking the limit in (22) as $k \to \infty$, $k \in \mathcal{K}$ we obtain

$$||\bar{x} - \bar{x}(s)|| \leq 0.$$  

Hence $\bar{x} = \bar{x}(s)$ and $\bar{x}$ is stationary. 

Q.E.D.
References


